

GENOMICS IN AQUACULTURE TO BETTER UNDERSTAND SPECIES BIOLOGY AND ACCELERATE GENETIC PROGRESS

EDITED BY: José Manuel Yáñez, Ross Houston and Scott Newman
PUBLISHED IN: Frontiers in Genetics



frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-957-0

DOI 10.3389/978-2-88919-957-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

GENOMICS IN AQUACULTURE TO BETTER UNDERSTAND SPECIES BIOLOGY AND ACCELERATE GENETIC PROGRESS

Topic Editors:

José Manuel Yáñez, University of Chile, Chile

Ross Houston, University of Edinburgh, UK

Scott Newman, Genus, plc, USA



Salmon eyed-eggs.

Image by Aquainnovo S.A.

artificial selection. Thus, there is considerable room for implementing breeding schemes aimed at improving productive traits having significant economic impact. By far the most economically relevant trait is growth rate, which can be efficiently improved by conventional genetic selection (i.e. based on breeding values of selection candidates). However, there are other important traits that cannot be measured directly on selection candidates, such as resistance against infectious and parasitic agents and carcass quality traits (e.g. fillet yield and meat color). However, these traits can be more efficiently improved using molecular tools to assist breeding programs by means of marker-assisted selection, using a few markers explaining a high proportion of the trait variation, or genomic selection, using thousands of markers to estimate genomic breeding values.

From a global perspective aquaculture is an activity related to food production with large potential for growth. Considering a continuously growing population, the efficiency and sustainability of this activity will be crucial to meet the needs of protein for human consumption in the near future. However, for continuous enhancement of the culture of both fish and shellfish there are still challenges to overcome, mostly related to the biology of the cultured species and their interaction with (increasingly changing) environmental factors. Examples of these challenges include early sexual maturation, feed meal replacement, immune response to infectious diseases and parasites, and temperature and salinity tolerance.

Moreover, it is estimated that less than 10% of the total aquaculture production in the world is based on populations genetically improved by means of

The development and implementation of new technologies applied to molecular biology and genomics, such as next-generation sequencing methods and high-throughput genotyping platforms, are allowing the rapid increase of availability of genomic resources in aquaculture species. These resources will provide powerful tools to the research community and will aid in the determination of the genetic factors involved in several biological aspects of aquaculture species. In this regard, it is important to establish discussion in terms of which strategies will be more efficient to solve the primary challenges that are affecting aquaculture systems around the world.

The main objective of this Research Topic is to provide a forum to communicate recent research and implementation strategies in the use of genomics in aquaculture species with emphasis on (1) a better understanding of fish and shellfish biological processes having considerable impact on aquaculture systems; and (2) the efficient incorporation of molecular information into breeding programs to accelerate genetic progress of economically relevant traits.

Citation: Yáñez, J. M., Houston, R., Newman, S., eds. (2016). *Genomics in Aquaculture to Better Understand Species Biology and Accelerate Genetic Progress*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-957-0

Table of Contents

06 *Genomics in aquaculture to better understand species biology and accelerate genetic progress*

José M. Yáñez, Scott Newman and Ross D. Houston

Reviews

09 *Genetics and genomics of disease resistance in salmonid species*

José M. Yáñez, Ross D. Houston and Scott Newman

22 *Applications in the search for genomic selection signatures in fish*

María E. López, Roberto Neira and José M. Yáñez

34 *Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture*

Paulino Martínez, Ana M. Viñas, Laura Sánchez, Noelia Díaz, Laia Ribas and Francesc Piferrer

Mini Reviews

47 *Genetic considerations for mollusk production in aquaculture: current state of knowledge*

Marcela P. Astorga

53 *RNA-seq as a powerful tool for penaeid shrimp genetic progress*

Camilla A. Santos, Danielly V. Blanck and Patrícia D. de Freitas

59 *Appearance traits in fish farming: progress from classical genetics to genomics, providing insight into current and potential genetic improvement*

Nelson Colihueque and Cristian Araneda

Original Research Articles

67 *Primary analysis of repeat elements of the Asian seabass (*Lates calcarifer*) transcriptome and genome*

Inna S. Kuznetsova, Natascha M. Thevasagayam, Prakki S. R. Sridatta, Aleksey S. Komissarov, Jolly M. Saju, Si Y. Ngoh, Junhui Jiang, Xueyan Shen and László Orbán

81 *Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*)*

Jørgen Ødegård, Thomas Moen, Nina Santi, Sven A. Korsvoll, Sissel Kjøglum and Theo H. E. Meuwissen

89 *Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum**

David Marancik, Guangtu Gao, Bam Paneru, Hao Ma, Alvaro G. Hernandez, Mohamed Salem, Jianbo Yao, Yniv Palti and Gregory D. Wiens

104 *Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress*

Jesús Fernández, Miguel Á. Toro, Anna K. Sonesson and Beatriz Villanueva

117 *Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes*

Ali Ali, Caird E. Rexroad, Gary H. Thorgaard, Jianbo Yao and Mohamed Salem

Perspectives

134 *Zebrafish as animal model for aquaculture nutrition research*

Pilar E. Ulloa, Juan F. Medrano and Carmen G. Feijoo

140 *Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals*

Marc Vandeputte and Pierrick Haffray

148 *Genetic improvement of Pacific white shrimp [Penaeus (Litopenaeus) vannamei]: perspectives for genomic selection*

Héctor Castillo-Juárez, Gabriel R. Campos-Montes, Alejandra Caballero-Zamora and Hugo H. Montaldo

Genomics in aquaculture to better understand species biology and accelerate genetic progress

José M. Yáñez^{1,2*}, Scott Newman³ and Ross D. Houston⁴

¹ Faculty of Veterinary and Animal Sciences, University of Chile, Santiago, Chile, ² Aquainnovo, Puerto Montt, Chile, ³ Genus plc, Hendersonville, TN, USA, ⁴ The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

Keywords: aquaculture, genome, breeding programs, QTL, single nucleotide polymorphisms, next-generation sequencing

The production of fish and shellfish through aquaculture is an increasingly important source of high-quality animal protein, with a worldwide production of 66.6 million tons in 2012 (FAO, 2014). Considering the continuously growing global human population and increasing demand for fish products, improvements in the scale, efficiency, and sustainability of aquaculture are essential. To achieve this, several challenges facing the culture of fish and shellfish species need to be overcome. These relate to the diverse biology of the cultured species and their interaction with environmental factors. Examples include outbreaks of infectious diseases, control of sexual maturation, sustainable feed for carnivorous species, and tolerance of diverse and changing environments. This “Frontiers in Livestock Genomics” Research Topic highlights the opportunities offered by recent developments in the field of genomics, and in particular high-throughput sequencing, to contribute to addressing these challenges, with a focus on selective breeding programmes.

The use of selective breeding as a tool to improve the biological efficiency of production in aquaculture generally lags behind plant and farm animal industries, and less than 10% of aquaculture production is based on genetically-improved stocks (Gjedrem et al., 2012). Encouragingly, annual genetic gains reported for aquatic species are in general substantially higher than that of terrestrial farm animals (Gjedrem et al., 2012) and there is considerable scope for achieving significant positive economic impact via improved breeding schemes. However, the status of breeding programs and the level of technology used for aquatic species production are wide-ranging, from use of wild seed stocks through to family-based selection incorporating genomic tools. Family selection and genomic tools can be applied to improve traits that are expensive or difficult to measure on the selection candidates themselves including disease resistance (Yáñez et al., 2014; Ødegård et al., 2014), flesh color (Colihueque and Araneda, 2014; Ødegård et al., 2014) and other appearance traits such as body shape and skin pigmentation (Colihueque and Araneda, 2014) in finfish species. In contrast, despite the global importance of mollusc species for aquaculture, few selective breeding programmes exist and the state of genomic tools and knowledge for these species is typically lacking (Astorga, 2014).

Genomics resources such as whole genome reference sequences, high-density SNP genotyping arrays and genotyping-by-sequencing are in development for several aquaculture species. Fuller characterisation of these resources is underway and is resulting in improved fundamental knowledge of the genome structure and biology, highlighted in this issue by the analysis of repeat elements in the Asian sea bass genome (Kuznetsova et al., 2014). These resources will provide powerful tools for the research community and will aid in the determination of the genetic factors involved in the regulation of complex traits. For example, high-throughput RNA sequencing can give a holistic view of the host response to infectious diseases, and help identify the important genes and pathways defining genetic resistance, as demonstrated in this issue for

OPEN ACCESS

Edited and reviewed by:

Max F. Rothschild,
Iowa State university, USA

*Correspondence:

José M. Yáñez,
jmayanez@uchile.cl

Specialty section:

This article was submitted to
Livestock Genomics, a section of the
journal *Frontiers in Genetics*

Received: 10 February 2015

Accepted: 17 March 2015

Published: 01 April 2015

Citation:

Yáñez JM, Newman S and Houston RD (2015) Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Front. Genet.* 6:128. doi: 10.3389/fgene.2015.00128

rainbow trout (Ali et al., 2014; Marancik et al., 2014) and panaeid shrimp (Santos et al., 2014). Sequencing technology has also facilitated the development of abundant genetic markers that have multi-faceted applications for selective breeding of aquatic species, including parentage assignment in mixed-family environments, providing greater control over family representation and inbreeding (Vandeputte and Haffray, 2014). Medium or high-density SNP arrays can be used to predict genomic breeding values for economically-important traits in well-developed breeding programmes, such as Atlantic salmon (Ødegård et al., 2014). For instance, based on simulations of a Pacific white shrimp breeding program, genetic progress of disease resistance traits is faster with genomic-enabled selection compared to conventional phenotype-based selection due to higher accuracy (Castillo-Juárez et al., 2015). Incorporation of genetic marker information can also be a useful asset to optimize genetic diversity and future genetic gain when establishing base populations for breeding programmes (Fernández et al., 2014). Furthermore, these genomic tools can be applied to investigate putative genomic signatures of selection during the domestication process of farmed fish species, thus potentially identifying genomic regions underlying variation in relevant phenotypes in wild and domestic fish populations (López et al., 2015).

Aquaculture species typically have several common features, for example high fecundity and external fertilization, plus a short evolutionary distance from their wild ancestors. The reproductive features enable flexible mating structures to be used for breeding programmes, and can provide a powerful resource for genetic studies of complex traits, such as disease resistance (Yáñez et al., 2014). However, the diversity between these species is enormous and often necessitates the establishment

of species-specific reproduction and breeding programmes. For example, there is a remarkable variety of sex-determination systems within aquatic farmed species, and the study of Martínez et al. (2014) highlights various methods of controlling sex ratio with aquaculture breeding programmes. This species diversity also presents an issue for choosing suitable model organisms to inform on the biology of the farmed species of interest. Model fin-fish species, such as zebrafish, have been well-characterized and Ulloa et al. (2014) highlight their utility for the evaluation of the response to alternative diets. However, due to the vast evolutionary distance between certain farmed aquatic and model species, it is clear that direct research on the species of interest can often be the most feasible and informative.

The aquaculture industry has often been innovative and visionary in their application of new technologies to improve production. Genomics present another major opportunity, and the research published in this special issue provides several excellent examples of their potential or realized application. Using genomic tools to more effectively utilize genetic variation in economically-important traits via sustainable breeding programmes is paramount to the continued successful growth and stability of aquaculture production.

Acknowledgments

The authors would like to acknowledge funding from Genus plc, CORFO (11IEI-12843 and 12PIE-17669), Government of Chile, Programa U-Inicia, Vicerrectoría de Investigación y Desarrollo, Universidad de Chile, the UK Biotechnology and Biological Sciences Research Council (BBSRC) (BB/H022007/1) and from the Roslin Institute's BBSRC Institute Strategic Funding Grant.

References

- Ali, A., Rexroad, C. E., Thorgaard, G. H., Yao, J., and Salem, M. (2014). Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes. *Front. Genet.* 5:348. doi: 10.3389/fgene.2014.00348
- Astorga, M. P. (2014). Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Front. Genet.* 5:435. doi: 10.3389/fgene.2014.00435
- Castillo-Juárez, H., Campos-Montes, G. R., Caballero-Zamora, A., and Montaldo, H. H. (2015). Genetic improvement of Pacific white shrimp (*Litopenaeus vannamei*): perspectives for genomic selection. *Front. Genet.* 6:93. doi: 10.3389/fgene.2015.00093
- Colihueque, N., and Araneda, C. (2014). Appearance traits in fish farming: progress from classical genetics to genomics, providing insight into current and potential genetic improvement. *Front. Genet.* 5:251. doi: 10.3389/fgene.2014.00251
- FAO. (2014). *The State of World Fisheries and Aquaculture Opportunities and challenges*. Rome: FAO, 243.
- Fernández, J., Toro, M. Á., Sonesson, A. K., and Villanueva, B. (2014). Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Front. Genet.* 5:414. doi: 10.3389/fgene.2014.00414
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Kuznetsova, I. S., Thevasagayam, N. M., Sridatta, P. S. R., Komissarov, A. S., Saju, J. M., Ngoh, S. Y., et al. (2014). Primary analysis of repeat elements of the Asian seabass (*Lates calcarifer*) transcriptome and genome. *Front. Genet.* 5:223. doi: 10.3389/fgene.2014.00223
- López, M. E., Neira, R., and Yáñez, J. M. (2015). Applications in the search for genomic selection signatures in fish. *Front. Genet.* 5:458. doi: 10.3389/fgene.2014.00458
- Marancik, D., Gao, G., Paneru, B., Ma, H., Hernandez, A. G., Salem, M., et al. (2014). Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum*. *Front. Genet.* 5:453. doi: 10.3389/fgene.2014.00453
- Martínez, P., Viñas, A. M., Sánchez, L., Díaz, N., Ribas, L., and Piferrer, F. (2014). Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front. Genet.* 5:340. doi: 10.3389/fgene.2014.00340
- Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., and Meuwissen, T. H. E. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402
- Santos, C. A., Blanck, D. V., and de Freitas, P. D. (2014). RNA-seq as a powerful tool for panaeid shrimp genetic progress. *Front. Genet.* 5:298. doi: 10.3389/fgene.2014.00298
- Ulloa, P. E., Medrano, J. F., and Feijoo, C. G. (2014). Zebrafish as animal model for aquaculture nutrition research. *Front. Genet.* 5:313. doi: 10.3389/fgene.2014.00313
- Vandeputte, M., and Haffray, P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* 5:432. doi: 10.3389/fgene.2014.00432

Yáñez, J. M., Houston, R. D., and Newman, S. (2014). Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5:415. doi: 10.3389/fgene.2014.00415

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Yáñez, Newman and Houston. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetics and genomics of disease resistance in salmonid species

José M. Yáñez^{1,2*}, Ross D. Houston³ and Scott Newman⁴

¹ Faculty of Veterinary and Animal Sciences, University of Chile, Santiago, Chile

² Aquainnovo, Puerto Montt, Chile

³ The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

⁴ Genus plc, Hendersonville, TN, USA

Edited by:

Peng Xu, Chinese Academy of Fishery Sciences, China

Reviewed by:

Zhi-Liang Hu, Iowa State University, USA

Yniv Palti, United States Department of Agriculture, USA

*Correspondence:

José M. Yáñez, Faculty of Veterinary and Animal Sciences, University of Chile, Avenue Santa Rosa 11735, P.O. Box 8820808, La Pintana, Santiago, Chile
e-mail: jmayanez@uchile.cl

Infectious and parasitic diseases generate large economic losses in salmon farming. A feasible and sustainable alternative to prevent disease outbreaks may be represented by genetic improvement for disease resistance. To include disease resistance into the breeding goal, prior knowledge of the levels of genetic variation for these traits is required. Furthermore, the information from the genetic architecture and molecular factors involved in resistance against diseases may be used to accelerate the genetic progress for these traits. In this regard, marker assisted selection and genomic selection are approaches which incorporate molecular information to increase the accuracy when predicting the genetic merit of selection candidates. In this article we review and discuss key aspects related to disease resistance in salmonid species, from both a genetic and genomic perspective, with emphasis in the applicability of disease resistance traits into breeding programs in salmonids.

Keywords: salmon, disease resistance, breeding programs, QTL, genomic selection

INTRODUCTION

Farming of salmonid species is one of the largest aquaculture industries, with a worldwide production of approximately 1.9 million tons of high value product in 2010 (Food and Agriculture Organization of the United Nations [FAO], 2012). As in other animal production systems, the success and sustainability of salmonid aquaculture largely depends on the control of diseases. A clear example of the negative impact of infectious diseases in salmon farming is the unprecedented economic loss caused by outbreaks of the viral disease infectious salmon anemia (ISA) between 2007 and 2009 in Chile (Asche et al., 2010).

Genetic improvement programs are focused on increasing economic return of aquaculture systems via selective breeding (Gjedrem, 2012). In this regard, all heritable and economically relevant traits should be included in the breeding objective. Thus, in salmonid species, traits such as growth rate, flesh color, and resistance to viral, bacterial, and parasitic diseases should be included (Gjedrem, 2000, 2012). Selective breeding can utilize trait information recorded on selection candidates themselves or, particularly in the case of disease or invasive traits, on relatives. Until now, salmon breeding programs have typically included disease resistance based only on information from relatives, which affects the degree of genetic progress achievable on each generation. This is because of the lower accuracy of estimated breeding values (EBVs) when using only sib information compared to the accuracy obtained when using information of the selection candidates themselves (Falconer and Mackay, 1996).

Recent advances in molecular biology techniques, such as next generation sequencing and high throughput genotyping methods, have helped identify genetic variants influencing phenotypic variation for different traits in a wide range of organisms

(Goddard and Hayes, 2009). Molecular markers can be used for a variety of applications in livestock and aquaculture species, such as strain and hybrid identification, genetic variability and genetic diversity evaluation, parentage analyses, quantitative trait loci (QTL) mapping, marker assisted selection (MAS), and genomic selection (GS; Liu and Cordes, 2004; Goddard and Hayes, 2009). Information from a few molecular markers linked to QTL (i.e., genomic regions harboring genes with a significant effect on the trait) might be implemented in breeding schemes through MAS, if they explain a high proportion of genetic variation in the trait. Additionally, the information of 1000s of markers might be simultaneously incorporated into genetic evaluation to estimate genomic breeding values (GEBVs; Meuwissen et al., 2001). These marker-based methods may be particularly useful for the improvement of traits that are complicated or impossible to measure directly on selection candidates, as is the case of resistance to disease (Sonesson and Meuwissen, 2009; Villanueva et al., 2011; Taylor, 2014). A typical first step to implement MAS or GS is to quantify the level of genetic variation in the trait by dissecting its genetic architecture. In salmonids, there is limited information on the genetic architecture of disease resistance. Nevertheless, it is expected that more knowledge on the QTL or genes affecting disease resistance traits will be revealed in the near future, facilitated by the increasing availability of genomic resources and better understanding of the biology of immune response in these species.

This paper reviews aspects of conventional breeding to improve disease resistance in salmonids and the application of molecular tools for the identification of genetic factors involved in these traits. Additionally, the incorporation of molecular information into breeding schemes to improve disease resistance is discussed.

IMPORTANCE OF DISEASE CONTROL IN SALMON FARMING

The health status of farmed fish is one of the main factors affecting the economic return in the salmon industry. Despite scientific, professional, and technical strategies aimed at improving health management, many novel pathological conditions have emerged in salmonid fish species worldwide in recent decades. A detailed description of each disease affecting culture of salmonid species would greatly exceed the purpose of this review. However, some particular examples are discussed to demonstrate the large economic impact diseases can cause in salmon production.

One of the most striking cases affecting salmon farming was the economic crisis triggered by ISA virus outbreaks since mid-2007 in Chile. The production of Chilean Atlantic salmon suffered a dramatic decrease due to increasingly frequent outbreaks between 2007 and 2009. In fact, total production of Atlantic salmon between 2005 and 2010 decreased by more than 60% in volume (Asche et al., 2010). Currently, ISA virus outbreaks appear to be controlled to a low number of events per year in Europe and North and South America. However, the prevalence and emergence of other viral diseases is still of concern. In Northern European countries including Norway and Scotland, ISA outbreaks have been rare in recent years. Nevertheless, infectious pancreatic necrosis (IPN), caused by an aquatic birnavirus, has caused large levels of mortality in Europe, particularly during the window of susceptibility following transfer from freshwater to seawater (Roberts and Pearson, 2005). Other viral diseases have also emerged and pose serious threats to salmon aquaculture, such as skeletal muscle inflammation (HSMI) – a piscine reovirus – and pancreas disease (PD) – an alphavirus, which has shown an increase in recent years (Biering et al., 2012). These viruses cause direct economic losses through mortality and indirect losses through reduced growth rate and treatment costs.

Among bacterial diseases with a negative impact on salmon farming, salmon rickettsial syndrome (SRS) caused by the Gram-negative bacterium *Piscirickettsia salmonis*, is one of the main sanitary challenges in Chilean salmon industry. This disease affects different salmonid species, including Atlantic salmon (*Salmo salar*), coho salmon (*Oncorhynchus kisutch*), and rainbow trout (*O. mykiss*; Fryer and Hedrick, 2003) and can generate economic losses equivalent to 25% of total profit in salmon exports in Chile (Rozas and Enríquez, 2014). Other bacterial diseases, such as those caused by *Aeromonas salmonicida*, *Vibrio anguillarum*, and *Vibrio/Aliivibrio salmonicida*, are recognized to be efficiently controlled by vaccination and do not currently represent a major economic threat for salmon (Biering et al., 2012).

In terms of parasitic diseases, two different species of sea lice, *Lepeophtheirus salmonis* and *Caligus rogercresseyi*, are most detrimental parasites for salmon farming at a worldwide level. In this regard, it has been estimated that on average, the economic impact of sea lice infestation is about 6% of the total value produced by the world salmon industry (Costello, 2009). An emerging threat to salmon production worldwide is Amoebic Gill Disease (AGD), which has been the major disease of farmed salmonid production in Tasmania for several decades (Mitchell and Rodger, 2011) and has appeared relatively recently in most major salmon-producing countries (Ruane and Jones, 2013).

The free-living amoebic protozoan *Neoparamoeba pemaquidensis* is the primary causative agent for the disease that can cause serious morbidity and reduced growth, in addition to increasing susceptibility to other pathogens (Mitchell and Rodger, 2011).

The measures used for prevention and treatment (vaccinations, antibiotics, and antiparasitic drugs, biosecurity measures) of some of the diseases presented above have typically been only partially effective in field conditions (Bravo et al., 2013; Jones et al., 2013; Rozas and Enríquez, 2014). Where effective vaccines do exist, administration typically requires individual handling and treatment of all production fish, which can be expensive and impractical in a large-scale production environment. Due to the fact that improvement in economic efficiency of salmon farming is dependent on disease prevention and control, (Asche and Roll, 2013) it is imperative to develop alternative effective and sustainable strategies. Genetic improvement of disease resistance represents a feasible solution to increase the sanitary status in animal production (Stear et al., 2001; Bishop, 2010). In this regard, there is increasing scientific literature aiming at both quantifying levels of host genetic variation for resistance against different diseases and identifying the specific genetic factors that influence these traits in salmonid species, as discussed below.

CONVENTIONAL BREEDING FOR DISEASE RESISTANCE IN SALMONIDS

Resistance to diseases can be defined as the ability of the host to limit infection by reducing pathogen replication (Råberg et al., 2007; Doeschl-Wilson et al., 2012). Selecting animals with increased resistance to specific diseases is a feasible method to improve productivity and animal welfare and offers advantages over other control methods against infection, such as the cumulative and permanent benefits of the improved resistance (Stear et al., 2001; Bishop, 2010). Disease resistance has been a target trait for the salmon breeding industry for at least 20 years, with a Norwegian salmon breeding program including resistance to bacterial and viral diseases into its breeding goal since 1993 (Gjøen and Bentsen, 1997). However, the study of disease resistance and its incorporation into breeding programs can be hindered by the difficulty in determining and measuring accurate and appropriate phenotypes (Bishop and Woolliams, 2014). This in turn influences the accuracy of disease resistance EBVs that can be achieved. Another limiting step is that disease information is typically only available from relatives of the selection candidates and not directly from the candidates themselves. In the following, we review the main aspects of breeding for resistance to infectious diseases in salmonids and discuss current status and future directions of research in this area.

CHALLENGE AGAINST PATHOGENS

Host resistance to viral and bacterial pathogens can often be measured, in practical terms, as survival (and/or mortality) of individuals during an outbreak (Ødegård et al., 2011). Data and samples from field outbreaks can be used opportunistically to make inference about genetic resistance to infectious diseases. For this purpose, it is necessary that the pedigree of the population be

accurately determined, often using genetic markers or electronic tagging of the fish (Guy et al., 2006). However, using the information from field outbreaks has some disadvantages, such as difficulty to identify the exact cause of death because the factors that influence survival under these conditions are likely to be diverse. Furthermore, the availability of information depends on the occurrence of high-mortality outbreaks, which are usually prevented or controlled to avoid serious economic loss. Moreover, the inference of pedigree using molecular markers can be expensive and laborious. Therefore, survival data are often obtained from experimental challenges, which can readily be standardized to control other variables and potentially allow a clearer interpretation of the results. In this case, it is necessary that a high genetic correlation between the trait measured in experimental and field conditions exists. High genetic correlations ($r_g \geq 0.95$) between field trials and experimental challenges to furunculosis in Atlantic salmon have been reported (Gjøen et al., 1997; Ødegård et al., 2006), suggesting that results from experimental challenges are likely to be directly applicable to commercial production systems. Therefore, challenge tests will often be more accurate and reliable than field outbreaks, due to decreased environmental variability and higher practical feasibility. In fact, challenge testing is currently used to select for resistance to viral, bacterial, and parasitic diseases in breeding programs for Atlantic salmon and rainbow trout (Gjøen and Bentsen, 1997; Leeds et al., 2010; Yáñez and Martínez, 2010; Ødegård et al., 2011; Gjedrem, 2012; Wiens et al., 2013a).

IMMUNOLOGICAL AND PHYSIOLOGICAL VARIABLES AS INDIRECT MEASURES OF RESISTANCE

Direct genetic selection for improved disease resistance based on challenge testing can be costly and time consuming, and has negative animal welfare implications. Furthermore, selection decisions using this strategy can only be carried out using information from relatives and not the candidates themselves. Indirect selection based on the measurement of other characteristics that are genetically correlated with disease resistance, would simplify the data collection and allow the incorporation of individual information. Some studies have aimed at determining the genetic variation of physiological and immunological variables, and the correlation between them and survival in challenge tests in salmon. Examples of variables that have been studied to date are hemolytic activity of serum and lysozyme activity (Røed et al., 1993; Lund et al., 1995), plasma levels of cortisol (Fevolden et al., 1993; Weber et al., 2008), and levels of IgM and antibody titer (Lund et al., 1995), serum $\alpha 2$ -antiplasmin (Salte et al., 1993), bactericidal and complement activity (Hollebecq et al., 1995). However, even when some studies show significant correlations between resistance and immune parameters, the proportion of the total variation in survival that could be explained by immune variables has been considered too low to be useful as a selection criterion. Hence, the prediction of breeding values for survival based on these variables may not be practically useful (Gjøen and Bentsen, 1997). This may be due in part to the complexity of the mechanisms involved in the immune response and the large number of factors that may be involved in disease resistance, which results in a great difficulty when trying to use the

information from a single parameter for the genetic evaluation of disease resistance.

GENETIC VARIATION IN RESISTANCE TO INFECTIOUS DISEASES

A requirement to improve a trait by means of artificial selection is that sufficient genetic variation for this trait exists in the population. Heritability is the proportion of the total phenotypic variance that is attributable to additive genetic variation (Falconer and Mackay, 1996). For disease resistance traits, heritability estimates can vary due to differences in trait definitions and the statistical models used in the analysis (Yáñez and Martínez, 2010; Ødegård et al., 2011). For example, some studies consider the binary trait of survival or mortality as a measure of resistance, whereas others consider the survival time following challenge. These require different statistical approaches to analysis, and are likely to inform on different components of the host response to infection. Nonetheless, there have been several studies aimed at determining levels of additive genetic variation for resistance to different diseases affecting salmonid species (see **Table 1**). The results from these studies show the feasibility of improving disease resistance through genetic improvement and the potential of this approach for helping in the control of disease problems in salmonids.

GENETIC CORRELATIONS BETWEEN DISEASE RESISTANCE AND OTHER TRAITS

The potential to simultaneously improve resistance to different diseases and other economically important traits is partly dependent on the genetic correlations between the traits. Few studies to date have aimed to determine the genetic correlations for disease resistance traits in salmon. Some studies in Atlantic salmon have indicated positive genetic correlations between resistance to bacterial diseases such as furunculosis, BKD, and cold water vibriosis (Gjedrem and Gjøen, 1995; Gjøen et al., 1997). Weak negative genetic correlations between resistance against ISA and bacterial diseases such as furunculosis, vibriosis, and cold-water vibriosis have been reported (Gjøen et al., 1997). However, positive genetic correlations between resistance against ISA virus and furunculosis have also been reported (Ødegård et al., 2007b). In rainbow trout, weak genetic correlations between viral hemorrhagic septicemia (VHS) and bacterial diseases such as enteric redmouth disease (ERM) and rainbow trout fry syndrome have been found (Henryon et al., 2005). Kjøglum et al. (2008) reported only weak genetic correlations when they estimated genetic correlations between resistance to IPN, ISA, and furunculosis. Verrier et al. (2013b) also failed to detect any genetic correlation between host resistance to two rhabdoviral pathogens (VHS and Infectious Hematopoietic Virus). Additionally, weak genetic correlations have also been calculated between resistance against SRS and *C. rogercresseyi* in Atlantic salmon (Yáñez et al., 2014a). In general, these results suggest no clear-cut relationship between genetic resistance to one pathogen and genetic resistance to another pathogen.

Moreover, it is important to know the genetic correlations between disease resistance and other economically important traits in salmon production, especially for selection index development. Previous reports of correlations between disease resistance and production traits have ranged from zero

Table 1 | Heritability values (h^2) and their SE for resistance to different infectious and parasitic diseases in salmonid species.

Salmonid species	Disease agent	$h^2 \pm SE$	Reference
<i>Salmo salar</i>	<i>Neoparamoeba</i> spp.	0.40 ± 0.08 – 0.49 ± 0.09	Taylor et al. (2009)
	<i>Renibacterium salmoninarum</i>	0.2 ± 0.1	Gjedrem and Gjøen (1995)
	<i>Aeromonas salmonicida</i>	0.48 ± 0.17	Gjedrem et al. (1991)
	<i>A. salmonicida</i>	0.34 ± 0.13	Gjøen et al. (1997)
	<i>A. salmonicida</i>	0.38 ± 0.09	Ødegård et al. (2006)
	<i>A. salmonicida</i>	0.43 ± 0.02	Ødegård et al. (2007b)
	<i>A. salmonicida</i>	0.59 ± 0.06	Olesen et al. (2007)
	<i>A. salmonicida</i>	0.62	Kjøglum et al. (2008)
	<i>A. salmonicida</i>	0.47 ± 0.05	Gjerde et al. (2009)
	<i>Gyrodactylus salaris</i>	0.32 ± 0.10	Salte et al. (2009)
	IPNV	0.43	Guy et al. (2006)
	IPNV	0.5	Wetten et al. (2007)
	IPNV	0.55	Kjøglum et al. (2008)
	ISAV	0.13 ± 0.03	Gjøen et al. (1997)
	ISAV	0.16 ± 0.01	Ødegård et al. (2007a)
	ISAV	0.32 ± 0.02	Ødegård et al. (2007b)
	ISAV	0.24 ± 0.03	Olesen et al. (2007)
	ISAV	0.37	Kjøglum et al. (2008)
	ISAV	0.40 ± 0.04	Gjerde et al. (2009)
	<i>Caligus elongatus</i>	0.22	Mustafa and MacKinnon (1999)
	<i>Lepeophtheirus salmonis</i>	0.14 ± 0.02	Kolstad et al. (2005)
	<i>Lepeophtheirus salmonis</i>	0.07 ± 0.02	Glover et al. (2005)
	<i>Vibrio anguillarum</i>	0.38 ± 1.07	Gjøen et al. (1997)
	<i>V. salmonicida</i>	0.13 ± 0.08	Gjedrem and Gjøen (1995)
	<i>Caligus royercesseyi</i>	0.10 ± 0.03	Yáñez et al. (2014a)
	<i>Caligus royercesseyi</i>	0.12 ± 0.07 – 0.34 ± 0.07	Lhorente et al. (2012)
	SPDV	0.21 ± 0.005	Norris et al. (2008)
<i>Piscirickettsia salmonis</i>	0.11 ± 0.02 - 0.41	Yáñez et al. (2013)	
<i>P. salmonis</i>	0.18 ± 0.03	Yáñez et al. (2014a)	
<i>Salvelinus fontinalis</i>	<i>Aeromonas salmonicida</i>	0.51 ± 0.03	Perry et al. (2004)
<i>Oncorhynchus tshawytscha</i>	<i>R. salmoninarum</i>	0 ± 0.05	Beacham and Evelyn (1992)
	<i>A. salmonicida</i>	0.21 ± 0.14	Beacham and Evelyn (1992)
	<i>V. anguillarum</i>	0.14 ± 0.11	Beacham and Evelyn (1992)
<i>O. kisutch</i>	<i>R. salmoninarum</i>	0.53 ± 0.16	Withler and Evelyn (1990)
<i>O. mykiss</i>	<i>Yersinia ruckeri</i>	0.21 ± 0.05	Henryon et al. (2005)
	<i>Flavobacterium psychrophilum</i>	0.35 ± 0.09	Silverstein et al. (2009)
	<i>F. psychrophilum</i>	0.23 ± 0.03	Leeds et al. (2010)
	<i>F. psychrophilum</i>	0.53 ± 0.09	Vallejo et al. (2010)
	<i>F. psychrophilum</i>	0.07 ± 0.02	Henryon et al. (2005)
	VHS	0.63 ± 0.26	Dorson et al. (1995)
	VHS	0.11 ± 0.1	Henryon et al. (2005)

(Silverstein et al., 2009), low to moderately negative (Henryon et al., 2002), inconsistent (Beacham and Evelyn, 1992; Henryon et al., 2002), and low to moderately positive (Gjedrem et al., 1991; Perry et al., 2004; Yáñez et al., 2014a). Thus, it is important to

evaluate these parameters for each particular breeding population to maximize the impact of selecting for improved disease resistance. Additionally, the relationship between ploidy levels and disease resistance is of interest because all female triploid rainbow

trout are advantageous for production. Weber et al. (2013) showed that triploid fish are generally slightly more susceptible to Bacterial Coldwater Disease than diploids, but that selection for improved resistance in diploids is also effective for triploid production.

RESISTANCE VERSUS TOLERANCE

If resistance is defined as an individual's ability to block the reproduction of a pathogen, then disease tolerance can be defined as the ability to limit the impact of infection on a host (Råberg et al., 2007; Doeschl-Wilson et al., 2012). Information relating to the genetic basis of disease tolerance has been sparse in animal genetics studies to date (Doeschl-Wilson et al., 2012). However, it is important to disentangle resistance from tolerance because the appropriate trait to select for may differ depending on the disease, host species, and environment. It is also possible that the two traits will be antagonistic, which could result in inadvertent undesirable outcomes of selection for disease resistance (Doeschl-Wilson et al., 2012). The implications of selection for resistance or tolerance on the host–pathogen interaction and pathogen evolution have also been considered. For example, selection for resistance *per se* may lead to selection pressure for higher virulence in the pathogen, whereas selection for tolerance could result in co-existence of pathogen and host with minimal impact on the performance of the host population. Therefore, in future disease studies and in selective breeding programs, it will be important to carefully consider the optimal disease trait to target for maximum benefit at the population level, but also the analytic issues of resistance versus tolerance.

STRATEGIES FOR GENETIC DISSECTION OF DISEASE RESISTANCE TRAITS

Major advances in nucleotide sequencing, ever-improving bioinformatics pipelines, and high-throughput genotyping tools have helped to identify genes associated with complex traits in various species of vertebrates. In salmonids, this is reflected in the substantial increase in genomic resources for these species during recent years and, for example, in the formation of an international collaboration to sequence the Atlantic salmon genome (Davidson et al., 2010). The salmon genome has undergone a relatively recent duplication and has a very high content of long repeat elements. This has hampered the sequencing and assembly of a reference genome for salmonids. However, a high-quality reference sequence has been published for rainbow trout (Berthelot et al., 2014) and is available for Atlantic salmon (Davidson et al., 2010). Further, high density SNP genotyping arrays were recently developed for Atlantic salmon (Houston et al., 2014; Yáñez et al., 2014b), and a lower density platform (Lien et al., 2011) has previously been used for QTL mapping and population genetics. Currently, these genomic resources are increasingly being used for the identification of the genetic factors involved in the resistance to different diseases in salmonids. The strategies used in the study of the genetic architecture of disease resistance can be classified as: (i) candidate gene approaches, (ii) QTL mapping, and (iii) gene expression studies.

CANDIDATE GENE APPROACH

Candidate gene theory states that a significant proportion of phenotypic variance of one trait in a population is determined by the

presence of polymorphisms within genes known to be involved in the physiological regulation of that trait (Rothschild and Soller, 1997). This approach requires previous knowledge on the biology of the species, biochemical pathways, and especially gene sequences, to study the variation within specific candidate genes. In aquaculture species, the availability of annotated gene sequences of known function is typically low, but is likely to increase in the short term with the use of high-throughput sequencing and ongoing genome sequencing projects.

In vertebrates, the major histocompatibility complex (MHC) has attracted much attention in studies of association between genetic variants and disease resistance. However, other genes are likely to play an important role in the mechanisms of disease resistance in production animals, model organisms, and humans (Hill, 1999; Qureshi et al., 1999). To our knowledge, there are no studies aimed at establishing association between candidate genes, other than the MHC, and resistance to infectious diseases in salmonid species.

Major histocompatibility complex

The MHC is a multigene family that acts at the interface between the immune system and infectious pathogens. The MHC gene family comprises two subfamilies: class I and II. Both classes are membrane glycoproteins involved in the processing and removal of pathogens (Thorgaard et al., 2002). MHC genes have been identified, cloned, and characterized in Atlantic salmon, rainbow trout, and other salmonids (Grimholt et al., 1993; Hordvik et al., 1993; Hansen et al., 1996; Shum et al., 1999, 2002). Furthermore, it has been shown that these genes are highly polymorphic in these species (Grimholt et al., 1994, 2002; Miller and Withler, 1996; Hansen et al., 1999; Garrigan and Hedrick, 2001; Aoyagi et al., 2002). As in other vertebrates there are two types of class I genes; the UAA which are highly divergent, non-polymorphic and expressed at low levels, and the UBA which are polymorphic, expressed at high levels in spleen, and with structural features similar to those of class Ia molecules (classical) which present antigen to T lymphocytes (Shum et al., 1999). The class II genes are divided into Class II A (DAA) and II B (DAB), depending on whether encoding α or β chain of the molecule, respectively (Grimholt et al., 2000; Stet et al., 2002). Both loci (DAA and DAB) co-segregate as haplotypes, suggesting a close physical linkage between them in Atlantic salmon (Stet et al., 2002).

The association between a polymorphism linked to MHC class II genes and resistance to virus infectious hematopoietic necrosis (IHN) in backcrosses of rainbow trout and cutthroat trout (*O. clarki*) has been found, but was relatively weak and dependent on the family analyzed (Palti et al., 2001). Suggestive associations between rainbow trout MHC class IB alleles and bacterial cold water disease have also been shown (Johnson et al., 2008). In Atlantic salmon, the association between MH class IIB alleles and resistance against *A. salmonicida* has been reported (Langefors et al., 2001; Lohm et al., 2002). In the same species, MHC class I and class II variants have been associated with susceptibility to IHN (Miller et al., 2004) and resistance to furunculosis and ISA (Grimholt et al., 2003; Kjøglum et al., 2006). Although associations between MH gene variants

and disease resistance have been established, the MHC is most likely not the only factor influencing genetic variation in disease resistance. For instance, in Atlantic salmon, a non-MHC effect for resistance to IPN, furunculosis and ISA has been detected (Kjøglum et al., 2005). Because disease resistance traits will typically be polygenic in nature, it is important to consider variants in a genome-wide context and possible interactions that can occur between genes (epistasis), which hinders the candidate gene approach as a comprehensive strategy for incorporating molecular information into the genetic evaluation of these traits.

QTL MAPPING

Quantitative trait loci mapping is a strategy providing information on the location and effect of the gene variants influencing complex quantitative traits, but without prior hypotheses. The QTL detection methodologies are based on the use of (typically anonymous) DNA markers dispersed throughout the genome to identify genomic regions involved in the genetic variation of a particular trait, by means of statistical analyses utilizing the co-segregation between markers and the (unknown) causative variants.

DNA markers

The development of DNA markers has had a major impact on studies of genetic variation in animals and fish. The categories of DNA markers widely used historically include Restriction Fragment Length Polymorphisms (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), microsatellites, and single nucleotide polymorphisms (SNPs). More recently, SNP markers have become the predominant marker due to their abundance, ease of discovery and low cost of genotyping per locus (Houston et al., 2014). However, each of these marker types vary from each other in their mode of inheritance (i.e., dominant or codominant), identification and detection methods, number of spanning loci and polymorphic information content (Liu and Cordes, 2004).

Microsatellites are tandem repeat nucleotide sequences that generally span between one to six base pairs. Different alleles are generated due to the variation in the number of repeats. Their main useful features for genetic studies include high variability, codominant inheritance, abundance, and wide distribution across the genome. One of the major advantages of microsatellites is their high degree of polymorphism, with 10s of alleles often observed at a single locus in an outbred population. Additionally, their genotyping is typically based on a simple DNA amplification by means of polymerase chain reaction (PCR). Thus, genotyping can be relatively rapid, cheap, and the amount of DNA required is minimal (nanograms). However, the scoring of microsatellites often requires optimization and manual input, which reduces their scalability for large genetic studies. There are a considerable number of microsatellites identified for different salmonid species, available for use in genetic studies (e.g., Cairney et al., 2000; Gilbey et al., 2004; Phillips et al., 2009).

Historically, AFLPs (which typically score SNPs) had the advantage of being generated more easily and cheaply than SNPs and microsatellites, because previous knowledge on genomic sequence

is not needed for their generation. However, their mode of inheritance, similar to RAPDs, is dominant, i.e., it is not possible to distinguish between heterozygous and one category of homozygous genotypes without the use of special equipment and software (Piepho and Koch, 2000). This reduces the amount of information provided by these kinds of markers.

Single nucleotide polymorphisms have several key advantages for genetic studies, which have led to a rapid increase in their popularity over recent years. Some of these include high abundance, amenability to automated scoring in large numbers, simultaneous assays of thousands of markers (e.g., using SNP arrays) and presence in both coding and non-coding regions. Therefore, SNPs have been applied recently for construction of dense genetic maps, which can be used for fine mapping of QTLs and facilitate the identification of causative genes involved in the genetic variation of specific characters (e.g., Lien et al., 2011; Gonen et al., 2014). Recently, there has been a rapid increase in available SNPs for salmonid species, mainly for Atlantic salmon and rainbow trout (Hayes et al., 2007; Sanchez et al., 2009; Everett et al., 2011, 2012; Hohenlohe et al., 2011; Lien et al., 2011; Houston et al., 2012, 2014; Salem et al., 2012). Additionally, SNP arrays are available for simultaneous genotyping of 10s to 100s of 1000s of markers in rainbow trout (Palti et al., 2014b) and Atlantic salmon (Houston et al., 2014; Yáñez et al., 2014b). These SNP resources are likely to be increasingly applied to high-resolution mapping of disease resistance genes in salmonid species.

In recent years, the same sequencing technology that has led to the increased ease of high-density SNP discovery described above has enabled direct genotyping of individual fish based on sequence information alone. Such techniques are collectively termed 'genotyping by sequencing' (GBS). Although GBS techniques encompass a diverse range of laboratory and bioinformatic pipelines, they are all based on the principle of using nucleotide barcodes ligated to the fragmented genomic DNA of individual fish and high-throughput sequencing in multiplexed pools (Davey et al., 2013). Partly due to the lack of established genomics resources for many farmed fish species, the aquaculture genetics community has been early adopters and developers of these techniques. In particular, RAD sequencing has been utilized in both Atlantic salmon (e.g., Houston et al., 2012) and rainbow trout (Palti et al., 2014a) as a conduit to incorporating genomic information into aquaculture breeding programs. While the application of GBS has huge potential for applied research in salmonids, there are also some challenges in managing and interpreting these datasets. For example, discovery of 'false' SNPs using sequencing can occur, particularly with the recent whole genome duplication of the salmonid species. Therefore, species-tailored bioinformatics pipelines are typically required to minimize these issues.

Detection of QTL affecting disease resistance in salmonid species

Salmon may be more likely than terrestrial farmed species to have QTL of major effect because they have had fewer generations of selection in the farmed environment and therefore the standing genetic variation for traits of economic importance may still be very large. The development of a genetic map based on linkage between genetic markers is the first step towards the

identification of QTL. To date, several linkage maps have been constructed using different marker types for rainbow trout (e.g., Young et al., 1998; Sakamoto et al., 2000; Nichols et al., 2003; Guyomard et al., 2006; Rexroad et al., 2008; Palti et al., 2011, 2012; Guyomard et al., 2012); Atlantic salmon (Gilbey et al., 2004; Moen et al., 2004b; Lien et al., 2011; Gonen et al., 2014); coho salmon (McClelland and Naish, 2008); brown trout (*Salmo trutta*; Gharbi et al., 2006); Arctic Trout (*Salvelinus alpinus*; Woram et al., 2004); and sockeye salmon (*Salmo nerka*; Everett et al., 2012).

Using microsatellite markers, two QTL with major effects on resistance to IPN in rainbow trout has been detected. These loci explained a large proportion (27 and 34%) of the phenotypic variation in a family from a backcross between a strain susceptible to IPN (YK-RT101) and a resistant one (YN-RT201; Ozaki et al., 2001). Using AFLP and microsatellite markers, QTLs for IHN resistance have been identified in three different linkage groups of the same species (Rodriguez et al., 2004). For the same disease, RFLP markers were associated with resistant and susceptible families in backcrosses of rainbow trout and cutthroat trout (Palti et al., 1999). Extensive research into the genetic architecture of resistance to Bacterial Coldwater Disease has also been undertaken, with evidence for QTL of major effect (Wiens et al., 2013b; Vallejo et al., 2014a,b). Additionally, major QTL have been detected for resistance to whirling disease caused by the myxosporean parasite *Myxobolus cerebralis* (Baerwald et al., 2010) and VHS (Verrier et al., 2013a).

In Atlantic salmon, using AFLP markers, two QTL associated with resistance to ISA have been detected in two full sib families (Moen et al., 2004a). One of these QTLs has been validated using microsatellite markers from a higher number of genotyped fish. This QTL explained 6% of the phenotypic variation for resistance to ISA and has been mapped to linkage group VIII of the Atlantic salmon genome (following SALMAP notation; Moen et al., 2007). In the same species, a major QTL for resistance to IPN has been identified using data collected from a 'field' seawater outbreak of the disease (Houston et al., 2008a). The QTL detection strategy involved utilizing the low male recombination rate observed in salmonids by using just two to three microsatellites per chromosome and male segregation to determine linkage groups with a significant effect, and secondly, a higher number of markers per linkage group. Female segregation was used to confirm the previously detected QTL and position it within the linkage group (Houston et al., 2008a). The major QTL, mapped to linkage group 21, was subsequently confirmed by analyzing nine additional families and a higher saturation of markers (Houston et al., 2008b). The same QTL was then confirmed and fine mapped in an independent population, and haplotypes of markers that could predict the genotype at the QTL were identified based on the linked microsatellites (Moen et al., 2009). Using a restriction-site associated DNA sequencing approach (RAD sequencing), several additional SNP markers linked to the QTL were identified and two SNP markers showed a significant population-level association with resistance in two year classes from an Atlantic salmon breeding population (Houston et al., 2012). The high proportion of the total variance for IPN resistance that this QTL explains has allowed the incorporation of markers linked to it into MAS

schemes for the genetic improvement of this trait in Atlantic salmon in both Norway (Moen et al., 2009) and Scotland (Houston et al., 2010).

In general, confidence intervals for mapped QTLs mapped are large. This issue has two consequences: first, widespread confidence intervals might contain a large number of genes (1000s) and, therefore, identification of the causative polymorphism is challenging. Second, the use of markers linked to QTLs in MAS programs is complicated, since the linkage phase between the marker and the QTL throughout the population may be different from family to family. An alternative for QTL fine mapping is to use information from linkage analysis in conjunction with information from linkage disequilibrium (LD) across the population (Meuwissen et al., 2002). Through simulation studies, power and accuracy of this combined approach has been successfully tested for QTL fine mapping, accounting for the structure of commercial salmon populations (Hayes et al., 2006). Additionally, the availability of both high-density SNPs marker panels and high-resolution genetic maps will contribute to detect association between markers and QTLs with higher precision by means of using across-population LD mapping (Goddard and Hayes, 2009). These strategies along with a reference sequence and a consolidated physical map of the genome, will facilitate the identification of causative mutations affecting disease resistance traits through positional studies in salmonid species.

GLOBAL GENE EXPRESSION

Functional genomics, defined as the application of experimental methods of genomic or systemic coverage to assess gene function using data from structural genomics (mapping and sequencing), has been recognized as an area of primary interest in disease studies (Hiendleder et al., 2005). These methodologies broaden the spectrum of biological research to study, simultaneously, the expression of thousands of genes at the transcriptional level. Currently, genomic resources and new sequencing technologies have helped to assess differential gene expression levels in the response against diseases in salmonid species. These data can help to pinpoint functional genetic variation underlying disease resistance.

An early example, using suppression subtractive hybridization (SSH) and liver samples from individuals injected with a *V. Anguillarum* bacterium and normal individuals, more than 25 genes important in the immune response in rainbow trout were identified, including sequences of proteins of acute phase of inflammation, complement, and coagulation system (Bayne et al., 2001). Using the same technique, genes involved in signal transduction and innate immunity (among others) have been identified as relevant factors in response to a challenge against *A. salmonicida* in Atlantic salmon (Tsoi et al., 2004).

The availability of ESTs (expression sequence tags) and cDNA libraries have allowed the development of DNA microarrays which can be used to study the differential expression patterns of a large number of genes simultaneously in salmonids (Rise et al., 2004b; Ewart et al., 2005; von Schalburg et al., 2005). Using a microarray of human cDNAs, differentially expressed transcripts against a challenge with *A. Salmonicida* have been identified in Atlantic salmon.

However, due to species divergence, only 6% of the sequences of the microarray showed detectable hybridization against salmon liver cDNA (Tsoi et al., 2003). Rise et al. (2004a) conducted the first study using a microarray constructed from salmon cDNA libraries. Following this, differential gene expression has been assessed in between macrophages infected and uninfected with *P. salmonis* and hematopoietic kidney from Atlantic salmon individuals challenged or not against the same pathogen. Differentially expressed genes were proposed to be relevant in immune response and as potential biomarkers of infection with *P. salmonis* (Rise et al., 2004a). Additionally, using an Atlantic salmon cDNA microarray including more than 4,000 genes extracted from liver, spleen and hematopoietic kidney, several differentially expressed genes in response to infection by *A. salmonicida* have been found (Ewart et al., 2005). Furthermore, the transcriptomic response against vaccination with an *A. salmonicida* bacterium has been assessed in Atlantic salmon, revealing temporal and tissue differences in terms of expression levels, which may be relevant in establishing protection (Martin et al., 2006). In addition, gene expression profiles in response to a DNA vaccine for the IHNV virus has been studied in rainbow trout, identifying 910 genes modulated in the injection site, and also determining the over-expression of genes of the type I Interferon system (IFN-1) in other tissues, suggesting that this system forms the basis of early antiviral immunity (Purcell et al., 2006). In the studies presented above, certain transcripts that showed variation in their expression levels during infection have low levels of homology with well-characterized genes available in public databases and, thus, do not have a known function (Rise et al., 2004a; Ewart et al., 2005; Martin et al., 2006; Purcell et al., 2006). Therefore, studies based on sequences of the encoded proteins aimed at decrypting their role in the immune response against infection are still needed.

To date, few studies have focused on the analysis of differential expression between resistant and susceptible fish for a particular disease. For example, Sutherland et al. (2014) analyzed gene expression profile differences between chum and pink salmon during infections with sea lice (*L. salmonis*) to gain insight into the functional mechanisms underlying the divergent resistance to lice observed in these two species. Zhang et al. (2011) took a similar approach to examine the differences between resistant and susceptible salmon families to the pathogen *A. salmonicida* to highlight the importance of several innate immune response genes. Cofre et al. (2014) also examined differential expression of several candidate genes in families with divergent resistance to IPNV. Finally, Langevin et al. (2012) compared the transcriptional response of resistant and susceptible clonal lines for Bacterial Coldwater Disease. The information provided by this type of analysis might be useful in the discovery of new sets of genes, with or without an assigned function, which may be associated with disease resistance (Walsh and Henderson, 2004). Another possibility is to consider the differential expression levels as a quantitative trait. This may allow identification of QTLs associated with differences in gene expression patterns between resistant and susceptible individuals (eQTL; Pomp et al., 2004; de Koning et al., 2005). However, it remains unclear how the information given by expression analysis can be used in breeding programs (Walsh and Henderson, 2004).

MOLECULAR MARKER-ASSISTED SELECTION (MAS) AND GENOMIC SELECTION (GS)

While selective breeding based on performance information from the selection candidate and its relatives is a highly successful means of genetic improvement in the trait of interest, utilizing genetic markers can provide an improvement in both genetic gain and selection accuracy (Goddard and Hayes, 2009). Historically, the use of genetic markers required mapping of QTL and therefore a key parameter for subsequent application in MAS programs is the level of LD between markers and causative mutations at a population level (Goddard and Meuwissen, 2005).

One factor reducing the level of LD each generation is recombination. When the LD between the QTL and the marker only exists within families and not across families, recombination can break the association between marker alleles and the QTL between families. Therefore, the linkage phase between the marker and the QTL should be determined in each generation and separately for each family if it is to be utilized in selection (Wientjes et al., 2013). To determine if the marker and QTL are in LD within each family phenotypic records and genotypes are needed on each generation. This makes unattractive the implementation of MAS exploiting only within-family LD (i.e., linkage) between the QTL and markers (Dekkers and Van der Werf, 2007). In the case of disease resistance traits, this means that all families comprising the breeding nucleus would need to be challenged each generation.

Marker assisted selection schemes are likely to be optimal when the markers explain a large proportion of the total variance of the trait, as in the case of the QTL for resistance to IPN in Atlantic salmon (Houston et al., 2008a, 2012; Moen et al., 2009). The information from markers linked to this QTL is being applied in MAS programs in different breeding populations of Atlantic salmon. Although SNP variants associated with IPN resistance in two different year classes of a breeding nucleus have been discovered (Houston et al., 2012), there is still a need to validate if these alleles are associated with resistance in independent Atlantic salmon populations. This will depend on the LD relationship between the markers and the causative mutation in the population(s) of interest. As such, testing the association between markers and IPNV mortality in these populations should be considered a prerequisite for effective commercial application.

Another alternative to exploit the LD at the population level is using information from high density panels of markers to predict GEBVs of selection candidates (Meuwissen et al., 2001). This approach effectively takes into account all markers when estimating the breeding value of the candidate, without the need to surpass a significance threshold for association with a particular trait of interest. The effectiveness of this strategy will also depend on the magnitude of the effects associated with the markers. When the effects of the markers across the entire genome are estimated, these effects can be used to select individuals lacking phenotypes (Meuwissen et al., 2001). As such, in the case of disease resistance, there would be fewer requirements for challenge testing of siblings for diseases of economic importance. Recombination results in decay of LD on each generation and the magnitude of this reduction depend on several population features (Porto-Neto et al., 2014). This means that in practice, it is necessary to corroborate the

accuracy in estimating the GEBVs and response selection in each generation.

Best linear unbiased prediction (BLUP) combines information from pedigree and phenotypes to predict breeding values (EBVs) of individuals. Molecular markers provide a new source of information, which will give higher accuracy when predicting EBVs. Thus, selection response will be potentially higher in traits in which the accuracy is low, i.e., traits with low heritability or traits that cannot be measured in the selection candidate, such as disease resistance (Sonesson and Meuwissen, 2009; Villanueva et al., 2011; Taylor, 2014). The relative increase in accuracy depends on the amount of variation explained by the markers. It is also possible to combine breeding values for disease resistance and other performance traits into a selection index with specified weightings based on their economic importance for multi-trait genetic improvement.

In livestock species the effects of QTL have been shown to exhibit a moderate leptokurtic gamma distribution, suggesting a small number of loci of large effect and high number of loci of small effect (Hayes and Goddard, 2001), which is likely to be expected in aquaculture species for most economically relevant traits. Therefore, it is expected that more than one marker is necessary to efficiently assist breeding programs using molecular information. The availability of high throughput SNP genotyping platforms will allow the use of these markers in selective breeding for aquaculture species by means of the implementation of GS schemes (Goddard and Hayes, 2009; Sonesson and Meuwissen, 2009; Taylor, 2014). There are currently at least three independent initiatives to generate high density SNP arrays for Atlantic salmon, run by groups from three of the leading producers of this species: Chile, Scotland and Norway. The first of these to be published is that of Houston et al. (2014) whereby approximately 132 K SNP markers were identified and verified in several populations of farmed and wild Atlantic salmon. The Chilean group validated almost 160 K SNP markers from a 200 K SNP platform for Atlantic salmon, genotyping fish representing wild and farmed populations from Europe, North America, and Chile (Yáñez et al., 2014b). Additionally, a 57 K SNP chip is available for rainbow trout (Palti et al., 2014b). Therefore, it can be expected that programs using a high number of markers in population LD will be implemented in the near future. However, it is necessary to determine the economic benefit of using high-density panels, versus the use of low-density panels supported by imputation methodologies and establish a strategy for GS taking into account technical and economic feasibility.

CONCLUSION

There is increasing information related to the determination of the genetic basis for disease resistance in salmonid species. The rapid development of genomic resources in these species will provide new tools for genetic dissection of these traits. The use of these tools will be of great help in identifying loci involved in the genetic variation of disease resistance. This information will increase our understanding of the underlying biology of resistance to disease. Further, it will be crucial for the implementation of MAS or GS programs that include disease resistance within the breeding objective. These methods will increase the accuracy of selection

candidates, thereby improving the selection response. However, the economic feasibility and profitability of the implementation of these new strategies and its comparison with conventional selection schemes must be studied for each particular breeding program in salmonids. It will also be necessary to assess the long-term impact of these strategies on the control of each specific disease, in an epidemiological context.

ACKNOWLEDGMENTS

This work was partially funded by grants from CORFO (11IEI-12843 and 12PIE-17669), Government of Chile and from Programa U-Inicia, Vicerrectoría de Investigación y Desarrollo, Universidad de Chile. José M. Yáñez would like to thank María E. López and Liane Bassini for their help in revising the manuscript.

REFERENCES

- Aoyagi, K., Dijkstra, J. M., Xia, C., Denda, I., Ototake, M., Hashimoto, K., et al. (2002). Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (*Oncorhynchus mykiss*). *J. Immunol.* 168, 260–273. doi: 10.4049/jimmunol.168.1.260
- Asche, F., Hansen, H., Tveteras, R., and Tveteras, S. (2010). The salmon disease crisis in Chile. *Marine Resour. Econ.* 24, 405–411. doi: 10.5950/0738-1360-24.4.405
- Asche, F., and Roll, K. H. (2013). Determinants of inefficiency in Norwegian salmon aquaculture. *Aquac. Econ. Manag.* 17, 300–321. doi: 10.1080/13657305.2013.812154
- Baerwald, M. R., Petersen, J. L., Hedrick, R. P., Schisler, G. J., and May, B. (2010). A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). *Heredity (Edinb.)* 106, 920–926. doi: 10.1038/hdy.2010.137
- Bayne, C. J., Gerwick, L., Fujiki, K., Nakao, M., and Yano, T. (2001). Immune relevant (including acute phase) genes identified in the livers of rainbow trout, *Oncorhynchus mykiss*, by means of suppression subtractive hybridization. *Dev. Comp. Immunol.* 25, 205–217. doi: 10.1016/S0145-305X(00)00057-4
- Beacham, T. D., and Evelyn, T. P. T. (1992). Genetic variation in disease resistance and growth and growth of chinook, coho, and chum salmon with respect to vibriosis, furunculosis, and bacterial kidney disease. *Trans. Am. Fish Soc.* 121, 456–485. doi: 10.1577/1548-8659(1992)121<0456:GVIDRA>2.3.CO;2
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5, 3657. doi: 10.1038/ncomms4657
- Biering, E., Madhun, A., Isachsen, C. H., Omdal, L. M., Einen, A. C. B., Garseth, Å. H., et al. (2012). *Annual Report on Health Monitoring of Wild Anadromous Salmonids in Norway. Annual Report, 2012, No 6-2013*. Bergen: Institute of Marine Research.
- Bishop, S. C. (2010). “Disease resistance: genetics,” in *Encyclopedia of Animal Science*, eds W. G. Pond and A. W. Bell (New York, NY: Marcel Dekker, Inc.) 288–290.
- Bishop, S. C., and Woolliams, J. A. (2014). Genomics and disease resistance studies in livestock. *Livest. Sci.* 166, 190–198. doi: 10.1016/j.livsci.2014.04.034
- Bravo, S., Nunez, M., and Silva, M. T. (2013). Efficacy of the treatments used for the control of *Caligus rogercresseyi* infecting Atlantic salmon, *Salmo salar* L., in a new fish-farming location in Region XI, Chile. *J. Fish Dis.* 36, 221–228. doi: 10.1111/jfd.12023
- Cairney, M., Taggart, J. B., and Høyheim, B. (2000). Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. *Mol. Ecol.* 9, 2175–2178. doi: 10.1046/j.1365-294X.2000.105312.x
- Cofre, C., Gonzalez, R., Moya, J., and Vidal, R. (2014). Phenotype gene expression differences between resistant and susceptible salmon families to IPNV. *Fish Physiol. Biochem.* 40, 887–896. doi: 10.1007/s10695-013-9894-3
- Costello, M. J. (2009). The global economic cost of sea lice to the salmonid farming industry. *J. Fish Dis.* 32, 115–118. doi: 10.1111/j.1365-2761.2008.01011.x
- Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., and Blaxter, M. L. (2013). Special features of RAD Sequencing data: implications for genotyping. *Mol. Ecol.* 22, 3151–3164. doi: 10.1111/mec.12084

- Davidson, W., Koop, B., Jones, S., Iturra, P., Vidal, R., Maass, A., et al. (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol.* 11, 403. doi: 10.1186/gb-2010-11-9-403
- de Koning, D. J., Carlborg, O., and Haley, C. S. (2005). The genetic dissection of immune response using gene-expression studies and genome mapping. *Vet. Immunol. Immunopathol.* 105, 343–352. doi: 10.1016/j.vetimm.2005.02.007
- Dekkers, J., and Van der Werf, J. (2007). “Strategies, limitations and opportunities for marker-assisted selection in livestock,” in *Marker-Assisted Selection – Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*, eds E. P. Guimarães, J. Ruane, B. D. Scherf, A. Sonnino, and J. D. Dargie (Rome: FAO), 167–184.
- Doeschl-Wilson, A. B., Bishop, S. C., Kyriazakis, I., and Villanueva, B. (2012). Novel methods for quantifying individual host response to infectious pathogens for genetic analyses. *Front. Genet.* 3:266. doi: 10.3389/fgene.2012.00266
- Dorson, M., Quillet, E., Hollebecq, M., Torhy, C., and Chevassus, B. (1995). Selection of rainbow trout resistant to viral haemorrhagic septicaemia virus and transmission of resistance by gynogenesis. *Vet. Res.* 26, 361–368.
- Everett, M., Grau, E., and Seeb, J. (2011). Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Mol. Ecol. Resour.* 11, 93–108. doi: 10.1111/j.1755-0998.2010.02969.x
- Everett, M. V., Miller, M. R., and Seeb, J. E. (2012). Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. *BMC Genomics* 13:521. doi: 10.1186/1471-2164-13-521
- Ewart, K., Belanger, J., Williams, J., Karakach, T., Penny, S., Tsoi, S., et al. (2005). Identification of genes differentially expressed in Atlantic salmon (*S. salar*) in response to infection by *Aeromonas salmonicida* using cDNA microarray technology. *Dev. Comp. Immunol.* 29, 333–347. doi: 10.1016/j.dci.2004.08.004
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4 Edn. Essex: Longman.
- Fevolden, S. E., Nordmo, R., Refstie, T., and Roed, K. H. (1993). Disease resistance in Atlantic salmon (*Salmo salar*) selected for high or low responses to stress. *Aquaculture* 109, 215–224. doi: 10.1016/0044-8486(93)90164-T
- Food and Agriculture Organization of the United Nations [FAO]. (2012). *The State of World Fisheries and Aquaculture*. Italy: FAO.
- Fryer, J., and Hedrick, R. (2003). *Piscirickettsia salmonis*: a Gram-negative intracellular bacterial pathogen of fish. *J. Fish Dis.* 26, 251–262. doi: 10.1046/j.1365-2761.2003.00460.x
- Garrigan, D., and Hedrick, P. W. (2001). Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics* 53, 483–489. doi: 10.1007/s002510100352
- Gharbi, K., Gautier, A., Danzmann, R. G., Gharbi, S., Sakamoto, T., Hoyheim, B., et al. (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics* 172, 2405–2419. doi: 10.1534/genetics.105.048330
- Gilbey, J., Verspoor, E., Mclay, A., and Houlihan, D. (2004). A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Anim. Genet.* 35, 98–105. doi: 10.1111/j.1365-2052.2004.01091.x
- Gjedrem, T. (2000). Genetic improvement of cold-water fish species. *Aquac. Res.* 31, 25–33. doi: 10.1046/j.1365-2109.2000.00389.x
- Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture* 344, 12–22. doi: 10.1016/j.aquaculture.2012.03.003
- Gjedrem, T., and Gjøen, H. (1995). Genetic variation in susceptibility of Atlantic salmon, *Salmo salar* L., to furunculosis, BKD and cold water vibriosis. *Aquac. Res.* 26, 129–134. doi: 10.1111/j.1365-2109.1995.tb00892.x
- Gjedrem, T., Salte, R., and Gjøen, H. M. (1991). Genetic variation in susceptibility of Atlantic salmon to furunculosis. *Aquaculture* 97, 1–6. doi: 10.1016/0044-8486(91)90274-B
- Gjerde, B., Evensen, Ø., Bentsen, H. B., and Storset, A. (2009). Genetic (co) variation of vaccine injuries and innate resistance to furunculosis (*Aeromonas salmonicida*) and infectious salmon anaemia (ISA) in Atlantic salmon (*Salmo salar*). *Aquaculture* 287, 52–58. doi: 10.1016/j.aquaculture.2008.10.028
- Gjøen, H., and Bentsen, H. (1997). Past, present, and future of genetic improvement in salmon aquaculture. *ICES J. Marine Sci.* 54, 1009–1014. doi: 10.1016/S1054-3139(97)80005-7
- Gjøen, H. M., Refstie, T., Ulla, O., and Gjerde, B. (1997). Genetic correlations between survival of Atlantic salmon in challenge and field tests. *Aquaculture* 158, 277–288. doi: 10.1016/S0044-8486(97)00203-2
- Glover, K., Aasmundstad, T., Nilsen, F., Storset, A., and Skaala, Ø. (2005). Variation of Atlantic salmon families (*Salmo salar* L.) in susceptibility to the sea lice *Lepeophtheirus salmonis* and *Caligus elongatus*. *Aquaculture* 245, 19–30. doi: 10.1016/j.aquaculture.2004.11.047
- Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575
- Goddard, M. E., and Meuwissen, T. H. E. (2005). The use of linkage disequilibrium to map quantitative trait loci. *Aust. J. Exp. Agric.* 45, 837–845.
- Gonen, S., Lowe, N. R., Cezard, T., Gharbi, K., Bishop, S. C., and Houston, R. D. (2014). Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* 15:166. doi: 10.1186/1471-2164-15-166
- Grimholt, U., Drablos, F., Jørgensen, S. M., Hoyheim, B., and Stet, R. J. (2002). The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics* 54, 570–581. doi: 10.1007/s00251-002-0499-8
- Grimholt, U., Getahun, A., Hermsen, T., and Stet, R. J. (2000). The major histocompatibility class II alpha chain in salmonid fishes. *Dev. Comp. Immunol.* 24, 751–763. doi: 10.1016/S0145-305X(00)00034-3
- Grimholt, U., Hordvik, I., Fosse, V. M., Olsaker, I., Endresen, C., and Lie, Ø. (1993). Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics* 37, 469–473. doi: 10.1007/BF00222473
- Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjøeplum, S., Storset, A., et al. (2003). MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55, 210–219. doi: 10.1007/s00251-003-0567-8
- Grimholt, U., Olsaker, I., Lindstrøm, C., and Lie, Ø. (1994). A study of variability in the MHC class II β 1 and class I α 2 domain exons of Atlantic salmon, *Salmo salar* L. *Anim. Genet.* 25, 147–153. doi: 10.1111/j.1365-2052.1994.tb00103.x
- Guy, D. R., Bishop, S. C., Brotherton, S., Hamilton, A., Roberts, R. J., McAndrew, B., et al. (2006). Analysis of the incidence of infectious pancreatic necrosis mortality in pedigreed Atlantic salmon, *Salmo salar* L., populations. *J. Fish Dis.* 29, 637–647. doi: 10.1111/j.1365-2761.2006.00758.x
- Guyomard, R., Boussaha, M., Krieg, F., Hervet, C., and Quillet, E. (2012). A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC Genet.* 13:15. doi: 10.1186/1471-2156-13-15
- Guyomard, R., Mauger, S., Tabet-Canale, K., Martineau, S., Genet, C., Krieg, F., et al. (2006). A type I and type II microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) with presumptive coverage of all chromosome arms. *BMC Genomics* 7:302. doi: 10.1186/1471-2164-7-302
- Hansen, J. D., Strassburger, P., and Du Pasquier, L. (1996). Conservation of an alpha 2 domain within the teleostean world, mhc class i from the rainbow trout *Oncorhynchus mykiss*. *Dev. Comp. Immunol.* 20, 417–425. doi: 10.1016/S0145-305X(96)00030-4
- Hansen, J. D., Strassburger, P., Thorgaard, G. H., Young, W. P., and Du Pasquier, L. (1999). Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. *J. Immunol.* 163, 774–786.
- Hayes, B., and Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Select. Evol.* 33, 209–230. doi: 10.1186/1297-9686-33-3-209
- Hayes, B., Laerdahl, J., Lien, S., Moen, T., Berg, P., Hindar, K., et al. (2007). An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. *Aquaculture* 265, 82–90. doi: 10.1016/j.aquaculture.2007.01.037
- Hayes, B. J., Gjuvsland, A., and Omholt, S. (2006). Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. *Heredity (Edinb.)* 97, 19–26. doi: 10.1038/sj.hdy.6800827
- Henryon, M., Berg, P., Olesen, N. J., Kjær, T. E., Slierendrecht, W. J., Jokumsen, A., et al. (2005). Selective breeding provides an approach to increase resistance of rainbow trout (*Oncorhynchus mykiss*) to the diseases, enteric redmouth disease, rainbow trout fry syndrome, and viral haemorrhagic septicaemia. *Aquaculture* 250, 621–636. doi: 10.1016/j.aquaculture.2004.12.022

- Henryon, M., Jokumsen, A., Berg, P., Lund, I., Pedersen, P. B., Olesen, N. J., et al. (2002). Genetic variation for growth rate, feed conversion efficiency, and disease resistance exists within a farmed population of rainbow trout. *Aquaculture* 209, 59–76. doi: 10.1016/S0044-8486(01)00729-3
- Hiendleder, S., Bauersachs, S., Boulesteix, A., Blum, H., Arnold, G., Frohlich, T., et al. (2005). Functional genomics: tools for improving farm animal health and welfare. *Rev. Sci. Tech.* 24, 355. doi: 10.1186/1476-4598-13-88
- Hill, A. V. (1999). Genetics and genomics of infectious disease susceptibility. *Br. Med. Bull.* 55, 401–213. doi: 10.1258/0007142991902457
- Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., and Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol. Ecol. Resour.* 11(Suppl. 1), 117–122. doi: 10.1111/j.1755-0998.2010.02967.x
- Hollebecq, M.-G., Faivre, B., Bourmaud, C., and Michel, C. (1995). Spontaneous bactericidal and complement activities in serum of rainbow trout (*Oncorhynchus mykiss*) genetically selected for resistance or susceptibility to furunculosis. *Fish Shellfish Immunol.* 5, 407–426. doi: 10.1006/fsim.1995.0040
- Hordvik, I., Grimholt, U., Fosse, V. M., Lie, Ø., and Endresen, C. (1993). Cloning and sequence analysis of cDNAs encoding the MHC class II β chain in Atlantic salmon (*Salmo salar*). *Immunogenetics* 37, 437–441. doi: 10.1007/BF00222467
- Houston, R. D., Davey, J. W., Bishop, S. C., Lowe, N. R., Mota-Velasco, J. C., Hamilton, A., et al. (2012). Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* 13:244. doi: 10.1186/1471-2164-13-244
- Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Mota-Velasco, J. C., Gheyas, A. A., et al. (2010). The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity (Edinb.)* 105, 318–327. doi: 10.1038/hdy.2009.171
- Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008a). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178, 1109–1115. doi: 10.1534/genetics.107.082974
- Houston, R. D., Gheyas, A., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008b). Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (*Salmo salar*). *Dev. Biol. (Basel)* 132, 199–204. doi: 10.1159/000317160
- Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., Downing, A., et al. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15:90. doi: 10.1186/1471-2164-15-90
- Johnson, N. A., Vallejo, R. L., Silverstein, J. T., Welch, T. J., Wiens, G. D., Hallerman, E. M., et al. (2008). Suggestive association of major histocompatibility IB genetic markers with resistance to bacterial cold water disease in rainbow trout (*Oncorhynchus mykiss*). *Mar. Biotechnol.* 10, 429–437. doi: 10.1007/s10126-007-9080-7
- Jones, P. G., Hammell, K. L., Gettinby, G., and Revie, C. W. (2013). Detection of emamectin benzoate tolerance emergence in different life stages of sea lice, *Lepeophtheirus salmonis*, on farmed Atlantic salmon, *Salmo salar* L. *J. Fish Dis.* 36, 209–220. doi: 10.1111/jfd.12022
- Kjøglum, S., Grimholt, U., and Larsen, S. (2005). Non-MHC genetic and tank effects influence disease challenge tests in Atlantic salmon (*Salmo salar*). *Aquaculture* 250, 102–109. doi: 10.1016/j.aquaculture.2005.03.021
- Kjøglum, S., Henryon, M., Aasmundstad, T., and Korsgaard, I. (2008). Selective breeding can increase resistance of Atlantic salmon to furunculosis, infectious salmon anaemia and infectious pancreatic necrosis. *Aquac. Res.* 39, 498–505. doi: 10.1111/j.1365-2109.2008.01904.x
- Kjøglum, S., Larsen, S., Bakke, H. G., and Grimholt, U. (2006). How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). *Fish Shellfish Immunol.* 21, 431–441. doi: 10.1016/j.fsi.2006.02.001
- Kolstad, K., Heuch, P. A., Gjerde, B., Gjedrem, T., and Salte, R. (2005). Genetic variation in resistance of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*. *Aquaculture* 247, 145–151. doi: 10.1016/j.aquaculture.2005.02.009
- Langefors, Å., Lohm, J., Grahm, M., Andersen, Ø., and Von Schantz, T. (2001). Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. Biol. Sci.* 268, 479–485. doi: 10.1098/rspb.2000.1378
- Langevin, C., Blanco, M., Martin, S. A. M., Jouneau, L., Bernardet, J.-F., Houel, A., et al. (2012). Transcriptional responses of resistant and susceptible fish clones to the bacterial pathogen *Flavobacterium psychrophilum*. *PLoS ONE* 7:e39126. doi: 10.1371/journal.pone.0039126
- Leeds, T. D., Silverstein, J. T., Weber, G. M., Vallejo, R. L., Palti, Y., Rexroad, C. E., et al. (2010). Response to selection for bacterial cold water disease resistance in rainbow trout. *J. Anim. Sci.* 88, 1936–1946. doi: 10.2527/jas.2009-2538
- Lhorente, J. P., Gallardo, J. A., Villanueva, B., Araya, A. M., Torrealba, D. A., Toledo, X. E., et al. (2012). Quantitative genetic basis for resistance to *Caligus rogercresseyi* sea lice in a breeding population of Atlantic salmon (*Salmo salar*). *Aquaculture* 324, 55–59. doi: 10.1016/j.aquaculture.2011.10.046
- Lien, S., Gidskehaug, L., Moen, T., Hayes, B. J., Berg, P. R., Davidson, W. S., et al. (2011). A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics* 12:615. doi: 10.1186/1471-2164-12-615
- Liu, Z., and Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37. doi: 10.1016/j.aquaculture.2004.05.027
- Lohm, J., Grahm, M., Langefors, Å., Andersen, Ø., Storset, A., and Von Schantz, T. (2002). Experimental evidence for major histocompatibility complex–allele-specific resistance to a bacterial infection. *Proc. Biol. Sci.* 269, 2029–2033. doi: 10.1098/rspb.2002.2114
- Lund, T., Gjedrem, T., Bentsen, H., Eide, D., Larsen, H., and Røed, K. (1995). Genetic variation in immune parameters and associations to survival in Atlantic salmon. *J. Fish Biol.* 46, 748–758. doi: 10.1111/j.1095-8649.1995.tb01598.x
- Martin, S. A., Blaney, S. C., Houlihan, D. F., and Secombes, C. J. (2006). Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*). *Mol. Immunol.* 43, 1900–1911. doi: 10.1016/j.molimm.2005.10.007
- McClelland, E. K., and Naish, K. A. (2008). A genetic linkage map for coho salmon (*Oncorhynchus kisutch*). *Anim. Genet.* 39, 169–179. doi: 10.1111/j.1365-2052.2008.01699.x
- Meuwissen, T. H. E., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genomewide dense marker maps. *Genetics* 157, 1819–1829.
- Meuwissen, T. H., Karlsen, A., Lien, S., Olsaker, I., and Goddard, M. E. (2002). Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161, 373–379.
- Miller, K. M., Winton, J. R., Schulze, A. D., Purcell, M. K., and Ming, T. J. (2004). Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. *Environ. Biol. Fishes* 69, 307–316. doi: 10.1023/B:EBF1.0000022874.48341.0f
- Miller, K. M., and Withler, R. E. (1996). Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics* 43, 337–351. doi: 10.1007/BF02199802
- Mitchell, S., and Rodger, H. (2011). A review of infectious gill disease in marine salmonid fish. *J. Fish Dis.* 34, 411–432. doi: 10.1111/j.1365-2761.2011.01251.x
- Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368
- Moen, T., Fjalestad, K. T., Munck, H., and Gomez-Raya, L. (2004a). A multistage testing strategy for detection of quantitative trait Loci affecting disease resistance in Atlantic salmon. *Genetics* 167, 851–858. doi: 10.1534/genetics.103.013227
- Moen, T., Hoyheim, B., Munck, H., and Gomez-Raya, L. (2004b). A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Anim. Genet.* 35, 81–92. doi: 10.1111/j.1365-2052.2004.01097.x
- Moen, T., Sonesson, A. K., Hayes, B., Lien, S., Munck, H., and Meuwissen, T. H. (2007). Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genet.* 8:53. doi: 10.1186/1471-2156-8-53
- Mustafa, A., and MacKinnon, B. M. (1999). Genetic variation in susceptibility of Atlantic salmon to the sea louse *Caligus elongatus* Nordmann, 1832. *Can. J. Zool.* 77, 1332–1335. doi: 10.1139/cjz-77-8-1332
- Nichols, K. M., Young, W. P., Danzmann, R. G., Robison, B. D., Rexroad, C., Noakes, M., et al. (2003). A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet.* 34, 102–115. doi: 10.1046/j.1365-2052.2003.00957.x

- Norris, A., Foyle, L., and Ratcliff, J. (2008). Heritability of mortality in response to a natural pancreas disease (SPDV) challenge in Atlantic salmon, *Salmo salar* L., post-smolts on a West of Ireland sea site. *J. Fish Dis.* 31, 913–920. doi: 10.1111/j.1365-2761.2008.00982.x
- Ødegård, J., Baranski, M., Gjerde, B., and Gjedrem, T. (2011). Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquac. Res.* 42, 103–114. doi: 10.1111/j.1365-2109.2010.02669.x
- Ødegård, J., Olesen, I., Gjerde, B., and Klemetsdal, G. (2006). Evaluation of statistical models for genetic analysis of challenge test data on furunculosis resistance in Atlantic salmon (*Salmo salar*): prediction of field survival. *Aquaculture* 259, 116–123. doi: 10.1016/j.aquaculture.2006.05.034
- Ødegård, J., Olesen, I., Gjerde, B., and Klemetsdal, G. (2007a). Evaluation of statistical models for genetic analysis of challenge-test data on ISA resistance in Atlantic salmon (*Salmo salar*): prediction of progeny survival. *Aquaculture* 266, 70–76. doi: 10.1016/j.aquaculture.2007.02.012
- Ødegård, J., Olesen, I., Gjerde, B., and Klemetsdal, G. (2007b). Positive genetic correlation between resistance to bacterial (furunculosis) and viral (infectious salmon anaemia) diseases in farmed Atlantic salmon (*Salmo salar*). *Aquaculture* 271, 173–177. doi: 10.1016/j.aquaculture.2007.06.006
- Olesen, I., Hung, D., and Ødegård, J. (2007). Genetic analysis of survival in challenge tests of furunculosis and ISA in Atlantic salmon. Genetic parameter estimates and model comparisons. *Aquaculture* 272(Suppl. 1), S297–S298. doi: 10.1016/j.aquaculture.2007.07.155
- Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M. R., Akutsu, T., et al. (2001). Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genomics* 265, 23–31. doi: 10.1007/s004380000392
- Palti, Y., Gao, G., Miller, M. R., Vallejo, R. L., Wheeler, P. A., Quillet, E., et al. (2014a). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Mol. Ecol. Res.* 14, 588–596. doi: 10.1111/1755-0998.12204
- Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., et al. (2014b). The development and characterization of a 57K SNP array for rainbow trout. *Mol. Ecol. Res.* doi: 10.1111/1755-0998.12337 [Epub ahead of print].
- Palti, Y., Genet, C., Gao, G., Hu, Y., You, F., Boussaha, M., et al. (2012). A second generation integrated map of the rainbow trout (*Oncorhynchus mykiss*) genome: analysis of conserved synteny with model fish genomes. *Mar. Biotechnol.* 14, 343–357. doi: 10.1007/s10126-011-9418-z
- Palti, Y., Genet, C., Luo, M.-C., Charlet, A., Gao, G., Hu, Y., et al. (2011). A first generation integrated map of the rainbow trout genome. *BMC Genomics* 12:180. doi: 10.1186/1471-2164-12-180
- Palti, Y., Nichols, K. M., Waller, K. I., Parsons, J. E., and Thorgaard, G. H. (2001). Association between DNA polymorphisms tightly linked to MHC class II genes and IHN virus resistance in backcrosses of rainbow and cutthroat trout. *Aquaculture* 194, 283–289. doi: 10.1016/S0044-8486(00)00526-3
- Palti, Y., Parsons, J. E., and Thorgaard, G. H. (1999). Identification of candidate DNA markers associated with IHN virus resistance in backcrosses of rainbow (*Oncorhynchus mykiss*) and cutthroat trout (*O. clarki*). *Aquaculture* 173, 81–94. doi: 10.1016/S0044-8486(98)00471-2
- Perry, G. M. L., Tarte, P., Croisetière, S., Belhumeur, P., and Bernatchez, L. (2004). Genetic variance and covariance for 0+ brook charr (*Salvelinus fontinalis*) weight and survival time of furunculosis (*Aeromonas salmonicida*) exposure. *Aquaculture* 235, 263–271. doi: 10.1016/j.aquaculture.2004.03.002
- Phillips, R., Keatley, K., Morasch, M., Ventura, A., Lubieniecki, K., Koop, B., et al. (2009). Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* 10:46. doi: 10.1186/1471-2156-10-46
- Piepho, H.-P., and Koch, G. (2000). Codominant analysis of banding data from a dominant marker system by normal mixtures. *Genetics* 155, 1459–1468.
- Pomp, D., Allan, M. F., and Wesolowski, S. R. (2004). Quantitative genomics: exploring the genetic architecture of complex trait predisposition. *J. Anim. Sci.* 82, E300–E312.
- Porto-Neto, L., Kijas, J., and Reverter, A. (2014). The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Select. Evol.* 46, 22. doi: 10.1186/1297-9686-46-22
- Purcell, M. K., Nichols, K. M., Winton, J. R., Kurath, G., Thorgaard, G. H., Wheeler, P., et al. (2006). Comprehensive gene expression profiling following DNA vaccination of rainbow trout against infectious hematopoietic necrosis virus. *Mol. Immunol.* 43, 2089–2106. doi: 10.1016/j.molimm.2005.12.005
- Qureshi, S. T., Skamene, E., and Malo, D. (1999). Comparative genomics and host resistance against infectious diseases. *Emerging Infect. Dis.* 5, 36–47. doi: 10.3201/eid0501.990105
- Råberg, L., Sim, D., and Read, A. F. (2007). Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science* 318, 812–814. doi: 10.1126/science.1148526
- Rexroad, C. E. III, Palti, Y., Gahr, S. A., and Vallejo, R. L. (2008). A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* 9:74. doi: 10.1186/1471-2156-9-74
- Rise, M. L., Jones, S. R. M., Brown, G. D., Von Schalburg, K. R., Davidson, W. S., and Koop, B. F. (2004a). Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiol. Genomics* 15, 21–35. doi: 10.1152/physiolgenomics.00036.2004
- Rise, M. L., Von Schalburg, K. R., Brown, G. D., Mawer, M. A., Devlin, R. H., Kuipers, N., et al. (2004b). Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res.* 14, 478–490. doi: 10.1101/gr.1687304
- Roberts, R. J., and Pearson, M. D. (2005). Infectious pancreatic necrosis in Atlantic salmon, *Salmo salar* L. *J. Fish Dis.* 28, 383–390. doi: 10.1111/j.1365-2761.2005.00642.x
- Rodriguez, M. F., Lapatra, S., Williams, S., Famula, T., and May, B. (2004). Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* 241, 93–115. doi: 10.1016/j.aquaculture.2004.08.003
- Røed, K. H., Fjalestad, K. T., and Strømshim, A. (1993). Genetic variation in lysozyme activity and spontaneous haemolytic activity in Atlantic salmon (*Salmo salar*). *Aquaculture* 114, 19–31. doi: 10.1016/0044-8486(93)90247-V
- Rothschild, M. F., and Soller, M. (1997). Candidate gene analysis to detect traits of economic importance in domestic livestock. *Probe* 8, 13–22.
- Rozas, M., and Enriquez, R. (2014). Piscirickettsiosis and *Piscirickettsia salmonis* in fish: a review. *J. Fish Dis.* 37, 163–188. doi: 10.1111/jfd.12211
- Ruane, N. M., and Jones, S. R. M. (2013). “Amoebic Gill Disease (AGD) of farmed Atlantic salmon (*Salmo salar* L.),” in *Proceedings of the ICES Identification Leaflets for Diseases and Parasites of Fish and Shellfish*. Leaflet No. 60 (Copenhagen: International Council for the Exploration), 6.
- Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., et al. (2000). A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155, 1331–1345.
- Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264
- Salte, R., Bentsen, H. B., Moen, T., Tripathy, S., Bakke, T. A., Ødegård, J., et al. (2009). Prospects for a genetic management strategy to control *Gyrodactylus salaris* infection in wild Atlantic salmon (*Salmo salar*) stocks. *Can. J. Fish. Aquat. Sci.* 67, 121–129. doi: 10.1139/F09-168
- Salte, R., Gjøn, H. M., Norberg, K., and Gjedrem, T. (1993). Plasma protein levels as potential marker traits for resistance to furunculosis. *J. Fish Dis.* 16, 561–568. doi: 10.1111/j.1365-2761.1993.tb00892.x
- Sanchez, C., Smith, T., Wiedmann, R., Vallejo, R., Salem, M., Yao, J., et al. (2009). Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* 10:559. doi: 10.1186/1471-2164-10-559
- Silverstein, J. T., Vallejo, R. L., Palti, Y., Leeds, T. D., Rexroad, C. E., Welch, T. J., et al. (2009). Rainbow trout resistance to bacterial cold-water disease is moderately heritable and is not adversely correlated with growth. *J. Anim. Sci.* 87, 860–867. doi: 10.2527/jas.2008-1157
- Shum, B., Mason, P., Magor, K., Flodin, L., Stet, R., and Parham, P. (2002). Structures of two major histocompatibility complex class I genes of the rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics* 54, 193–199. doi: 10.1007/s00251-002-0450-z
- Shum, B. P., Rajalingam, R., Magor, K. E., Azumi, K., Carr, W. H., Dixon, B., et al. (1999). A divergent non-classical class I gene conserved in salmonids. *Immunogenetics* 49, 479–490. doi: 10.1007/s002510050524

- Sonesson, A., and Meuwissen, T. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Select. Evol.* 41, 37. doi: 10.1186/1297-9686-41-37
- Stear, M. J., Bishop, S. C., Mallard, B. A., and Raadsma, H. (2001). The sustainability, feasibility and desirability of breeding livestock for disease resistance. *Res. Vet. Sci.* 71, 1–7. doi: 10.1053/rvsc.2001.0496
- Stet, R., De Vries, B., Mudde, K., Hermesen, T., Van Heerwaarden, J., Shum, B., et al. (2002). Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics* 54, 320–331. doi: 10.1007/s00251-002-0477-1
- Sutherland, B., Koczka, K., Yasuike, M., Jantzen, S., Yazawa, R., Koop, B., et al. (2014). Comparative transcriptomics of Atlantic *Salmo salar*, chum *Oncorhynchus keta* and pink salmon *O. gorbuscha* during infections with salmon lice *Lepeophtheirus salmonis*. *BMC Genomics* 15:200. doi: 10.1186/1471-2164-15-200
- Taylor, J. F. (2014). Implementation and accuracy of genomic selection. *Aquaculture* 420–421 (Suppl. 1), S8–S14. doi: 10.1016/j.aquaculture.2013.02.017
- Taylor, R. S., Kube, P. D., Muller, W. J., and Elliott, N. G. (2009). Genetic variation of gross gill pathology and survival of Atlantic salmon (*Salmo salar* L.) during natural amoebic gill disease challenge. *Aquaculture* 294, 172–179. doi: 10.1016/j.aquaculture.2009.06.007
- Thorgaard, G. H., Bailey, G. S., Williams, D., Buhler, D. R., Kaattari, S. L., Ristow, S. S., et al. (2002). Status and opportunities for genomics research with rainbow trout. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 133, 609–646. doi: 10.1016/S1096-4959(02)00167-7
- Tsoi, S. M., Cale, J., Bird, I., Ewart, V., Brown, L., and Douglas, S. (2003). Use of human cDNA microarrays for identification of differentially expressed genes in Atlantic salmon liver during *Aeromonas salmonicida* infection. *Mar. Biotechnol.* 5, 545–554. doi: 10.1007/s10126-002-0112-z
- Tsoi, S. C. M., Ewart, K. V., Penny, S., Melville, K., Liebscher, R. S., Brown, L. L., et al. (2004). Identification of immune-relevant genes from Atlantic salmon using suppression subtractive hybridization. *Mar. Biotechnol.* 6, 199–214. doi: 10.1007/s10126-002-0101-2
- Vallejo, R., Palti, Y., Liu, S., Evenhuis, J., Gao, G., Rexroad, C. III, et al. (2014a). Detection of QTL in rainbow trout affecting survival when challenged with *Flavobacterium psychrophilum*. *Mar. Biotechnol.* 16, 349–360. doi: 10.1007/s10126-013-9553-9
- Vallejo, R. L., Palti, Y., Liu, S., Marancik, D. P., and Wiens, G. D. (2014b). Validation of linked QTL for bacterial cold water disease resistance and spleen size on rainbow trout chromosome Omy19. *Aquaculture* 432, 139–143. doi: 10.1016/j.aquaculture.2014.05.003
- Vallejo, R. L., Wiens, G. D., Rexroad, C. E. III, Welch, T. J., Evenhuis, J. P., Leeds, T. D., et al. (2010). Evidence of major genes affecting resistance to bacterial cold water disease in rainbow trout using Bayesian methods of segregation analysis. *J. Anim. Sci.* 88, 3814–3832. doi: 10.2527/jas.2010-2951
- Verrier, E. R., Dorson, M., Mauger, S., Torhy, C., Ciobotaru, C., Hervet, C., et al. (2013a). Resistance to a Rhabdovirus (VHSV) in rainbow trout: identification of a major QTL related to innate mechanisms. *PLoS ONE* 8:e55302. doi: 10.1371/journal.pone.0055302
- Verrier, E. R., Ehanno, A., Biacchesi, S., Le Guillou, S., Dechamp, N., Boudinot, P., et al. (2013b). Lack of correlation between the resistances to two rhabdovirus infections in rainbow trout. *Fish Shellfish Immunol.* 35, 9–17. doi: 10.1016/j.fsi.2013.03.369
- Villanueva, B., Fernández, J., García-Cortés, L. A., Varona, L., Daetwyler, H. D., and Toro, M. A. (2011). Accuracy of genome-wide evaluation for disease resistance in aquaculture breeding programs. *J. Anim. Sci.* 89, 3433–3442. doi: 10.2527/jas.2010-3814
- von Schalburg, K., Rise, M., Cooper, G., Brown, G., Gibbs, A. R., Nelson, C., et al. (2005). Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6:126. doi: 10.1186/1471-2164-6-126
- Walsh, B., and Henderson, D. (2004). Microarrays and beyond: what potential do current and future genomics tools have for breeders? *J. Animal Sci.* 82, E292–E299.
- Weber, G. M., Vallejo, R. L., Lankford, S. E., Silverstein, J. T., and Welch, T. J. (2008). Cortisol response to a crowding stress: heritability and association with disease resistance to *Yersinia ruckeri* in rainbow trout. *N. Am. J. Aquac.* 70, 425–433. doi: 10.1577/A07-059.1
- Weber, G. M., Wiens, G. D., Welch, T. J., Hostuttler, M. A., and Leeds, T. D. (2013). Comparison of disease resistance between diploid, induced-triploid, and intercross-triploid rainbow trout including trout selected for resistance to *Flavobacterium psychrophilum*. *Aquaculture* 410–411, 66–71. doi: 10.1016/j.aquaculture.2013.06.014
- Wetsten, M., Aasmundstad, T., Kjøglum, S., and Storset, A. (2007). Genetic analysis of resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar* L.). *Aquaculture* 272, 111–117. doi: 10.1016/j.aquaculture.2007.08.046
- Wiens, G. D., Lapatra, S. E., Welch, T. J., Evenhuis, J. P., Rexroad, C. E. III, and Leeds, T. D. (2013a). On-farm performance of rainbow trout (*Oncorhynchus mykiss*) selectively bred for resistance to bacterial cold water disease: effect of rearing environment on survival phenotype. *Aquaculture* 388–391, 128–136. doi: 10.1016/j.aquaculture.2013.01.018
- Wiens, G. D., Vallejo, R. L., Leeds, T. D., Palti, Y., Hadidi, S., Liu, S., et al. (2013b). Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout: identification of QTL on chromosome Omy19. *PLoS ONE* 8: e75749. doi: 10.1371/journal.pone.0075749
- Wientjes, Y. C. J., Veerkamp, R. F., and Calus, M. P. L. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193, 621–631. doi: 10.1534/genetics.112.146290
- Withler, R. E., and Evelyn, T. P. T. (1990). Genetic variation in resistance to bacterial kidney disease within and between two strains of coho salmon from British Columbia. *Trans. Am. Fish. Soc.* 119, 1003–1009. doi: 10.1577/1548-8659(1990)119<1003:GVIRTB>2.3.CO;2
- Woram, R. A., McGowan, C., Stout, J. A., Gharbi, K., Ferguson, M. M., Hoyheim, B., et al. (2004). A genetic linkage map for Arctic charr (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome* 47, 304–315. doi: 10.1139/g03-127
- Yáñez, J. M., Bangera, R., Lhorente, J. P., Oyarzún, M., and Neira, R. (2013). Quantitative genetic variation of resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *Aquaculture* 414–415, 155–159. doi: 10.1016/j.aquaculture.2013.08.009
- Yáñez, J. M., Lhorente, J. P., Bassini, L. N., Oyarzún, M., Neira, R., and Newman, S. (2014a). Genetic co-variation between resistance against both *Caligus rogercresseyi* and *Piscirickettsia salmonis*, and body weight in Atlantic salmon (*Salmo salar*). *Aquaculture* 433, 295–298. doi: 10.1016/j.aquaculture.2014.06.026
- Yáñez, J. M., Naswa, S., López, M. E., Bassini, L., Cabrejos, M. E., Gilbey, J., et al. (2014b). “Development of a 200K SNP array for Atlantic Salmon: exploiting across continents genetic variation,” in *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*, Vancouver.
- Yáñez, J. M., and Martínez, V. (2010). Factores genéticos que inciden en la resistencia a enfermedades infecciosas en salmónidos y su aplicación en programas de mejoramiento. *Arch. Med. Vet.* 42, 1–13.
- Young, W. P., Wheeler, P. A., Coryell, V. H., and Keim, P. (1998). A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics* 148, 839–850.
- Zhang, Z., Niu, C., Storset, A., Bøgvold, J., and Dalmo, R. A. (2011). Comparison of *Aeromonas salmonicida* resistant and susceptible salmon families: a high immune response is beneficial for the survival against *Aeromonas salmonicida* challenge. *Fish Shellfish Immunol.* 31, 1–9. doi: 10.1016/j.fsi.2010.12.019

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 September 2014; accepted: 06 November 2014; published online: 26 November 2014.

Citation: Yáñez JM, Houston RD and Newman S (2014) Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5:415. doi: 10.3389/fgene.2014.00415 This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Yáñez, Houston and Newman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Applications in the search for genomic selection signatures in fish

María E. López^{1,2*}, Roberto Neira¹ and José M. Yáñez^{2,3*}

¹ Faculty of Agricultural Sciences, University of Chile, Santiago, Chile

² Aquainnovo, Puerto Montt, Chile

³ Faculty of Veterinary and Animal Sciences, University of Chile, Santiago, Chile

Edited by:

Peng Xu, Chinese Academy of Fishery Sciences, China

Reviewed by:

Magadan Momo Susana, Institut National de la Recherche Agronomique Unité de Virologie et Immunologie Moleculaires, France
William S. Davidson, Simon Fraser University, Canada

*Correspondence:

María E. López and José M. Yáñez, Faculty of Agricultural Sciences and Faculty of Veterinary and Animal Sciences, University of Chile, Avenue Santa Rosa 11735, La Pintana, P. O. Box 8820808, Santiago, Chile
e-mail: me.lopez.dinamarca@gmail.com; jmayanez@uchile.cl

Selection signatures are genomic regions harboring DNA sequences functionally involved in the genetic variation of traits subject to selection. Selection signatures have been intensively studied in recent years because of their relevance to evolutionary biology and their potential association with genes that control phenotypes of interest in wild and domestic populations. Selection signature research in fish has been confined to a smaller scale, due in part to the relatively recent domestication of fish species and limited genomic resources such as molecular markers, genetic mapping, DNA sequences, and reference genomes. However, recent genomic technology advances are paving the way for more studies that may contribute to the knowledge of genomic regions underlying phenotypes of biological and productive interest in fish.

Keywords: fish, selection, domestication, single nucleotide polymorphisms, genome

INTRODUCTION

Selection signatures are genomic regions that harbor DNA sequences involved in genetic variation of traits subject to natural or artificial selection (Qanbari et al., 2012). Currently, due to advances in genomic technologies and statistical methods, such signatures can be identified in the genomes of various species.

Most studies in this field of genetics are based on the concept of *hitchhiking*, which suggests that selection affects the genome at a specific region, leaving “signatures” around the selected gene(s) (Smith and Haigh, 1974). Specifically, the hitchhiking theory focuses on the spread of new variants in a population due to selection for their favorable effects (Przeworski, 2002; Kim and Nielsen, 2004). Selection involving alleles from the population’s standing genetic variation produces specific and detectable DNA sequence patterns (Hermisson and Pennings, 2005).

The search for these molecular signatures has been the subject of intense research in recent years in both domesticated and wild populations of plants and animals, as well as in humans. These studies have been motivated by two main objectives: (1) a strong interest in the evolutionary past of the species and basic molecular mechanisms governing this evolution and (2) the expectation of an association between these genomic regions and biological functions or phenotypes of interest, since these regions should have some functional or adaptive importance underlying their selection (Nielsen et al., 2007). These studies are possible due to the development of various methods aimed at detecting selection at the molecular level in population samples. Information on allelic frequencies or haplotype patterns segregated in the population

can be used to identify signatures, since selection modifies the patterns of genetic variation expected under the neutral theory of molecular evolution.

Most studies in domesticated populations have focused on detecting relatively old selection signatures dating back hundreds or thousands of generations, e.g., (Flori et al., 2009), with few studies on genetic changes during early domestication stages (Trut et al., 2009).

Certain fish species provide unique models for studying the effects of selection and domestication, as their populations were domesticated recently and are available as both wild and domesticated populations simultaneously.

In this article we present different aspects involved in studying selection signatures at a genomic level in different species and discuss about the potential application of these studies in fish populations to unravel recent selection and domestication processes in these species.

IDENTIFICATION OF LOCI ASSOCIATED WITH TRAITS OF INTEREST

The search for genes controlling phenotypic variation can be performed in two different ways. First, the “top-down” approach which begins with knowledge of the phenotype of interest and uses genetic analysis to identify genes or causal regions. These approaches include candidate gene studies, identification of Quantitative trait loci (QTLs) and association mapping. These studies have certain limitations, including the need for an *a priori* hypothesis about which genes underlie the trait of interest, information about family relationship between individuals, as well as, access

to a large number of relatives with phenotypic records (Gu et al., 2009). Second, the “bottom-up” approach, in contrast, begins with genomic information and involves statistical evaluation of molecular information to identify regions subject to selection (Ross-Ibarra et al., 2007). This approach searches for patterns of linkage disequilibrium, genetic differentiation, or frequency spectrum that are inconsistent with the neutral evolution model to identify selection signatures (Qanbari et al., 2010). Recent advances in genomics provide a new paradigm for the “bottom-up” strategy concerning population genomics, a discipline that infers genetic and evolutionary parameters of a population based on datasets from the whole genome (Black et al., 2001).

In this context, population genomics relies on two basic principles or assumptions. First, neutral loci will be equally affected by demographic effects and by the evolutionary history of the population. Second, loci under selection will tend to behave distinctively, revealing atypical variation patterns (Luikart et al., 2003).

MODELS OF SELECTION

Natural selection can be defined as the differential contribution of genetic variation to future generations (Aquadro et al., 2001) due to differential reproduction of some phenotypes/genotypes over others under prevailing environmental conditions at a given time (Futuyma, 1998). It is the driving force behind Darwinian evolution and can be subdivided into different types, depending on the evolutionary outcome (Hurst, 2009).

Directional selection tends to decrease variation *within* a population but may increase or decrease variation *among* populations. Positive selection is a type of directional selection that favors alleles that increase fitness of individuals. When directional selection eliminates unfavorable mutations, it is called purifying selection (also known as negative selection).

Diversifying (or disruptive) selection favors variety and benefits individuals with extreme phenotypes over intermediate. In this type of selection, the propagation of an allele never reaches fixation, and therefore it may occur when an allele is initially subject to positive selection, and then negative selection when the frequency becomes too high (Nielsen, 2005).

Balanced selection, which helps to maintain an equilibrium point at which both alleles remain in the population, has several forms, including frequency-dependent selection and overdominance, which occurs when the heterozygote has the higher biological fitness, and therefore variability is maintained in the population (Nielsen, 2005).

SELECTION SIGNATURES

In the classic “hitchhiking” scenario, first described by Smith and Haigh (1974), a new allelic variant that represents a favorable adaptive substitution originates within the population as a new mutation, and its frequency increases as a result of constant selection pressure. When a favorable allele is selected, and its frequency increases to fixation in a population, genetic variation in the surrounding DNA segment is altered; that is, the increased frequency of the selected allele also produces increased frequency of closely-linked alleles (Pennings and Hermisson, 2006).

The ancestral variation, i.e., genetic variation present in a population prior to a selection process, is maintained only if

recombination during this phase disrupts the association between an adjacent locus and the selected site. The resulting pattern of such a selective event is a strong reduction in genetic variation around the selected site, known as a “*hard sweep*,” which corresponds to the classic selective sweep (Pritchard et al., 2010).

There is a second scenario in which an adaptive substitution involves multiple copies of a favorable allele in the population. This may occur for two reasons. First, when an adaptation arises from genetic variation, many copies of the favorable allele may be present in the population. Fixation of this allele may involve descendants of more than one of these copies. Second, a favorable allele can be introduced in the population by recurrent mutation or migration during a selection phase, and again, several descendants of independent origin may contribute to the allelic fixation. In both cases, different alleles of loci adjacent to any such favorable copies will be retained in the population, resulting in different haplotypes (Pennings and Hermisson, 2006).

Selection signatures involving descendants of more than one copy of the selected allele and, therefore with different haplotypes at closely-linked sites, are called “*soft sweeps*.” This type of selection signature results in different haplotype patterns than the “*hard sweeps*” described above and it is more difficult to detect as it only produces a slight reduction in the levels of adjacent polymorphisms (Cutter and Payseur, 2013).

Furthermore, when adaptation occurs by polygenic selection, it induces an increase in the allelic frequency of several loci which have a favorable effect on a particular phenotype; however, these polygenic alleles do not necessarily achieve fixation, and the resulting haplotype pattern corresponds to several partial selection signatures or multiple “*partial sweeps*” (Pritchard et al., 2010).

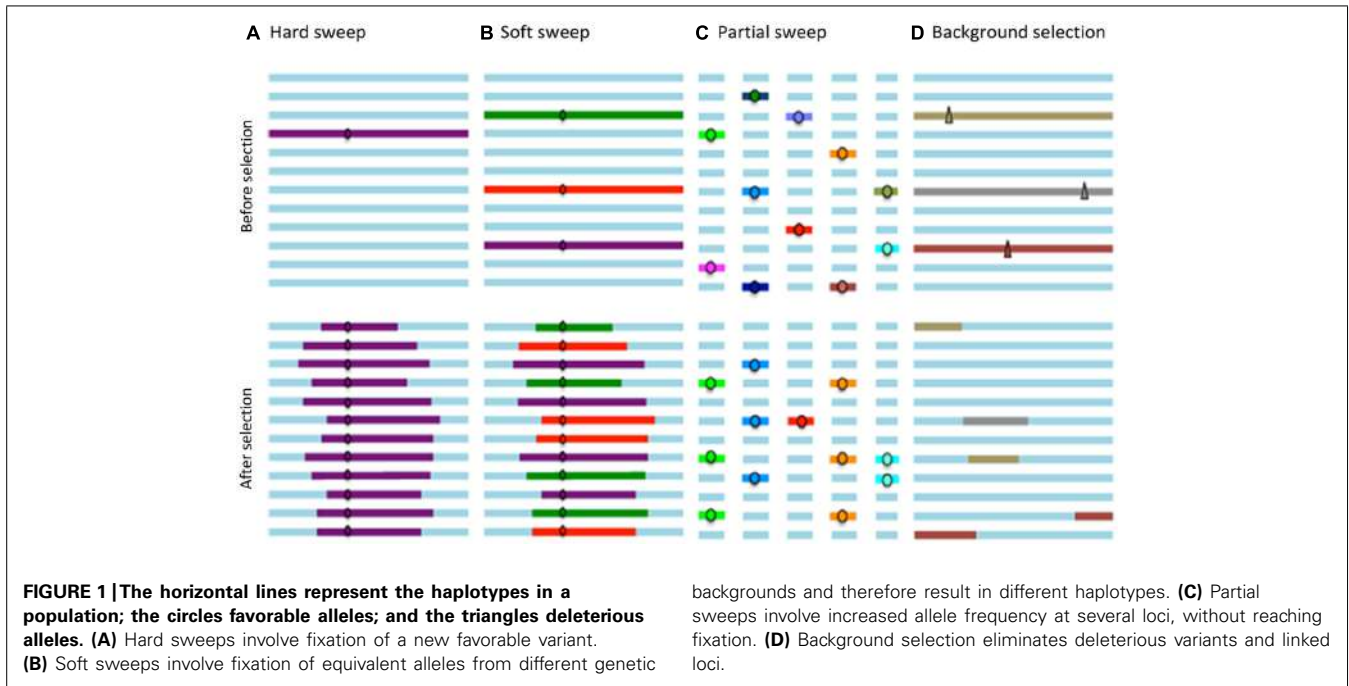
Finally, when purifying or negative selection reduces the frequency or eliminates a deleterious allele, the genetic diversity at linked loci also decreases, which is known as “*background selection*” (Charlesworth, 1993).

Figure 1 schematically summarizes the patterns caused by “*hard sweeps*,” “*soft sweeps*,” and “*partial sweeps*” that correspond to selective events for favorable variants in a population, as well as the pattern produced by “*background selection*.”

DOMESTICATION AND RECENT ARTIFICIAL SELECTION IN FISH

Domestication is the process by which various species have been adapted to a captive environment by humans. Such adaptation is accomplished through systematic breeding over generations and is characterized by changes in behavior, morphology, and physiology, as well as adaptive genetic changes caused by artificial and natural selection (Price, 1999).

In fish, domestication occurred very recently as compared to other land animals. One theory to explain the late domestication of aquatic species suggests that, due to the high fertility of these species, a small number of broodstock were required to obtain a sufficiently large progeny in subsequent generations. After a few generations, the inbreeding depression increases considerably; therefore, fitness and productive behavior decrease. As a result, fish farmers were forced to repeatedly take new broodstock from the wild environment, interrupting the continuity of domestication and breeding (Gjedrem et al., 2012). For this reason, aquaculture



has lagged behind land animal and plant culture in the use of breeding to enhance biological production efficiency.

On the other hand, it is estimated that less than 10% of aquaculture production is based on genetically improved stocks (Gjedrem, 2012), although the annual genetic gains reported for aquaculture species are substantially higher than for land animals. For example, selection response reported for grow-related traits are even higher than 10% in fish populations, which can substantially enhance aquaculture production by selective breeding (Gjedrem et al., 2012). Recent high selective pressures in farmed fish populations may have shaped genome variation in regions harboring causative mutations of selected traits. The identification of these regions may help in the understanding of the effect of selection events and identification of genetic variants involved in phenotypic variation in fish populations.

EXAMPLES OF DOMESTICATION AND BREEDING PROGRAMS IN AQUATIC SPECIES

The fish belonging to the family *Ciprinidae* are most likely the first fish species to be domesticated. For instance, the goldfish (*Carrassius auratus*) is an ornamental fish, believed to have been domesticated in China before the XVI century and later taken to Japan and Europe (Purdom, 1993). Another important group of ornamental fish is the koi carp, a variety derived from common carp (*Cyprinus carpio*) and mainly cultivated in Japan. The large variety of colors and forms among koi carp resulted from directed selection and crossbreeding (Gjedrem, 2005).

There is evidence in common carp showing a large response to selection for furunculosis survival rates (Schaperclaus, 1962). In 1987, Ilyassov (1987) showed results from 4 to 5 generations of selection in this species for resistance to dropsy disease, which increased survival by 30–40% as compared to unselected carps.

On the other hand, selection responses reported for growth rate in rohu carp (*Labeo rohita*) have been particularly high, reaching almost a 30% per generation (Gjedrem, 2012).

Furthermore, salmonid species are the most intensively selected fish populations. In this regard, the rainbow trout (*Oncorhynchus mykiss*) has a long history of domestication and breeding in the United States, Norway, Finland, and Denmark (McAndrew and Napier, 2011). In 1932, investigators began to select individuals to improve growth rate, number of eggs, and characteristics of sexual maturity (Donaldson and Olson, 1957). Currently there are 13 breeding programs worldwide aimed at improving growth rate, age at sexual maturity, fillet quality, and disease resistance in this species (Rye et al., 2010).

In the case of Atlantic salmon (*Salmo salar*), breeding programs exist in Norway, Scotland, Ireland, Australia, and Chile (Norris et al., 1999; Metcalfe et al., 2003; Glover et al., 2009; Dominik et al., 2010; Rye et al., 2010; McAndrew and Napier, 2011). Several traits of commercial interest such as growth, sexual maturity, meat quality, and disease resistance have been incorporated into breeding objectives. Furthermore, findings from genomic technologies have been incorporated into these breeding programs, for example, the use of QTLs to assist selection for resistance against the viral disease named infectious pancreatic necrosis (Houston et al., 2008; Moen et al., 2009).

Among Pacific salmon, the chinook salmon (*Oncorhynchus tshawytscha*) originating in British Columbia (BC), Canada was one the first species of salmon to be domesticated (Kim et al., 2004). Currently, its farming is limited and there are two breeding programs in operations (Rye et al., 2010). Moreover, genetic improvement programs for coho salmon (*Oncorhynchus kisutch*) have been successful in selecting for harvest weight and early spawning, with selection responses of about 10% per generation (Neira et al., 2006).

Tilapias are the second-most important group of cultivated fish in the world. The dominant species is the Nile tilapia (*Oreochromis niloticus*); however, other species of the genus *Oreochromis* (Neira, 2010) are also cultivated. The GIFT (Genetic Improvement of Farmed Tilapias) program, begun in 1987 in the Philippines, systematically compared wild and commercial strains in various aquatic environments and established a family-based selection system to improve growth rate (Eknath et al., 1993). The program is currently managed by the World Fish Center in Malaysia and genetic gains for growth-related traits are among 10–15% (Ponzoni et al., 2011).

Breeding programs have recently been established for other important species such as, sea bass (*Dicentrarchus labrax*; Vandeputte et al., 2009), sea bream (*Sparus aurata*), turbot (*Scophthalmus maximus*), Atlantic cod (*Gadus morhua*; Glover et al., 2011), halibut (Glover et al., 2007), and tuna (Owen, 2011).

All of these domestication and artificial selection processes shape the genomes of cultured fish populations, resulting in selection signatures that could potentially be identified using molecular and statistical methods.

APPROACHES USED FOR DETECTING SELECTION SIGNATURES

When a new allelic variant that does not affect the fitness of individuals originates in a population, it is not affected by natural selection and is said to be neutral. Statistical tests aimed at testing a neutral evolution model can be divided into three main classes: (1) tests based on polymorphisms within species; (2) tests based on the differences between species; and (3) tests that use information within and between species. A description of these three approaches is given below.

TESTS BASED ON POLYMORPHISMS WITHIN SPECIES

Frequency spectrum

The frequency spectrum is defined as the allele frequency distribution of a large number of independent loci in a given sample (Nielsen, 2005; Vogl and Clemente, 2012). Deviations from expectations of the neutral model (no selection, recombination, population subdivision, or changes in the effective population size) could be indicative of selection: purifying or negative selection tends to increase the fraction of mutations segregating at low frequencies, while positive selection increases the number of alleles observed at high frequencies (Hurst, 2009).

Many tests for detecting selection signatures are based on information provided by the frequency spectrum obtained from DNA sequence data. One of the most commonly used is the Tajima's (1989) D test, which compares two measures of genetic variation (θ). The first is obtained from the average of nucleotide differences between pairs of sequences, and the second is the total number of segregating sites (Nielsen, 2005). If the difference between these two measures is greater than expected under neutral evolution, this model is rejected. Other tests have incorporated phylogenetic information in order to estimate the direction of change and increase power to detect deviations from the null hypothesis of the neutral model (Perfectti et al., 2009). One such test is that of Fu and Li (1993), which also calculates a statistic based on the comparison of two genetic variation estimates, adding phylogenetic

information. For example, a related species may be added as an outgroup, such as the inclusion of the chimpanzee in an analysis of human genetic variation (Nielsen, 2005). Likewise, Fay and Wu (2000) developed a test based on the concept that the frequency spectrum expected under neutrality must be enriched with mutations at low frequencies, and that therefore, mutations at high frequencies are atypical.

Researchers have used this approach to detect selection signatures in several species. In humans, for example, evidence of selection has been found in genes related to the immune system and social behavior (Sabeti et al., 2002; Williamson, 2007). In other species such as chickens, it has been possible to identify genomic regions related to production-related traits such as eggshell hardness and immune system characteristics (Qanbari et al., 2012).

Linkage disequilibrium (LD) and haplotype structure

Linkage disequilibrium (LD) refers to the non-random association of alleles at two or more loci. That is, if two alleles at two loci segregate together in greater proportion than expected by chance, it is said that these loci are in linkage disequilibrium. This measure has been widely used to study various demographic events and evolutionary processes in plants and animals, such as breeding systems, patterns of geographic subdivision, events of natural, and artificial selection, gene conversion, mutation, and other forces that can cause changes in gene frequency (Slatkin, 2008). The LD is affected by different evolutionary factors, including recombination, admixture, bottlenecks, gene flow, genetic drift, inbreeding, and selection (Slatkin, 2008). As a consequence, LD across the genome can vary within and between populations.

Thus, another approach to detect genomic selection signatures is based on statistical comparisons of atypical LD patterns at specific haplotypes of certain genomic regions that are inconsistent with the neutral evolution model (Mueller, 2004). This approach has been used in numerous studies to detect selection signatures in humans and in domesticated species (Sabeti et al., 2002; Voight et al., 2006; Hayes et al., 2008). These studies are based on the concept that in a large population, a neutral variant, which by definition is not under selection, will take many generations to become fixed or lost. Recombination and the passing of generations act with stronger intensity, and therefore, LD around these neutral alleles erodes quickly, leaving a smaller surrounding haplotype (Kimura, 1983; Nielsen et al., 2005b; Sabeti et al., 2006).

Conversely, alleles under positive or balanced selection carry other linked alleles with them, generating increased LD in the genomic region, as described for the hitchhiking effect (Smith and Haigh, 1974). LD between these alleles is slowly eroded, such that the adjacent haplotype is longer than expected by chance (Sabeti et al., 2002). Thus, large haplotypes reflect positive selection. This forms the basis of the EHH statistic ("extended haplotype homozygosity") suggested by Sabeti et al. (2002), which is defined as the probability that two randomly selected chromosomes carrying the core haplotype are identical by descent, and also measures the decay of haplotype homozygosity as a function of the distance. EHH allows for identification of regions with atypical frequencies of extended haplotypes and has been effectively used to detect

signatures of recent positive selection within a population (Tang et al., 2004; Walsh et al., 2006).

Voight et al. (2006) developed the statistic iHS or “integrated Haplotype Score” which allows to compare the area under the curve of EHH distribution between ancestral and derived alleles. This approach is based on the fact that the EHH area of an allele under selection will be greater than that of a neutral allele; therefore, the integral of EHH captures this effect. iHS corresponds to a standardized ratio between the areas under the curve of ancestral and derived alleles, which is equal to 0 when the EHH decay is similar for both types of alleles. A negative iHS value near -1 indicates extended haplotype around a derived allele, whereas positive values near one indicate extended haplotype around an ancestral allele.

The iHS statistic is more sensitive for detecting rapid increases in frequencies of the derived allele produced by selection. However, it cannot detect selection signatures resulting from complete or nearly complete fixation of a beneficial allele in the population, and therefore cannot detect a significant fraction of variants under positive selection (Qanbari et al., 2011). For this reason, Tang et al. (2007) reported a new method involving comparison of EHH at the same site, but between populations, i.e., an approach based on the genetic diversity among divergent populations. These statistics are called site-specific EHH ($EHHS$); the area under the $EHHS$ curve (iES); and the standardized ratio of iES between two populations (Hellmann et al., 2003), which reflect haplotype variation among populations. The search for selection signatures from EHH statistical derivatives has been performed in several species such as cattle (Qanbari et al., 2011), poultry (Li et al., 2012; Zhang et al., 2012), swine (Ai et al., 2013) and humans (Sabeti et al., 2007).

Index of population differentiation

The F_{ST} (Wright, 1951) is a statistical measure of genetic variation due to differences in allele frequencies between and within populations (Holsinger and Weir, 2009). The F_{ST} statistic has been one of the most widely used methods for detecting genomic regions that have been under selection (Gianola et al., 2010; Qanbari et al., 2011). The F_{ST} for a locus that has been selected in one population but not another will be higher than in other loci not affected by selection, where genetic diversity is mainly caused by genetic drift (Holsinger and Weir, 2009). Genetic drift affects all loci in the genome similarly; however, loci under selection often behave differently and therefore may present atypical patterns of variation. These atypical patterns can be determined by genotyping, for example, a large number of single nucleotide polymorphisms (SNP) throughout the whole genome, where loci influenced by selection may be identified by deviations from the empirical distribution of F_{ST} statistic (Cavalli-Sforza, 1966; Akey et al., 2002). That is, relative to a neutral model, outliers with value below a certain level suggest the effect of balanced selection, while outliers with values above a certain level are indicative of directional selection.

Various estimates of the F_{ST} statistic have been developed and applied in a number of studies to search for selection signatures (Akey et al., 2002; Hayes et al., 2009; Amaral et al., 2011). However, although the outlier approach may be effective in identifying genes

under selection, it poses several challenges, such as susceptibility to genotyping errors, population stratification, and false positives, as well as variations in mutation rate and low sensitivity (Narum and Hess, 2011). It is also well known that the outlier detection methods have limited power to detect disruptive selection (Beaumont and Balding, 2004) and weak forms of divergent selection (Wright and Gaut, 2005).

TESTS BASED ON DIFFERENCES BETWEEN SPECIES

The statistical methodology to detect selection signatures by comparing information between species relies on the fact that genomic substitutions in coding regions are present in two forms: non-synonymous mutations (d_n), which can lead to the replacement of amino acids in the resulting proteins, and synonymous mutations (d_s), which do not cause amino acid substitution because of the redundancy of the genetic code (Nielsen, 2005; Biswas and Akey, 2006).

The d_n/d_s ratio provides information about evolutionary forces acting upon a particular gene. For example, at loci under neutrality, the d_n/d_s ratio will be equal to 1. Genes subject to functional limitations, such that a non-synonymous substitution is detrimental, will tend to be eliminated from the population by negative selection; therefore, $d_n/d_s < 1$. Conversely, an excess of non-synonymous mutations over synonymous mutations ($d_n/d_s > 1$) provides evidence for the action of positive selection in favor of non-synonymous substitution, which could provide a comparative advantage at the protein level (Nielsen et al., 2005a).

Based on these concepts, several studies have detected selection in many genes and organisms, such as genes related to immune response (Endo et al., 1996; Hughes, 1997; Sawyer et al., 2004), viral receptor genes (Fitch et al., 1997; Nielsen and Yang, 1998; Bush et al., 1999), genes associated with fertility (Swanson et al., 2001, 2003), and genes involved in sensory perception and smell in humans (Gilad et al., 2000).

TESTS THAT USE INFORMATION WITHIN AND BETWEEN SPECIES

The neutral theory of molecular evolution indicates that genomic regions that evolve rapidly and, thus, have high divergence between species, will also show high levels of polymorphisms within species. The Hudson–Kreitman–Aguade (HKA) test compares the level of polymorphisms within each species and observed divergence between related species for two or more loci. The test can determine if it is likely that the observed difference is due to neutral or adaptive evolution (Hudson et al., 1987). The HKA test is the precursor to the McDonald–Kreitman test (Howe et al., 2013), which compares synonymous (P_S) and non-synonymous (P_N) mutations at a specific locus that are polymorphic within a species and synonymous (D_S) and non-synonymous (D_N) mutations that are fixed between species. Under neutrality, the ratios between P_N/P_S and D_N/D_S should be the same, while positive selection leads to increased divergence of synonymous substitutions ($D_N/D_S > P_N/P_S$; McDonald and Kreitman, 1991).

GENOMIC RESOURCES IN FISH

In recent decades, the development of DNA markers has greatly contributed to the study of animal genetics. DNA markers allow us

to observe and exploit variation across the genome of an individual (Liu and Cordes, 2004; Tier, 2010).

In fish, a wide range of DNA markers have been used, including amplified fragment length polymorphisms (AFLP), random amplified polymorphic DNA (RAPD), sequence tagged sites (STS), variable number of tandem repeats (VNTR), microsatellites or simple sequence repeats (SSR), SNP, and expressed sequence tags (EST; Liu, 2007). Currently, with the development of high-throughput sequencing technologies many gigabases of nucleotide sequences can be generated in a short period of time, and many SNP and other polymorphisms can be detected using bioinformatics methods (Liu, 2011). These techniques provide an affordable and reliable scale of DNA sequencing in several organisms (Mardis, 2008). They are extensively used in *de novo* sequencing, quantification of gene expression by RNA-seq (“RNA sequencing”; Wang et al., 2009), massive identification of SNP markers using RAD-sequencing (“restriction site associated DNA sequencing”; Rowe et al., 2011), and population genomics studies (Hohenlohe et al., 2010; De Wit et al., 2012).

Although teleost fish are the largest group of vertebrates (about 27,000 species), they are underrepresented in genome sequencing projects (Spaink et al., 2013). **Table 1** shows some of the species that have undergone genome sequencing projects to date.

Extracted and modified from Spaink et al. (2013). The terms scaffolds or contigs indicate that the genome of the species has been partially sequenced, and the term chromosome indicates that sequencing has been anchored to the existing physical map of the species.

SELECTION SIGNATURES IN FISH

In fish, studies aimed at detecting selection signatures are performed mainly in the context of molecular ecology disciplines. Most of them have been limited to a low level of resolution and restricted to specific genomic regions.

MODEL FISH SPECIES

Using SNP markers from ESTs, loci with outlier F_{ST} values were identified in wild populations of zebrafish (*Danio rerio*), suggesting directional selection in genes associated with energy metabolism, homeostasis regulation, and signal transduction, which could be associated with local adaptation among different populations. Further, evidence was found to suggest balanced selection of the gene encoding the receptor for the NS1A influenza virus (Whiteley et al., 2011). In the same study, outlier F_{ST} values were found for loci in laboratory strains related to oxidoreductase activity, chromatin condensation, immune response, and induction of apoptosis, among other processes, which could be associated with the domestication process of cultured strains (Whiteley et al., 2011).

CICHLIDS

Keller et al. (2013) detected outlier SNP patterns between five cichlid species from the Lake Victoria area in East Africa, identifying signatures of divergent selection between the two genera that include these species. These selection signals were associated with male color, depth distribution, feeding patterns, and

morphological traits that distinguish the genera. Moreover, evidence has been found to suggest selection in the homeobox genes (*dlx*) involved in the development of the nervous system, the craniofacial skeleton, and the formation of connective tissue and appendages (Diepeveen et al., 2013).

SALMONIDS

In lake whitefish (*Coregonus clupeaformis*), a fish of the salmon family distributed along northern Alaska and all of Canada, 24 loci were identified that revealed selection signatures associated with QTL of certain adaptive traits such as natatorium behavior, growth rate, morphology, and reproduction characters (Rogers and Bernatchez, 2007).

In Atlantic salmon, Vasemägi et al. (2012) used microsatellite markers and SNP to locate 10 genomic regions showing signatures of directional selection related to characteristics such as growth rate and morphology. Martinez et al. (2013) found strong evidence of selection in a microsatellite marker on chromosome 3, which harbored QTL for body weight. Furthermore, there is evidence that genes associated with immune response have been subject to greater selection pressure compared with other regions of the genome (Tonteri et al., 2010; Portnoy et al., 2014). Other studies in genera *Oncorhynchus*, *Salmo*, and *Salvelinus* have revealed signatures of balanced selection for genes of the major histocompatibility complex IIB (Aguilar and Garza, 2007; Limborg et al., 2012). In brown trout, analysis with markers linked to genes related to the immune response showed evidence of having been subjected to selection (Jensen et al., 2008). Finally, evidence was found in both brown trout and sockeye salmon of disruptive selection at two loci within the major histocompatibility complex IIB (Hansen et al., 2010; Gomez-Uchida et al., 2011; Meier et al., 2011).

OTHER FAMILIES

In guppies (*Poecilia reticulata*), outlier F_{ST} values suggest that between 3.5 and 6.5% of SNP markers are under directional selection. Some of these loci are near QTL associated with ornamental traits, and they are also in EST (Willing et al., 2010).

In Atlantic cod in 1960, Sick detected evidence of selection in the locus encoding Hemoglobin (*Hb I*) and the *Pan I* locus, which encodes a protein related to the neuroendocrine system and has recently been associated with vesicle transport in adipocytes (Pogson, 2001). In the same species, Moen et al. (2008) identified 29 SNP with outlier F_{ST} , suggesting that these loci are or have been under selection. These loci were found in genes involved in muscle contraction, immune response, and production of ribosomal proteins. Moreover, Nielsen (Nielsen et al., 2009) found evidence of directional selection for local adaptation to various environmental conditions, such as loci with outlier F_{ST} values associated with genes involved in the production of proteins for thermal shock (*Hsp90*), determination of sexual behavior (*Aromatasa*), and formation of photoreceptor cells for perception of light (*rhodopsin*).

In stickleback (*Gasterosteus aculeatus*), a fish of the *Gasterosteidae* family, studies using microsatellite markers to assess genetic diversity among marine and freshwater populations revealed evidence of directional selection that might be associated with

Table 1 | Sequenced genomes available in fish as of 2013.

Order	Family	Species	Common name	Size estimation (Mb)	Chromosomes	Scaffolds	Status	Reference
Anguilliformes	Anguillidae	<i>Anguilla anguilla</i>	European eel	1,000	–	–	Scaffolds or contigs	Henkel et al. (2012a)
		<i>Anguilla japonica</i>	Japanese eel	1,150	–	323,776	Scaffolds or contigs	Henkel et al. (2012b)
Beloniformes	Adrianchthyidae	<i>Oryzias latipes</i>	Japanese rice fish	585.33	24	82,496	Chromosomes	Kawahara et al. (2007)
Characiformes	Characidae	<i>Astyanax mexicanus</i>	Mexican Tetra	964.31	–	10,735	Scaffolds or contigs	Hinaux et al. (2013)
Coelacanthiformes	Latimeriidae	<i>Latimeria chalumnae</i>	African coelacanth	2,183.72	–	22,818	Scaffolds or contigs	Amemiya et al. (2013)
Cypriniformes	Cyprinidae	<i>Danio rerio</i>	Zebra fish	1,412.47	25	4,560	Chromosomes	Hinaux et al. (2013), Howe et al. (2013)
		<i>Cyprinus carpio</i>	Common carp	875	25	–	Chromosomes	Xu et al. (2014)
Cyprinodontiformes	Nothobranchiidae	<i>Nothobranchius furzeri</i>	Turquoise killifish	1,500	–	5,299	Scaffolds or contigs	Reichwald et al. (2009)
		<i>Nothobranchius kuhntae</i>	Beira killifish	–	–	5,934	Scaffolds or contigs	
Gadiformes	Poeciliidae	<i>Xiphophorus maculatus</i>	Southern platyfish	652.84	–	20,640	Scaffolds or contigs	Schartl et al. (2013)
		<i>Gadus morhua</i>	Atlantic cod	608.29	–	427,427	Scaffolds or contigs	Star et al. (2011)
Gasterosteiformes	Gasterosteidae	<i>Gasterosteus aculeatus</i>	Three-spined stickleback	446.62	–	–	Scaffolds or contigs	
Lepisosteiformes	Lepisosteidae	<i>Lepisosteus oculatus</i>	Spotted gar	945.86	29	2,105	Chromosomes	
Perciformes	Cichlidae	<i>Oreochromis niloticus</i>	Nile tilapia	816.12	–	5,901	Scaffolds or contigs	
		<i>Pundamilia nyererei</i>	Python island	698.8	–	7,236	Scaffolds or contigs	
		<i>Rhamphochromis esox</i>	–	1,100	–	55,751	Scaffolds or contigs	Loh et al. (2008)
		<i>Mchenga conophoros</i>	–	1,100	–	61,923	Scaffolds or contigs	Loh et al. (2008)
		<i>Melanochromis auratus</i>	Golden mbuna	1,100	–	63,297	Scaffolds or contigs	Loh et al. (2008)
		<i>Neolamprologus brichardi</i>	Princess Burundi	685.96	–	9,098	Scaffolds or contigs	
		<i>Haplochromis burtoni</i>	–	698.98	–	8,001	Scaffolds or contigs	
		<i>Labeotropheus fuelleborni</i>	Blue mbuna	1,100	–	58,245	Scaffolds or contigs	Loh et al. (2008)
		<i>Maylandia zebra</i>	Zebra mbuna	1,100	–	65,094	Scaffolds or contigs	Loh et al. (2008)
Perciformes	Moronidae	<i>Dicentrarchus labrax</i>	European sea bass	98.25	–	–	Scaffolds or contigs	
Pleuronectiformes	Cynoglossidae	<i>Cynoglossus semilaevis</i>	Flatfish	477	21	80,677	Chromosomes	Chen et al. (2014)
Salmoniformes	Salmonidae	<i>Oncorhynchus mykiss</i>	Rainbow trout	1,877.5	29	79,941	Chromosomes	Berthelot et al. (2014)
		<i>Salmo salar</i>	Atlantic salmon	2,435.31	29	843,055	Chromosomes	Davidson et al. (2010)
Tetraodontiformes	Tetraodontidae	<i>Takifugu flavidus</i>	Sansai fugu	314.95	–	34,332	Scaffolds or contigs	
		<i>Takifugu rubripes</i>	Pufferfish	281.57	22	7091	Chromosomes	Aparicio et al. (2002)
		<i>Tetraodon nigroviridis</i>	Green spotted puffer	308.45	–	–	Scaffolds or contigs	

adaptation of certain populations to freshwater environments (Mäkinen et al., 2008).

FUTURE DIRECTIONS

The search for and detection of genomic signatures produced by selection has provided valuable information that contributes to the understanding of evolutionary forces affecting the genome and gene functions that control phenotypes of biological and economic interest (Nielsen, 2005; Nielsen et al., 2007).

Some fish species provide the great advantage of simultaneous availability as both a wild and a cultivated population. Additionally, these species have unique characteristics in terms of population structure and intra-specific adaptive divergence, mainly due to the diversity of environmental conditions that fish populations inhabit, resulting in populations that exhibit characteristics of strong local adaptation. Comparative studies among these populations would provide benefits in terms of elucidating the effects of selective processes and recent domestication events, which could improve the understanding regarding the impact of the interaction between domesticated and wild populations, the identification of genetic factors involved in economically important traits for aquaculture and unravelling the actual phenotypic variation within and between fish populations.

In domesticated species, the main motivation behind the search for selection signatures lies in the possibility of finding genes or genomic regions associated with traits of economic interest. The development of next-generation sequencing technologies and high-throughput genotyping has made it possible to investigate the effect of selective pressures on genome variation in several domesticated species. In cattle and sheep, researchers have detected selection signatures associated with carcass yield traits, tail fat deposition, dairy traits (Moradi et al., 2012; Rothhammer et al., 2013), reproductive traits (Gautier and Naves, 2011; Qanbari et al., 2011), immune response (Gautier and Naves, 2011), coat color, and horn development (Druet et al., 2013), among other characters of interest. Also, in swine, selection signatures have been identified in genomic regions associated with traits such as coat color, ear morphology, reproductive characteristics, and fat deposition (Wilkinson et al., 2013). In chickens, researchers have identified selection signatures associated with eggshell hardness and immune system characteristics (Qanbari et al., 2012), to mention just a few examples. These studies may provide a basis for conducting similar research that allows for investigation of the genomic regions affected by the processes of domestication and natural or artificial selection in fish populations, allowing for discovery of new genes that underlie phenotypic traits of interest and understanding processes relevant for conservation purposes.

There is currently little genomic information for fish species as compared to humans or domesticated animals. This is one of the reasons why selection signatures studies have been conducted in only a few species and generally limited to a low level of genomic coverage (Vasemägi et al., 2005, 2012). However, recent advances in genomic technologies, including high quality reference genome sequences, construction of genetic maps, and development of high-density SNP arrays are paving the way for systematic study of genetic variation in these species.

The development and application of next-generation sequencing approaches will represent a powerful strategy to improve the resolution and accuracy when detecting regions under selection in several species. This may lead to determination of the causative genetic factors involved in several biological aspects of aquaculture species. However, the application of these results in the aquaculture development requires further studies aiming at determining effective and practical applications of this technology. Candidate disciplines to be benefited from the discovery of selected regions using next-generation sequencing are, for example, genetic improvement, vaccine and pharmaceutical development and fish nutrition.

CONCLUSION

The development of genomic methodologies has contributed greatly to the study of genetic variation between and within species. High-resolution studies at the level of the whole genome can identify selection signatures explaining phenotypic variation between and within populations, and therefore potentially identify genetic variants underlying characteristics of biological and economic interest.

Although the application and utility of these techniques in aquaculture species has been limited by a lack of genomic information, there is a great potential for conducting such studies, especially in species for which there are genome sequencing projects and high-density molecular markers platforms availability.

ACKNOWLEDGMENTS

This work was supported by the Programa Formación Capital Humano Avanzado from CONICYT, Government of Chile, and partially funded by grants from CORFO (11IEI-12843 and 12PIE-17669), Government of Chile.

REFERENCES

- Aguilar, A., and Garza, J. (2007). Patterns of historical balancing selection on the salmonid major histocompatibility complex class II β gene. *J. Mol. Evol.* 65, 34–43. doi: 10.1007/s00239-006-0222-8
- Ai, H., Huang, L., and Ren, J. (2013). Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS ONE* 8:e56001. doi: 10.1371/journal.pone.0056001
- Akey, J., Zhang, G., Zhang, K., Jin, L., and Shriver, M. (2002). Interrogating a high density SNP map for signatures of natural selection. *Genome Res.* 12, 1805–1814. doi: 10.1101/gr.631202
- Amaral, A. J., Ferretti, L., Megens, H.-J., Crooijmans, R. P. M. A., Nie, H., Ramos-Onsins, S. E., et al. (2011). Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. *PLoS ONE* 6:e14782. doi: 10.1371/journal.pone.0014782
- Amemiya, C. T., Alfoldi, J., Lee, A. P., Fan, S., Philippe, H., Maccallum, I., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496, 311–316. doi: 10.1038/nature12027
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-M., Dehal, P., et al. (2002). Whole-genome shotgun assembly and analysis of the genome of *fugu rubripes*. *Science* 297, 1301–1310. doi: 10.1126/science.1072104
- Aquadro, C. F., Bauer Dumont, V., and Reed, F. A. (2001). Genome-wide variation in the human and fruitfly: a comparison. *Curr. Opin. Genet. Dev.* 11, 627–634. doi: 10.1016/S0959-437X(00)00245-248
- Beaumont, M. A., and Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13, 969–980. doi: 10.1111/j.1365-294X.2004.02125.x

- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5:3657. doi: 10.1038/ncomms4657
- Biswas, S., and Akey, J. (2006). Genomic insights into positive selection. *Trends Genet.* 22, 437–436. doi: 10.1016/j.tig.2006.06.005
- Black, W. C. IV, Baer, C. F., Antolin, M. F., and Duteau, N. M. (2001). POPULATION GENOMICS: genome-wide sampling of insect populations. *Annu. Rev. Entomol.* 46, 441–469. doi: 10.1146/annurev.ento.46.1.441
- Bush, R. M., Fitch, W. M., Bender, C. A., and Cox, N. J. (1999). Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol. Biol. Evol.* 16, 1457–1465. doi: 10.1093/oxfordjournals.molbev.a026057
- Cavalli-Sforza, L. (1966). Population structure and human evolution. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 164, 362–379. doi: 10.1098/rspb.1966.0038
- Charlesworth, B. (1993). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* 63, 213–227. doi: 10.1017/S0016672300032365
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46, 253–260. doi: 10.1038/ng.2890
- Cutter, A. D., and Paysseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* 14, 262–274. doi: 10.1038/nrg3425
- Davidson, W. S., Koop, B. F., Jones, S. J., Iturra, P., Vidal, R., Maass, A., et al. (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol.* 11:403. doi: 10.1186/gb-2010-11-9-403
- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., et al. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12, 1058–1067. doi: 10.1111/1755-0998.12003
- Diepeveen, E., Kim, E., and Salzburger, W. (2013). Sequence analyses of the distal-less homeobox gene family in East African cichlid fishes reveal signatures of positive selection. *BMC Evol. Biol.* 13:153. doi: 10.1186/1471-2148-13-153
- Dominik, S., Henshall, J. M., Kube, P. D., King, H., Lien, S., Kent, M. P., et al. (2010). Evaluation of an Atlantic salmon SNP chip as a genomic tool for the application in a Tasmanian Atlantic salmon (*Salmo salar*) breeding population. *Aquaculture* 308(Suppl. 1), S56–S61. doi: 10.1016/j.aquaculture.2010.05.038
- Donaldson, L. R., and Olson, P. R. (1957). Development of rainbow trout brood stock by selective breeding. *Trans. Am. Fish. Soc.* 85, 93–101. doi: 10.1577/1548-8659(1955)85[93:DORTBS]2.0.CO;2
- Druet, T., Pérez-Pardal, L., Charlier, C., and Gautier, M. (2013). Identification of large selective sweeps associated with major genes in cattle. *Anim. Genet.* 44, 758–762. doi: 10.1111/age.12073
- Eknath, A. E., Tayamen, M. M., Palada-De Vera, M. S., Danting, J. C., Reyes, R. A., Dionisio, E. E., et al. (1993). Genetic improvement of farmed tilapia: the growth performance of eight strains of *Oreochromis niloticus* tested in different farm environments. *Aquaculture* 111, 171–188. doi: 10.1016/0044-8486(93)90035-W
- Endo, T., Ikeo, K., and Gjobori, T. (1996). Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* 13, 685–690. doi: 10.1093/oxfordjournals.molbev.a025629
- Fay, J. C., and Wu, C.-I. (2000). Hitchhiking under positive darwinian selection. *Genetics* 155, 1405–1413.
- Fitch, W. M., Bush, R. M., Bender, C. A., and Cox, N. J. (1997). Long term trends in the evolution of H(3) HA1 human influenza type. *A. Proc. Natl. Acad. Sci. U.S.A.* 94, 7712–7718. doi: 10.1073/pnas.94.15.7712
- Flori, L., Fritz, S., Jaffre Zic, F., Boussaha, M., Gut, I., Heath, S., et al. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE* 4:e6595. doi: 10.1371/journal.pone.0006595
- Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.
- Futuyma, D. J. (1998). *Evolutionary Biology*. Sunderland, MA: Sinauer Associates Inc.
- Gautier, M., and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 3128–3143. doi: 10.1111/j.1365-294X.2011.05163.x
- Gianola, D., Simianer, H., and Qanbari, S. (2010). A two-step method for detecting selection signatures using genetic markers. *Genet. Res. (Camb.)* 92, 141–155. doi: 10.1017/S0016672310000121
- Gilad, Y., Segre, D., Skorecki, K., Nachman, M. W., Lancet, D., and Sharon, D. (2000). Dichotomy of single-nucleotide polymorphism haplotypes in olfactory receptor genes and pseudogenes. *Nat. Genet.* 26, 221–224. doi: 10.1038/79957
- Gjedrem, T. (2005). *Selection and Breeding Programs in Aquaculture*. Dordrecht: Springer. doi: 10.1007/1-4020-3342-7
- Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: A personal opinion review. *Aquaculture* 344–349, 12–22. doi: 10.1016/j.aquaculture.2012.03.003
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350–353, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Glover, K. A., Dahle, G., and Jørstad, K. E. (2011). Genetic identification of farmed and wild Atlantic cod, *Gadus morhua*, in coastal Norway. *ICES J. Mar. Sci.* 68, 901–910. doi: 10.1093/icesjms/fsr048
- Glover, K. A., Otterå, H., Olsen, R. E., Slinde, E., Taranger, G. L., and Skaala, Ø. (2009). A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. *Aquaculture* 286, 203–210. doi: 10.1016/j.aquaculture.2008.09.023
- Glover, K. A., Svåsand, T., Olesen, L., and Rye, M. (2007). *Atlantic Halibut – Hippoglossus Hippoglossus: Genetic Impact of Aquaculture Activities on Native Populations*. GenImpact Final Scientific Report (EU contract n. RICA-CT-2005-022802), 176. Available at: <http://genimpact.imr.no>
- Gomez-Uchida, D., Seeb, J., Smith, M., Habicht, C., Quinn, T., and Seeb, L. (2011). Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evol. Biol.* 11:48. doi: 10.1186/1471-2148-11-48
- Gu, J., Orr, N., Park, S., Katz, L., Sulimova, G., Machugh, D., et al. (2009). A genome scan for positive selection in thoroughbred horses. *PLoS ONE* 4:e5767. doi: 10.1371/journal.pone.0005767
- Hansen, M. M., Meier, K., and Mensberg, K.-L. D. (2010). Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Mol. Ecol.* 19, 1787–1800. doi: 10.1111/j.1365-294X.2010.04615.x
- Hayes, B., Chamberlain, A., Maceachern, S., Savin, K., Mcpartlan, H., Macleod, I., et al. (2009). A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. *Anim. Genet.* 40, 176–184. doi: 10.1111/j.1365-2052.2008.01815.x
- Hayes, B., Lien, S., Nilsen, H., Olsen, H., Berg, P., Maceachern, S., et al. (2008). The origin of selection signatures on bovine chromosome six. *Anim. Genet.* 39, 105–111. doi: 10.1111/j.1365-2052.2007.01683.x
- Hellmann, I., Ebersberger, I., Ptak, S. E., Paabo, S., and Przeworski, M. (2003). A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72, 1527–1535. doi: 10.1086/375657
- Henkel, C. V., Burgerhout, E., De Wijze, D. L., Dirks, R. P., Minegishi, Y., Jansen, H. J., et al. (2012a). Primitive duplicate hox clusters in the European Eel's genome. *PLoS ONE* 7:e32231. doi: 10.1371/journal.pone.0032231
- Henkel, C. V., Dirks, R. P., De Wijze, D. L., Minegishi, Y., Aoyama, J., Jansen, H. J., et al. (2012b). First draft genome sequence of the Japanese eel, *Anguilla japonica*. *Gene* 511, 195–201. doi: 10.1016/j.gene.2012.09.064
- Hermisson, J., and Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169, 2335–2352. doi: 10.1534/genetics.104.036947
- Hinaux, H., Poulain, J., Da Silva, C., Noirot, C., Jeffery, W. R., Casane, D., et al. (2013). De novo sequencing of astyanax mexicanus surface fish and páchón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. *PLoS ONE* 8:e53553. doi: 10.1371/journal.pone.0053553
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genetics* 6:e1000862. doi: 10.1371/journal.pgen.1000862
- Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* 10, 639–650. doi: 10.1038/nrg2611
- Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008). Major quantitative trait loci affect resistance to infectious

- pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178, 1109–1115. doi: 10.1534/genetics.107.082974
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–503. doi: 10.1038/nature12111
- Hudson, R. R., Kreitman, M., and Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153–159.
- Hughes, A. L. (1997). Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. *Mol. Biol. Evol.* 14, 1–5. doi: 10.1093/oxfordjournals.molbev.a025694
- Hurst, L. D. (2009). Genetics and the understanding of selection. *Nat. Rev. Genet.* 10, 83–93. doi: 10.1038/nrg2506
- Ilyassov, Y. I. (1987). *Selection, Hybridization, and Genetic Engineering in Aquaculture*. Berlin: H. Heenemann GmbH & Company.
- Jensen, L. F., Hansen, M. M., Mensberg, K. L. D., and Loeschke, V. (2008). Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.). *Heredity* 100, 79–91. doi: 10.1038/sj.hdy.6801067
- Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447, 714–719. doi: 10.1038/nature05846
- Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., et al. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol. Ecol.* 22, 2848–2863. doi: 10.1111/mec.12083
- Kim, J., Withler, R., Ritland, C., and Cheng, K. (2004). “Genetic variation within and between domesticated chinook salmon, *Oncorhynchus tshawytscha* strains and their progenitor populations,” in *Genetics of Subpolar Fish and Invertebrates*, eds A. Gharrett, R. Gustafson, J. Nielsen, J. Seeb, L. Seeb, W. Smoker, et al. (Dordrecht: Springer), 371–378.
- Kim, Y., and Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167, 1513–1524. doi: 10.1534/genetics.103.025387
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press. doi: 10.1038/nrg2473
- Li, D. F., Liu, W. B., Liu, J. F., Yi, G. Q., Lian, L., Qu, L. J., et al. (2012). Whole-genome scan for signatures of recent selection reveals loci associated with important traits in White Leghorn chickens. *Poult. Sci.* 91, 1804–1812. doi: 10.3382/ps.2012-02275
- Limborg, M. T., Blankenship, S. M., Young, S. F., Utter, F. M., Seeb, L. W., Hansen, M. H. H., et al. (2012). Signatures of natural selection among lineages and habitats in *Oncorhynchus mykiss*. *Ecol. Evol.* 2, 1–18. doi: 10.1002/ece3.59
- Liu, Z. E. (2011). *Next Generation Sequencing and Whole Genome Selection in Aquaculture*. Des Moines, IA: Blackwell Publishing Ltd. doi: 10.1002/9780470958964
- Liu, Z. J. (2007). *Aquaculture Genome Technologies*. Oxford, UK: Blackwell Publishing Ltd. doi: 10.1002/9780470277560
- Liu, Z. J., and Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37. doi: 10.1016/j.aquaculture.2004.05.027
- Loh, Y.-H., Katz, L., Mims, M., Kocher, T., Yi, S., and Streebman, J. T. (2008). Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biol.* 9:R113. doi: 10.1186/gb-2008-9-7-r113
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4, 981–994. doi: 10.1038/nrg1226
- Mäkinen, H. S., Shikano, T., Cano, J. M., and Merilä, J. (2008). Hitchhiking mapping reveals a candidate genomic region for natural selection in three-spined stickleback chromosome VIII. *Genetics* 178, 453–465. doi: 10.1534/genetics.107.078782
- Mardis, E. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:15. doi: 10.1146/annurev.genom.9.081307.164359
- Martinez, V., Dettleff, P., Lopez, P., Fernandez, G., Jedlicki, A., Yañez, J. M., et al. (2013). Assessing footprints of selection in commercial Atlantic salmon populations using microsatellite data. *Anim. Genet.* 44, 223–226. doi: 10.1111/j.1365-2052.2012.02387.x
- McAndrew, B., and Napier, J. (2011). Application of genetics and genomics to aquaculture development: current and future directions. *J. Agric. Sci.* 149, 143–151. doi: 10.1017/S0021859610001152
- McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654. doi: 10.1038/351652a0
- Meier, K., Hansen, M. M., Bekkevold, D., Skaala, O., and Mensberg, K. L. D. (2011). An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. *Heredity* 106, 488–499. doi: 10.1038/hdy.2010.164
- Metcalfe, N. B., Valdimarsson, S. K., and Morgan, I. J. (2003). The relative roles of domestication, rearing environment, prior residence and body size in deciding territorial contests between hatchery and wild juvenile salmon. *J. Appl. Ecol.* 40, 535–544. doi: 10.1046/j.1365-2664.2003.00815.x
- Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368
- Moen, T., Hayes, B., Nilsen, F., Delghandi, M., Fjalestad, K., Fevolden, S. -E., et al. (2008). Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genet.* 9:18. doi: 10.1186/1471-2156-9-18
- Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K., and Mcewan, J. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.* 13:10. doi: 10.1186/1471-2156-13-10
- Mueller, J. C. (2004). Linkage disequilibrium for different scales and applications. *Brief. Bioinform.* 5, 355–364. doi: 10.1093/bib/5.4.355
- Narum, S. R., and Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* 11, 184–194. doi: 10.1111/j.1755-0998.2011.02987.x
- Neira, R. (2010). “Breeding in aquaculture species: genetic improvement programs in developing countries,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, 8.
- Neira, R., Díaz, N. F., Gall, G. A. E., Gallardo, J. A., Lhorente, J. P., and Manteola, R. (2006). Genetic improvement in Coho salmon (*Oncorhynchus kisutch*). I: selection response and inbreeding depression on harvest weight. *Aquaculture* 257, 9–17. doi: 10.1016/j.aquaculture.2006.03.002
- Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., and Bekkevold, D. (2009). Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol. Ecol.* 18, 3128–3150. doi: 10.1111/j.1365-294X.2009.04272.x
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218. doi: 10.1146/annurev.genet.39.073003.112420
- Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., et al. (2005a). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170. doi: 10.1371/journal.pbio.0030170
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M., Clark, A., and Bustamante, C. (2005b). Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575. doi: 10.1101/gr.4252305
- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* 8, 857–868. doi: 10.1038/nrg2187
- Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Norris, A. T., Bradley, D. G., and Cunningham, E. P. (1999). Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture* 180, 247–264. doi: 10.1016/S0044-8486(99)00212-214
- Owen, M. (2011). South Australian company completes bluefin tuna breeding program. *The Australian*.
- Pennings, P. S., and Hermisson, J. (2006). Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* 23, 1076–1084. doi: 10.1093/molbev/msj117
- Perfectti, F., Picó, F. X., and Gómez, J. M. (2009). La huella genética de la selección natural. *Ecosistemas* 18, 10–16.
- Pogson, G. H. (2001). Nucleotide polymorphism and natural selection at the panthophysin (Pan I) Locus in the Atlantic Cod, *Gadus morhua* (L.). *Genetics* 157, 317–330.
- Ponzone, R. W., Nguyen, N. H., Khaw, H. L., Hamzah, A., Bakar, K. R. A., and Yee, H. Y. (2011). Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Rev. Aquaculture* 3, 27–41. doi: 10.1111/j.1753-5131.2010.01041.x
- Portnoy, D. S., Hollenbeck, C. M., Vidal, R. R., and Gold, J. R. (2014). A comparison of neutral and immune genetic variation in Atlantic Salmon,

- Salmo salar* L. in Chilean aquaculture facilities. *PLoS ONE* 9:e99358. doi: 10.1371/journal.pone.0099358
- Price, E. O. (1999). Behavioral development in animals undergoing domestication. *Appl. Anim. Behav. Sci.* 65, 245–271. doi: 10.1016/S0168-1591(99)00087-8
- Pritchard, J. K., Pickrell, J. K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20, R208–R215. doi: 10.1016/j.cub.2009.11.055
- Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics* 160, 1179–1189.
- Purdum, C. E. (1993). *Genetics and Fish Breeding*. London: Chapman and Hall, 277.
- Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., et al. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12:318. doi: 10.1186/1471-2164-12-318
- Qanbari, S., Pimentel, E., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A., et al. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41, 377–389. doi: 10.1111/j.1365-2052.2009.02016.x
- Qanbari, S., Strom, T. M., Haberer, G., Weigend, S., Gheyas, A. A., Turner, F., et al. (2012). A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. *PLoS ONE* 7:e49525. doi: 10.1371/journal.pone.0049525
- Reichwald, K., Lauber, C., Nanda, I., Kirschner, J., Hartmann, N., Schories, S., et al. (2009). High tandem repeat content in the genome of the short-lived annual fish *Nothobranchius furzeri*: a new vertebrate model for aging research. *Genome Biol.* 10:R16. doi: 10.1186/gb-2009-10-2-r16
- Rogers, S., and Bernatchez, L. (2007). The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. *Salmonidae*) *Species Pairs*. *Mol. Biol. Evol.* 24, 1423–1438. doi: 10.1093/molbev/msm066
- Ross-Ibarra, J., Morrell, P., and Gaut, B. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8641–8648. doi: 10.1073/pnas.0700643104
- Rothhammer, S., Seichter, D., Forster, M., and Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics* 14:908. doi: 10.1186/1471-2164-14-908
- Rowe, H. C., Renaut, S., and Guggisberg, A. (2011). RAD in the realm of next-generation sequencing technologies. *Mol. Ecol.* 20, 3499–3502. doi: 10.1111/j.1365-294X.2011.05197.x
- Rye, M., Gjerde, B., and Gjedrem, T. (2010). “Genetic improvement programs for aquaculture species in developed countries,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, 8.
- Sabeti, P., Reich, D., Higgins, J., Levine, H., Richter, D., Schaffner, S., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140
- Sabeti, P., Schaffner, S., Fry, B., Lohmueller, J., Vailly, P., Shamovsky, O., et al. (2006). Positive natural selection in the human lineage. *Science* 3129, 1614–1620. doi: 10.1126/science.1124309
- Sabeti, P. C., Vailly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250
- Sawyer, S. L., Emerman, M., and Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* 2:e275. doi: 10.1371/journal.pbio.0020275
- Schaperclaus, W. (1962). *Trate de Pisciculture en Etang*. Paris: Vigot Freres, 208.
- Schartl, M., Walter, R. B., Shen, Y., Garcia, T., Catchen, J., Amores, A., et al. (2013). The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* 45, 567–572. doi: 10.1038/ng.2604
- Slatkin, M. (2008). Linkage disequilibrium [mdash] understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 9, 477–485. doi: 10.1038/nrg2361
- Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genet. Res.* 23, 23–35. doi: 10.1017/S0016672300014634
- Spaink, H. P., Jansen, H. J., and Dirks, R. P. (2013). Advances in genomics of bony fish. *Brief. Funct. Genomics* 13, 144–156. doi: 10.1093/bfpg/elt046
- Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrom, M., Gregers, T. F., et al. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477, 207–210. doi: 10.1038/nature10342
- Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F., and Aquadro, C. F. (2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7375–7379. doi: 10.1073/pnas.131568198
- Swanson, W. J., Nielsen, R., and Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* 20, 18–20. doi: 10.1093/oxfordjournals.molbev.a004233
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5:e171. doi: 10.1371/journal.pbio.0050171
- Tang, K., Wong, L. P., Lee, E. J. D., Chong, S. S., and Lee, C. G. L. (2004). Genomic evidence for recent positive selection at the human MDR1 gene locus. *Hum. Mol. Genet.* 13, 783–797. doi: 10.1093/hmg/ddh099
- Tier, B. (2010). Genomic selection: promises and propriety. *J. Animal Breeding and Genetics* 127, 169–170. doi: 10.1111/j.1439-0388.2010.00876.x
- Tonteri, A., Vasemägi, A., Lumme, J., and Primmer, C. R. (2010). Beyond MHC: signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci. *Mol. Ecol.* 19, 1273–1282. doi: 10.1111/j.1365-294X.2010.04573.x
- Trut, L., Oskina, I., and Kharlamova, A. (2009). Animal evolution during domestication: the domesticated fox as a model. *Bioessays* 31, 349–360. doi: 10.1002/bies.200800070
- Vandeputte, M., Dupont-Nivet, M., Haffray, P., Chavanne, H., Cenadelli, S., Parati, K., et al. (2009). Response to domestication and selection for growth in the European sea bass (*Dicentrarchus labrax*) in separate and mixed tanks. *Aquaculture* 286, 20–27. doi: 10.1016/j.aquaculture.2008.09.008
- Vasemägi, A., Nilsson, J., McGinnity, P., Cross, T., O’Reilly, P., Glebe, B., et al. (2012). Screen for footprints of selection during domestication/captive breeding of Atlantic Salmon. *Comp. Funct. Genomics* 2012:14. doi: 10.1155/2012/628204
- Vasemägi, A., Nilsson, J., and Primmer, C. R. (2005). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol. Biol. Evol.* 22, 1067–1076. doi: 10.1093/molbev/msi093
- Vogl, C., and Clemente, F. (2012). The allele-frequency spectrum in a decoupled Moran model with mutation, drift, and directional selection, assuming small mutation rates. *Theor. Popul. Biol.* 81, 197–209. doi: 10.1016/j.tpb.2012.01.001
- Voight, B., Kudaravalli, S., Wen, X., and Pritchard, J. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72. doi: 10.1371/journal.pbio.0040072
- Walsh, E., Sabeti, P., Hutcheson, H., Fry, B., Schaffner, S., Bakker, P. W., et al. (2006). Searching for signals of evolutionary selection in 168 genes related to immune function. *Hum. Genet.* 119, 92–102. doi: 10.1007/s00439-005-0090-90
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Whiteley, A. R., Bhat, A., Martins, E. P., Mayden, R. L., Arunachalam, M., Uusi-Heikkilä, S., et al. (2011). Population genomics of wild and laboratory zebrafish (*Danio rerio*). *Mol. Ecol.* 20, 4259–4276. doi: 10.1111/j.1365-294X.2011.05272.x
- Wilkinson, S., Lu, Z. H., Megens, H. -J., Archibald, A. L., Haley, C., Jackson, I. J., et al. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genet* 9:e1003453. doi: 10.1371/journal.pgen.1003453
- Williamson, S. H. (2007). Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3:e90. doi: 10.1371/journal.pgen.0030090
- Willing, E.-M., Bentzen, P., Van Oosterhout, C., Hoffmann, M., Cable, J., Breden, F., et al. (2010). Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Mol. Ecol.* 19, 968–984. doi: 10.1111/j.1365-294X.2010.04528.x
- Wright, S. (1951). The genetical structure of populations. *Ann. Eugen.* 15, 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x
- Wright, S. I., and Gaut, B. S. (2005). Molecular Population Genetics and the Search for Adaptive Evolution in Plants. *Mol. Biol. Evol.* 22, 506–519. doi: 10.1093/molbev/msi035
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., et al. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genet.* 46, 1212–1219. doi: 10.1038/ng.3098

Zhang, H., Wang, S.-Z., Wang, Z.-P., Da, Y., Wang, N., Hu, X.-X., et al. (2012). A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. *BMC Genomics* 13:704. doi: 10.1186/1471-2164-13-704

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 October 2014; accepted: 15 December 2014; published online: 14 January 2015.

Citation: López ME, Neira R and Yáñez JM (2015) Applications in the search for genomic selection signatures in fish. *Front. Genet.* 5:458. doi: 10.3389/fgene.2014.00458
This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 López, Neira and Yáñez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture

Paulino Martínez¹* Ana M. Viñas², Laura Sánchez¹, Noelia Díaz³, Laia Ribas³ and Francesc Piferrer³

¹ Departamento de Genética, Facultad de Veterinaria, Universidad de Santiago de Compostela, Lugo, Spain

² Departamento de Genética, Facultad de Biología, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

³ Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas, Barcelona, Spain

Edited by:

Ross Houston, University of Edinburgh, UK

Reviewed by:

Lior David, The Hebrew University of Jerusalem, Israel

Shaojun Liu, Hunan Normal University, China

Christos Palaiokostas, University of Stirling, UK

*Correspondence:

Paulino Martínez, Departamento de Genética, Facultad de Veterinaria, Universidad de Santiago de Compostela, Campus de Lugo, 27002 Lugo, Spain
e-mail: paulino.martinez@usc.es

Controlling the sex ratio is essential in finfish farming. A balanced sex ratio is usually good for broodstock management, since it enables to develop appropriate breeding schemes. However, in some species the production of monosex populations is desirable because the existence of sexual dimorphism, primarily in growth or first time of sexual maturation, but also in color or shape, can render one sex more valuable. The knowledge of the genetic architecture of sex determination (SD) is convenient for controlling sex ratio and for the implementation of breeding programs. Unlike mammals and birds, which show highly conserved master genes that control a conserved genetic network responsible for gonad differentiation (GD), a huge diversity of SD mechanisms has been reported in fish. Despite theory predictions, more than one gene is in many cases involved in fish SD and genetic differences have been observed in the GD network. Environmental factors also play a relevant role and epigenetic mechanisms are becoming increasingly recognized for the establishment and maintenance of the GD pathways. Although major genetic factors are frequently involved in fish SD, these observations strongly suggest that SD in this group resembles a complex trait. Accordingly, the application of quantitative genetics combined with genomic tools is desirable to address its study and in fact, when applied, it has frequently demonstrated a multigene trait interacting with environmental factors in model and cultured fish species. This scenario has notable implications for aquaculture and, depending upon the species, from chromosome manipulation or environmental control techniques up to classical selection or marker assisted selection programs, are being applied. In this review, we selected four relevant species or fish groups to illustrate this diversity and hence the technologies that can be used by the industry for the control of sex ratio: turbot and European sea bass, two reference species of the European aquaculture, and salmonids and tilapia, representing the fish for which there are well established breeding programs.

Keywords: sex determination, fish, genetic architecture, sex ratio, aquaculture

INTRODUCTION

Fish represent the most diverse group of vertebrates including more than 28,000 species (Nelson, 2006). This diversity is a reflection of their high capacity for adaptation to a broad spectrum of environmental conditions. As a result, fish show amazing morphological, physiological, and behavioral adaptations to live in the highly diverse aquatic environment. Fish also show all types of reproductive strategies, including gonochorism, proterandrous, protogynous, and simultaneous hermaphroditism, and unisexuality (Devlin and Nagahama, 2002). These reproductive strategies emerged independently in different lineages during evolution demonstrating a polyphyletic origin (Awise and Mank, 2009).

Domestication of fish for production is an ancient practice and again shows the high adaptation capacity of this group, especially considering that more than 354 fish species are cultivated all over the world (Food and Agriculture Organization [FAO], 2014). Production of domestic fish largely relies on reproduction, and a vast

amount of information has been gathered for its control. Reproduction techniques include production of monosex populations because the existence of sexual growth dimorphism, either in favor of males or females depending on the species, and also because sometimes the most valuable trait is associated with one sex (e.g., color, shape, secondary sexual ornaments).

Here, we review the available data on the genetic architecture of sex determination (SD) in fish and how sex ratio is controlled in aquaculture production. We contrast this information with models emerging from the classical studies in *Drosophila*, mammals, or birds with highly conserved mechanisms associated to marked sex chromosome heteromorphisms. We show the huge intra and interspecific diversity of SD systems in fish associated to a high evolutionary turnover. So, its genetic architecture, although commonly supported by major genes, is also influenced by minor genes, and environmental factors approaching it to a complex trait. We illustrate this diversity and its influence on the strategies aimed at the production of the most

desired sex in the last section of the paper by analyzing the genetic basis of SD in two important species of the European aquaculture, turbot (*Scophthalmus maximus*), and European sea bass (*Dicentrarchus labrax*). We also consider two of the main fish groups with established breeding programs, salmonids, and tilapias, all of them with remarkable differences in sex determination.

GENETIC ARCHITECTURE OF SEX DETERMINATION IN FISH

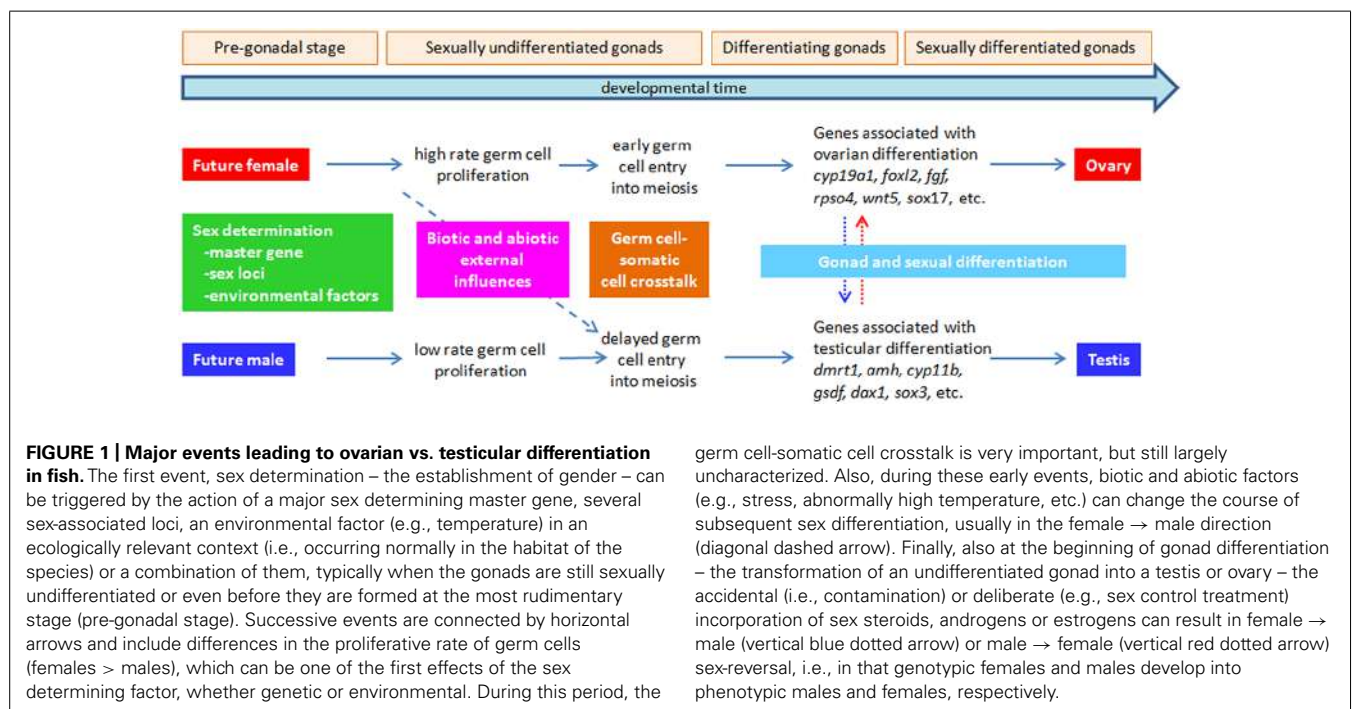
The genetic architecture of a complex trait refers to the genes involved in that trait, their influence and interactions to establish the phenotype. It takes also into account the influence of environmental factors and their interactions with the genotype on the final phenotype (McKay, 2001). Although getting all this information is a very ambitious enterprise, the more we approximate to this goal the better we shall understand key questions related to the genetic variation for its exploitation in genetic breeding programs.

MAIN FEATURES OF GONAD DEVELOPMENT

Gonad development includes all developmental processes aimed to transform an undifferentiated primordium into a mature gonad, either ovary or testis. It is the result of two concatenated development processes controlled by a hierarchical genetic network: SD and gonad differentiation (GD; **Figure 1**). The sex of gonads is essentially determined by processes, either genetic or environmental, operating at the beginning of development, where a binary decision is taken related to the fate of the undifferentiated primordium (Kobayashi and Nagahama, 2009; Siegfried, 2010; Uller and Helanterä, 2011). Once the future of the gonad has been established, morphogenetic GD processes work until maturation is completed.

As a consequence of the hierarchical nature of gonad development, a single gene or environmental cue operating at the beginning of development can drive the gonad pathway towards one direction or another. Thus, in contrast to other characters influenced by many genes operating in different routes with important additive effects, here gene interactions likely represent an important genetic component. Particularly, epistatic effects may be relevant because a single gene acting upstream or even downstream of a preexisting SD gene (SDg) may take the control of gonad fate, thus, masking extant genetic variation at other involved loci. Epistatic interactions have been reported between major SD loci in different fish groups (Cnaani et al., 2008; Ser et al., 2010; Parnell and Streelman, 2013), and also epistatic allelic variants have been reported segregating in populations of species with a well known SD genetic system like medaka (*Oryzias latipes*; Shinomiya et al., 2010). In fact, notable interactions occur between gene products at the initial stages of gonad development such as the suppression of *wnt4* and β -*catenin*, key genes for ovarian development, by *sox9* and *fgf9* (Nef and Vassalli, 2009); the modulation of gonadal aromatase (*cypb19a*), responsible of the balance between androgens and estrogens, by the action of other genes or environmental factors such as temperature (Navarro-Martín et al., 2011); or the interaction between the anti-müllerian hormone (*amh1*) and its receptor (*amhr2*), which triggers an essential signaling pathway for testis development (Kamiya et al., 2012).

Gonad development of fish is unusual in the sense that the sexually undifferentiated period can last from weeks until years (Saito and Tanaka, 2009; Berbejillo et al., 2012) opening a large developmental window in which the sexual fate can be influenced by abiotic or biotic environmental factors (Penman and Piferrer, 2008; Baroiller et al., 2009). In such a long period, it is tempting to speculate that the brain may be involved through



the hypothalamic–pituitary–gonadal axis (Baroiller et al., 2009). However, despite the fact that the brain certainly integrates environmental stimuli and, in particular, social interactions, which have been shown to be implicated in the process of sex-change in hermaphrodites, currently there is no convincing proof that the brain plays any role in the SD process in gonochoristic fishes.

Homoeothermic vertebrates show a conserved morphogenetic development supported by a strongly canalized genetic sex determining system (Charlesworth et al., 2005). However, fish show inter-specific differences in the morphogenetic events occurring along gonad development. Variation exists both regarding the general pattern of differentiation, the interaction between somatic and germ cells, and in the time of occurrence and relative weight of the different steps. The amount of primordial germ cells have been reported to be the first development difference between males and females in species such as medaka (*O. latipes*; Kondo et al., 2009) and stickleback (*Gasterosteus aculeatus*; Lewis et al., 2008), and this has been related to the possible influence of growth-related factors on SD (Schlueter et al., 2007; Baroiller et al., 2009). Also, gonad development has been classified as undifferentiated or differentiated type, respectively, depending on the existence of an initial transitory female stage that can subsequently revert to testis like in zebrafish (*Danio rerio*; Orban et al., 2009), or the lack of that stage like in medaka or tilapia (*Oreochromis niloticus*; Saito and Tanaka, 2009). Interaction between somatic and germ cells is also recognized as an important feature for gonad development. In fact, several key genes related to the initial steps of differentiation like *dmrt1*, *amh*, or *sox9* in males (Lee et al., 2009; Wu et al., 2010) and *cyp19a1a* or *foxl2* in females (Nakamura et al., 2009) are expressed in Sertoli or granulosa/theca cells, respectively, and thus, communication between somatic and germ cells is essential for GD. In this communication there are species-specific differences and, for example, the ablation of the female germ cells determines the reversal of development towards a testis in zebrafish (*D. rerio*) and tilapia (Siegfried and

Nusslein-Volhard, 2008; Fujimoto et al., 2010), while in goldfish (*Carassius auratus*) the female pathway is maintained (Goto et al., 2012).

SEX DETERMINATION: ORIGIN AND EVOLUTION

A consensus existed until recently on the high conservation of the gene network controlling gonad development among vertebrates, differences being mainly related to changes in the switching mechanism. This hierarchical development controlled process would facilitate the control of sex ratio by a single-gene mechanism, but at the same time it would open the opportunity for changing the SD factor in response to new evolutionary scenarios (Marín and Baker, 1998).

Theories on the evolution and genetic architecture of SD in animals have been largely influenced by studies on *Drosophila*, mammals, and birds, all of them showing convergent patterns, with a heteromorphic sex chromosome pair and, as a consequence, a particular sex-linked inheritance model (Bachtrog et al., 2014). A generalized theory on the origin and evolution of SD systems emerged from these data, which assumed a sexual conflict between antagonistic alleles at specific loci favorable to one sex but detrimental to the other (Figure 2; Charlesworth et al., 2005). To maintain the beneficial association between the antagonistic allele and the SD locus, recombination would be restricted, giving rise to the permanent heterozygous state of that portion of the sexual pair (Bergero and Charlesworth, 2009). That circumstance would promote the accumulation of repetitive elements and deleterious variants in the SDg-bearing chromosome, contributing to its progressive degeneration and the typical heteromorphic shape of the sexual pair (Charlesworth et al., 2005). Mathematical models based on this theory suggested that only one gene should underlie the SD system, and if more than one gene were segregating, this should represent an unstable equilibrium towards a new SD mechanism (Rice, 1986).

Initial data in ectothermic vertebrates, particularly fish, demonstrated a sharply different picture (Devlin and Nagahama, 2002).

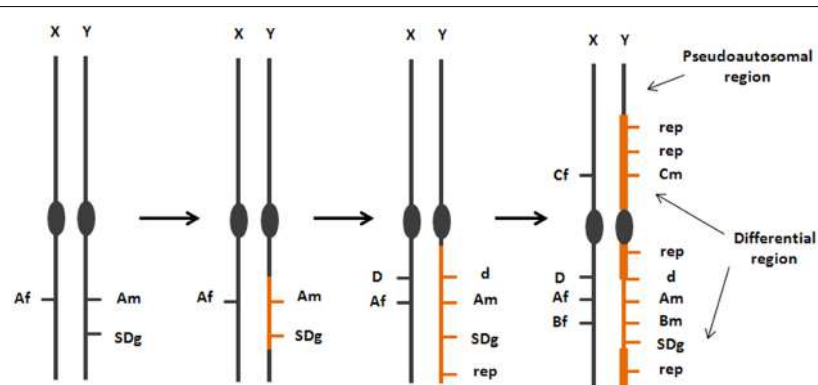


FIGURE 2 | Model on the origin and evolution of the SD region-bearing chromosome from studies in mammals, birds and *Drosophila*. This model, although has been demonstrated in some fish, shows a large variation on its progression, which is reflected on the degree of differentiation between the chromosomes of the sexual pair. The origin of a new sexual pair is related

to the origin of genes (A,B) with antagonistic alleles favorable to females (Af, Bf) or to males (Am, Bm) associated with a new SDg. Subsequent steps involve accumulation of repetitive elements (rep) and the degeneration of the Y chromosome because of its permanent heterozygotic state at the differential region (d: recessive non functional variant of a sex-linked gene).

More than one sex-associated gene has been reported in many fish species (Cnaani et al., 2008; Ser et al., 2010). In addition, the conservation of the GD cascade has been to some degree questioned, and notable genetic differences have been observed not only at the top, but also downstream of the GD network (Böhne et al., 2013; Herpin et al., 2013). For example, the well-known female-associated aromatase genes (*cyp19a1a* and *cyp19a1b*) have shown a new role in testis differentiation in African cichlids (Böhne et al., 2013). Although antagonistic alleles demonstrated to be associated with the SD region in fish (Roberts et al., 2009; Parnell and Streelman, 2013), only 7% species showed heteromorphic sex chromosomes (Penman and Piferrer, 2008; Oliveira et al., 2009). This has been related to the huge evolutionary turnover of SD mechanism which limits chromosome differentiation and, as a consequence, different SD systems have been reported between closely related species and even between populations of the same species. However, degeneration and differentiation of sex chromosomes can evolve very quickly (Charlesworth et al., 2005) and a broad heteromorphism degree has been observed among species of the same genus in Neotropical fish of the genera *Eigenmannia* (Henning et al., 2011), *Characidium* (Vicari et al., 2008), and *Leporinus* (Parise-Maltempi et al., 2013). Thus, sex heteromorphism can involve the whole chromosome and be detectable with the usual cytogenetic techniques, as reported in Neotropical fish (Henning et al., 2011; Parise-Maltempi et al., 2013); be cryptic at cytogenetic level but involving an important degeneration of the SDg-bearing chromosome, as in stickleback (Ross and Peichel, 2008; Shikano et al., 2011); embrace no more than a few kilobases as in medaka (*O. latipes*; Matsuda et al., 2002); or show a very tiny differentiation as the single SNP observed in the *amhr2* receptor, the only detectable difference between X and Y chromosomes in fugu (*Takifugu rubripes*; Kamiya et al., 2012).

To understand the origin of SD regions both external pressures (i.e., sexual selection) and the internal context (available genes and genomic structure) should be considered. Several genes with different functions have been recruited along evolution as SDg in fish, which shows the opportunistic nature of selection to face new evolutionary pressures. In this regard, the specific genome duplication occurred within teleosts may have provided a suitable raw material for new sex determinants (Mank and Avise, 2009). A remarkable case which exemplifies the many options available to change SD mechanisms is the association of sex to B chromosomes, usually considered as junk DNA, in species of the genus *Astyanax* and in two species of pufferfish (Vicente et al., 1996; Noletto et al., 2012). The high turnover of the SD region in fish has led to suggest that changes in the SD mechanism may be associated with speciation (Ser et al., 2010; Böhne et al., 2013; Parnell and Streelman, 2013).

THE GENETIC BASIS OF SEX DETERMINATION IN FISH

High genetic variation has also been described between fish species regarding the gene responsible for SD, the number of genes involved in such decision and the relationships between them. Currently, five different master genes have been documented in fish: *dmY*, *gsdf*, *amhy*, *amhr2*, and *sdY*. *dmY* (DM-domain gene on the Y chromosome), the SDg of medaka and the first one described in fish (Matsuda et al., 2002), is a transcription factor expressed in

the somatic cells surrounding germ cells before sex differentiation and in the testis thereafter and it is involved in germ cell proliferation and development of pre-Sertoli cells into Sertoli cells. It originated from a segmental duplication of a small autosomal region containing the precursor *dmrt1*, followed by an insertion of the duplicated region on the proto-Y chromosome (Matsuda et al., 2002). *gsdf*, *amhy*, and *amhr2* are members of the TGF- β superfamily involved in cell signaling controlling cell proliferation (Heule et al., 2014). *gsdf* (gonadal soma derived growth factor) is a downstream gene of *dmY* in the SD cascade that has taken the role of master SDg in *Oryzias curvinotus* (Myosho et al., 2012). *amhy* (Y chromosome-specific anti-müllerian hormone) is expressed in the presumptive Sertoli cells of XY males of *Odonthestes hatcheri* during the onset and subsequent GD. This gene has been inserted upstream of *amh* in the cascade of male development, becoming a male SDg (Hattori et al., 2012). In *T. rubripes*, *amhr2* (anti-müllerian hormone receptor type 2) is expressed in somatic cells surrounding germ cells and it is thought to be the SDg in this species. This gene contains a specific SNP variant in the kinase domain of *amhr2* in the X chromosome which determines lower affinity for the amh hormone, thus fating the female pathway when homozygous (XX; Kamiya et al., 2012). Finally, *sdY* (sexually dimorphic on the Y chromosome) is linked to the SEX locus of salmonids and is necessary and sufficient to induce testicular differentiation. It has evolved through neofunctionalization from *irf9* by losing its role in *ifn* signaling pathway and acquiring a new role in SD (Yano et al., 2012).

From an evolutionary perspective different SDgs have been reported either in closely related species like *O. latipes* (*dmY*) and *Oryzias luzonensis* (*gsdf1*; Myosho et al., 2012) or in divergent ones such as rainbow trout (*Oncorhynchus mykiss*; *sdY*; Yano et al., 2012), fugu (*amhr2*; Kamiya et al., 2012), or pejerrey (*O. hatcheri*; *amhY*; Hattori et al., 2012). Also, the same SDg has been identified in closely related species like *O. latipes* and *O. curvinotus* (*dmY*; Matsuda et al., 2002); *T. rubripes*, *Takifugu pardalis*, and *Takifugu poecilonotus* (*amhr2*; Kamiya et al., 2012); and in most salmonid species studied so far (*sdY*; Yano et al., 2013). Very recently, *dmrt1* has been suggested as a strong candidate in the half-smooth tongue sole (*Cynoglossus semilaevis*) based on its association with sex and its pseudogenization in the W chromosome (Chen et al., 2014). This would constitute the first SDg reported in a ZW system species and within flatfish, a group of species of great commercial value and with a particular metamorphosis to adapt to demersal life. However, no functional demonstration has been reported to date thus requiring further investigation.

Additional information exists from marker and QTL sex-associated studies which are usually unraveling variation on major genes controlling the fate of the undifferentiated primordium, and thus, basically related to SD. The relationships between the different genomic regions identified through this approach has been established through comparative mapping using model fish as a bridge, taking advantage of the conserved macrosynteny pattern observed in teleosts (Kai et al., 2011; Bouza et al., 2012). In most fish groups analyzed, these SD regions demonstrated to be non-homologous (Piferrer et al., 2012), so in the *Oryzias* (medaka) genus up to five different genes/genomic regions seem to be involved in sex determination, and only two species among

the six analyzed show the same SDg (Tanaka et al., 2007); in the Gasterosteidae (stickleback) family four different genomic regions have been identified in the five species analyzed, and in the fifth, a fusion of two previously reported SD chromosomes gave rise to the sexual chromosome (Ross et al., 2009); in the tilapiine (tilapia) cichlid tribe two chromosomes have been identified [linkage group 1 (LG1) and LG3], two species associated with LG1, other two with LG3, and the remaining two showing both linkage groups, but other studies also demonstrated association with LG23 (Cnaani, 2013); in the Salmoniformes (salmonids) order most species showed different non-homologous sex-associated genomic regions (Phillips et al., 2001); and finally, in the Poeciliidae (guppy and platyfish) family, up to four different chromosomes are involved in SD (Tripathi et al., 2009). Very likely these non-homologous regions include different SDgs, although in salmonid species showing non-homologous SD genomic regions, the SDg appears to be the same (Yano et al., 2013).

The diversity of SDg in fish highlights the many options available at the undifferentiated stage of gonads to switch and drive gonad fate, although some genes have been recurrently used because of their prominent position in the development cascade (Graves and Peichel, 2010; Heule et al., 2014). Among them, *dmrt1* and related genes found in medaka and likely in half-smooth tongue sole have also been reported in different vertebrates, including birds and amphibians (Smith and Sinclair, 2004; Yoshimoto et al., 2008), which illustrate processes of convergent evolution related to the suitability of some genes acting at the beginning of development. Furthermore, a great plasticity has been shown, since it can be found in XY or ZW systems and acting on a presence/absence model (medaka and frog) or in a dose-dependent manner (birds and likely tongue sole; Koopman, 2009; Chen et al., 2014). Three other genes, *gsdf1*, *amh1*, and *amhr2*, have also been reported to be activated at the beginning of development in the male pathway, and thus their SD role fits to the top of the GD invoked by theory (Heule et al., 2014). Contrary to previous findings in vertebrates (*sry* and *dmrt*-derived genes), *gsdf1*, *amh1*, and *amhr2* are not transcription factors being involved in cell signaling controlling cell proliferation (Heule et al., 2014). Finally, the *sdY* gene, the major SD factor in rainbow trout, represents an unexpected SDg of unknown function, whose carboxy-terminal extreme is homologous to an interferon-related gene, thus exemplifying the vast source of genes available for leading the SD process (Yano et al., 2012).

In addition to species with a single major SDg, many fish have shown more than one gene of big effect involved in sex determination. Indeed, in the most investigated fish groups at least two major genes (or genomic regions) related to SD have been reported in the same species: within tilapiinid, two major male and female determinant genes on LG1 and LG3, respectively, (Cnaani et al., 2008); in the platyfish, a poeciliid species, a multifactorial SD system with X, Y, and W sex chromosomes (Schultheis et al., 2009); in the sticklebacks, *Gasterosteus weathlandi* a multiple chromosome system due to the fusion of two major SD chromosomes presented in other species of Gasterosteidae (Ross et al., 2009); and in cichlids from lake Malawi, two main SD systems on LG5 and LG7, the first one representing a female epistatically dominant factor (Ser et al., 2010). Finally, a polygenic SD system has also been documented in

other species like European sea bass (Vandeputte et al., 2007) and zebrafish (Liew et al., 2012)

The lowering cost of new generation sequencing (NGS) methodologies will allow obtaining much more information in the near future to get a more accurate picture of SD in fish. Restriction-site associated DNA (RAD) sequencing, a technique which combines the powerful of NGS with the simplification of genomes through restriction enzyme digestion (Baird et al., 2008; Davey et al., 2011), is enabling to perform dense genomic screening to study the genetic architecture of multi-gene traits (Hohenlohe et al., 2010). This methodology is being applied for the study of SD in zebrafish (Anderson et al., 2012) and to identify sex-associated genomic regions in Nile tilapia (*O. niloticus*; Palaikostas et al., 2013a) and Atlantic halibut (*Hippoglossus hippoglossus*; Palaikostas et al., 2013b) for its application in aquaculture production. In addition, the high capacity of RAD sequencing for SNP discovery and constructing genetic maps will aid to get dense maps at candidate regions to narrow them and facilitate the identification SDgs (Taboada et al., 2014).

Gene expression studies linked to NGS methodologies will also be essential to understand the relationship between morphogenetic effects and the underlying genetic network. Microarrays have been used as a powerful tool for assessing gene expression profiles along GD (Gardner et al., 2012; Sreenivasan et al., 2014), but more recently, RNA-Seq is being applied due to its higher sensitivity, accuracy, and also because it provides additional information on genetic variants linked to expression differences (Sun et al., 2013; Tao et al., 2013). Finally, the evaluation of the pattern of methylation through bisulfite sequencing and other methodologies is providing quick genomic evaluation of the epigenomic maps along development (Cokus et al., 2008) constituting a valuable tool for understanding the regulation of SD and gonadogenesis (Piferrer, 2013).

ENVIRONMENTAL FACTORS ON SEX DETERMINATION

In environmental sex determination (ESD), the first difference between the two sexes is established by differences in the value of an environmental factor. ESD has been well studied in reptiles like crocodiles and turtles (Valenzuela and Lance, 2004). In *Menidia menidia* (Conover and Kynard, 1981), the first fish species where temperature-dependent sex determination (TSD), a form of ESD, was first described, some populations have shown a genetic component underlying SD (Conover and Heins, 1987). Much of the literature in this field demonstrated the influence of environmental factors in laboratory conditions, extreme in some cases, but not necessarily reflecting the conditions that species experience in nature (Ospina-Álvarez and Piferrer, 2008). Nevertheless, the presence of TSD in fish demonstrates the plasticity of gonad development (Baroiller et al., 2009).

Social interactions represent important environmental cues for SD in hermaphroditic species (Godwin, 2009). In these cases, the brain has shown to be a major player translating social cues into a physiological signal (Kobayashi et al., 2010). Although this is not a mechanism driving the gonad fate at the beginning of development, it is a good example of the plasticity of gonad development in fish and also an evidence on the existence of bipotential

primordium cells in the differentiated gonads of adult fish (Zhou and Gui, 2010).

Temperature is the environmental factor with highest influence in SD in fish (Ospina-Álvarez and Piferrer, 2008; Baroiller et al., 2009). Temperature influence on sex shifting can be exerted at several points of the differentiation cascade. High temperatures usually tend to produce more males and low temperatures have no effects or produce more females in some cases (Ospina-Álvarez and Piferrer, 2008). The ultimate mechanism (if there were a single one) connecting temperature and sex ratio is not known, and several have been proposed. The influence of temperature on SD has been related to a higher stress, giving rise to changes in circulating cortisol levels. In fact, the administration of cortisol in the diet has demonstrated a significant influence on sex ratio (Mankiewicz et al., 2013). However, it is far from clear the ultimate molecular mechanism connecting cortisol levels and masculinization. On one hand, Hayashi et al. (2010) proposed a direct up-regulation of the follicle stimulant hormone (FSH) receptor, which would be connected to germ cell proliferation. On the other, Fernandino et al. (2013) suggested an up-regulation of *hsd11b2*, a steroidogenic enzyme implicated both in the metabolism of cortisol into cortisone, and in the synthesis of biologically active androgens such as 11-ketotestosterone. Epigenetic regulation of aromatase expression mediated by temperature has also been proposed. Thus, Navarro-Martín et al. (2011) demonstrated that hypermethylation of aromatase promoter is correlated with high temperature during the thermosensitive period in the European sea bass, strongly suggesting that sex differentiation is under epigenetic control in this species.

GENETIC VARIATION WITHIN SPECIES: SEX DETERMINATION AS A COMPLEX TRAIT

Most studies in SD in fish have been focused on identifying the master SDg expected according to previous SD models. However, new data are consistently showing that other minor genetic and environmental factors are also involved in SD, even in species with well studied master genes like *dmY* of medaka (Matsuda et al., 2002), and thus, more effort should be devoted to investigate this variation to get the closest picture as possible on SD in fish.

The information exposed so far shows that: (i) although environmental factors may influence sex ratio, usually the genetic component represent the main SD factor in most studied species; (ii) major genes, those which explain a high proportion of the trait phenotypic variance, are on the basis of SD in many species, likely an expected fact, because of the hierarchical nature of the sex development pathway; and (iii) the interaction of other genes involved in SD with the major factors and the environment. Although major genes are involved in SD of many fish species, available data suggest that SD studies and their applications for fish aquaculture should emphasize the complex nature of this trait and thus, using appropriate quantitative genetics tools for its study.

In fact, consistent variation among families has been reported on sex ratio in European sea bass (Vandeputte et al., 2007), turbot (Haffray et al., 2009; Martínez et al., 2009), Nile tilapia (Lozano et al., 2013) and zebrafish (Liew et al., 2012). In some species, the additive genetic component underlying SD or sex ratio was even estimated (Vandeputte et al., 2007; Lozano et al., 2013). In

zebrafish, a polygenic SD system was suggested based on inter-family and inter-strain variation, and consequently several QTL at different genomic locations were identified, some of them associated to the different strains studied (Bradley et al., 2011; Anderson et al., 2012; Howe et al., 2013). In other species where major loci are involved, the application of QTL screening or genomic association analysis to look for the SD region demonstrated to be efficient and new SD-related genomic regions were identified (Martínez et al., 2009; Hermida et al., 2013), sometimes denoting important intraspecific variation such as in *Eigenmannia* or in cichlid species complex (Ser et al., 2010; Henning et al., 2011; Parnell and Streelman, 2013). As a consequence, selection has demonstrated to be efficient to change sex ratio in progenies of several species and in other related traits like the sensitivity to temperature on sex ratio (Baroiller et al., 2009; Liew et al., 2012; Lüthmann et al., 2012; Lozano et al., 2013).

SEX-ASSOCIATED TRAITS IN FINFISH AQUACULTURE: SOME RELEVANT EXAMPLES

In many fish species one sex grows faster or matures earlier than the other and these differences may be accentuated under aquaculture conditions (Breder and Rosen, 1966; Parker, 1992; Piferrer et al., 2012). Sex-associated growth differences generate size dispersion, and therefore, classification must be performed for feeding and to avoid cannibalism or size hierarchies affecting social relations (Dou et al., 2004). This represents more work in animal husbandry, and a higher number of production units to adjust different growth groups (Piferrer et al., 2012). Sexual growth dimorphism can favor males (e.g., tilapias) or, more frequently, females (flatfish, sea bass, among others). In some cases, as in the turbot, females can be 50% larger than males (Imstrand et al., 1997). In other cases, as in the European sea bass, the rearing conditions result in highly male-biased stocks (Piferrer et al., 2005). A great deal of research towards the development of sex control methods has been carried out in fish (Piferrer, 2001; Cnaani and Levavi-Sivan, 2009).

Sex-associated markers are very useful in this context for precocious sex identification, especially in those species lacking morphological sexual dimorphism. This can aid to identify the sex of potential broods in genetic breeding programs and to avoid sex bias in the selected population. However, the most relevant application of sex-associated markers is to identify the genetic sex of sex-reversed individuals after hormonal treatment to accelerate the processes for establishing monosex populations (Penman and Piferrer, 2008). The availability of sex-associated markers or even better the SD master gene makes it possible to shorten this process using marker assisted selection (MAS) or gene assisted selection (GAS), respectively.

TURBOT

The strong sexual growth dimorphism of turbot has promoted the interest of industry for all-female populations. No sex-associated karyotype heteromorphism have been detected in turbot, either after analyzing the mitotic or the 11-fold longer and higher resolution meiotic chromosomes (Bouza et al., 1994; Cuiñado et al., 2002). This suggests that the SD region in turbot is small or not large enough to be detected with these cytogenetic techniques. A QTL screening performed with 100 homogeneously distributed

microsatellites identified a major SD region in the proximal region of LG5 between two markers separated by 17.4 cM (Martínez et al., 2009). Assuming a single SD region with full penetrance, the SD master gene (SDg) was located at 2.6 cM of the Sma-USC30 microsatellite locus, representing 1.4 Mb according to the general relationship between genetic and physical maps in turbot (Bouza et al., 2007). Also, the analysis of segregation of Sma-USC30 in all families demonstrated that the mother is responsible for sex, supporting a ZZ/ZW system (Martínez et al., 2009) in accordance with the sex ratios observed in progenies from hormonal sex reversed parents (Haffray et al., 2009). However, sex association of Sma-USC30 showed variation among families (between 84 and 100%) and, in addition, other minor QTL were detected at LG6, LG8, and LG21. Temperature also showed some influence on sex ratios (Haffray et al., 2009), although without the general trend reported in most species where the proportion of males increased with temperature (Ospina-Álvarez and Piferrer, 2008).

Using the Sma-USC30 marker, it was possible to classify correctly 98.4% of the individuals in four out of five families analyzed. This information was essential to develop a molecular tool for precocious sex identification in turbot, currently under a Spanish patent (Ref. number: 2 354 343). Since sex cannot be identified in turbot until fish maturation, precocious sex identification is relevant in breeding programs to estimate sex ratio in selected progenies. This molecular tool is also essential to facilitate the achievement of all-female populations. Because turbot displays a ZW mechanism, getting all-female populations requires a three-generation pedigree starting from hormonal sex-reversed ZW neomales until obtaining WW superfemales in the progeny of the second cross (Figure 3). These superfemales would produce all-female offspring after being crossed with normal ZZ males. However, the chromosome constitution of ZW neomales or WW superfemales require individual progeny testing of

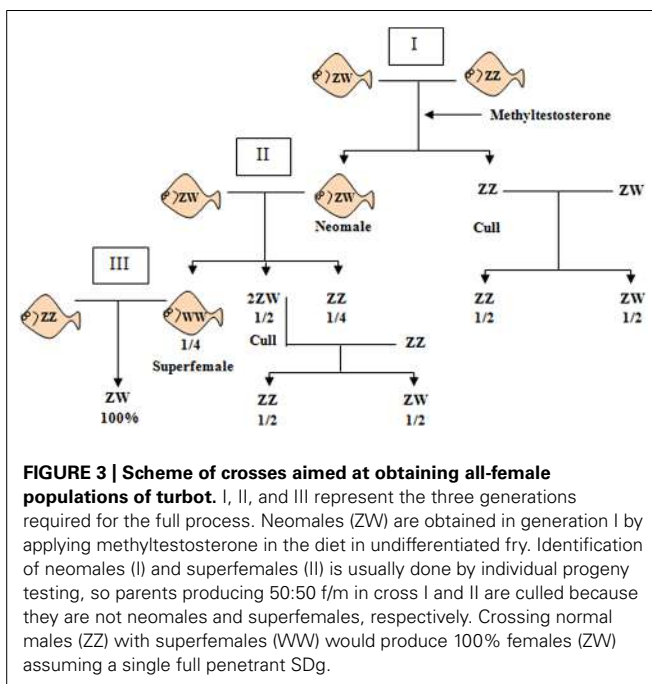
hormone-treated larvae (ZZ or ZW) and of female offspring in cross II (ZW or WW).

Progeny testing is long and involves at least 2–3 years until fish are mature for checking sex ratio in their progenies. Additionally, sex can only be visually identified 4–6 months after hatching in sacrificed offspring, which globally represents a minimum of 2.5 years for progeny testing per generation. The availability of a genetic marker closely linked to the SDg has enabled assessment of the genotypic sex of fish just after 4–6 months when a fin clip can be obtained without damage, thus saving a minimum of 5 years for the production of all-female progenies. However, some limitations still remain, so this technology should be further refined. First, production of 100% females is not often fulfilled because turbot sex also depends on other minor genetic and environmental factors. In addition, because Sma-USC30 is a linked marker to the SDg, we need to establish association of SmaUSC-E30 to sex at family level because this is not a sex-specific marker. Finally, because crossovers can take place between the marker and the SDg, some of the selected ZW neomales or WW females would not show the expected genetic constitution. Nevertheless, this tool is being used by turbot companies with encouraging results.

The availability of a denser genetic map with around 600 markers (Bouza et al., 2012; Hermida et al., 2013), the enriched database with reproduction and immune genes (Pereiro et al., 2012; Ribas et al., 2013), and the recently assembled turbot genome (Figueras et al., in preparation) is enabling a more refined analysis of the SD region to identify the SDg, to analyze its relationship with the previous suggestive QTL, and to study the evolution of sex chromosomes (Taboada et al., 2014). The fine mapping performed has narrowed the genomic position of the SDg to a few kilobases and now much more genetic markers closely associated to the SDg are available, thus facilitating the precocious evaluation of sex. However, although several strong candidates were identified in that region (*sox2*, *dnajc19*, and *fxr1*), none of them were associated with sex at the species level, which illustrates the difficulty of such enterprise (Hermida et al., 2013). In addition, this work has provided additional support for the existence of minor factors at LG6, LG8, and LG21 in new families and demonstrated an interactive rather than an additive component between minor and the major SD QTL.

EUROPEAN SEA BASS

The European sea bass is a gonochoristic marine teleost of the family Moronidae that is present in the NE Atlantic, from Norway to NW Africa and in the Mediterranean up to the Black Sea. As in most teleosts, sea bass does not show sex chromosomes (Devlin and Nagahama, 2002). Female homogamety (XX) has been ruled out since sex ratios of normal diploid and gynogenetic offspring are equivalent (Felip et al., 2002), and offspring from masculinized females is not female-biased (Blázquez et al., 1999). Other data from hormone-treated fish suggested that male homogamety with environmentally male-biased sex ratio would still be a possibility (Vandeputte et al., 2007). However, a polygenic sex determining system (Vandeputte et al., 2007) influenced by temperature (Piferrer et al., 2005) seems to fit better to data. Sex ratio shows interfamily differences with specific and measurable paternal and maternal components and at least two genes would



be necessary to explain the variation pattern observed (Vandeputte et al., 2007), although no sex-associated QTL screening was carried out. The temperature can greatly influence European sea bass sex ratios (Piferrer et al., 2005) and a genetic component determining differential sensitivity to the masculinizing power of temperature has been reported (Saillant et al., 2002). Variation found on female percent due to thermal treatments show that genetic and environmental components are of comparable magnitude, supporting the notion of a continuum between GSD and ESD components of SD in the European sea bass.

In the European sea bass, gonads remain undifferentiated during post-larval stages until fish attain about 8 cm standard length, usually at about 5–6 months of age (Bruslé and Roblin, 1984; Blázquez et al., 1999; Navarro-Martín et al., 2009a; Díaz et al., 2013). European sea bass females differentiate earlier, are bigger, and mature later than males (Blázquez et al., 1999; Navarro-Martín et al., 2009a; Díaz et al., 2013). While still sexually undifferentiated, European sea bass gonads can be influenced by environmental abiotic factors or external factors such as sex steroids (Navarro-Martín et al., 2009a), but once sex is determined remains throughout life (Zanuy et al., 2001).

One of the main genes involved in fish GD as outlined above is the aromatase (cytochrome P450). This gene is present in two paralogous copies as a result of the teleost genome duplication, one predominantly expressed in the ovary (*cyp19a1a*; Dalla Valle et al., 2002) and the other in the brain (*cyp19a1b*; Blázquez and Piferrer, 2004). Interestingly, the *cyp19a1a* promoter exhibits important conserved binding sites for several genes of the GD network such as *sfl*, *sox*, *foxl2* or *ar*.

Recent work by Navarro-Martín et al. (2011) showed how temperature during early development is linked to the production of male-biased populations through differences in the methylation levels specifically on the gonadal aromatase promoter at one year. Different CpGs loci within the *cyp19a1a* promoter showed different sensitivities to temperature, suggesting a different role on regulation of aromatase expression. Methylation of gonadal aromatase promoter is thought to be the cause of the lower expression of aromatase in the temperature-masculinized fish linked to the suppressed ability of *sfl* and *foxl2* transcription factors to stimulate gonadal aromatase expression (Navarro-Martín et al., 2011).

While 15°C has been proposed as the optimal temperature for larval rearing in European sea bass (Koumoundouros et al., 2001), optimal growth in juveniles is found at 26°C and 13°C is considered as detrimental (Person-Le Ruyet et al., 2004). European sea bass hatcheries usually apply high temperatures (>20°C) to speed up growth rates, but male-bias progenies determine a loss of biomass. In the European sea bass the thermosensitive period includes from half-epiboly to mid-metamorphosis (~17–18 mm total length; ~70 dph; Koumoundouros et al., 2002), and treating fish with high temperatures masculinize a high proportion of the population (Navarro-Martín et al., 2009b), while temperatures never surpassing 17°C until metamorphosis yielded the maximum female proportion (Pavlidis et al., 2000). However, raising fish at low temperatures (15°C) for a long period also masculinize the population even more than a high temperature thermal treatment (Saillant et al., 2002). A thermal protocol to maximize the number of females was developed (Navarro-Martín et al., 2009a)

and recently patented (patent no N200802927). Such protocol consisted on maintaining 17°C water temperature until the end of the thermosensitive period, and then increasing the temperature as a ratio of 0.5°C/day until 21°C to allow high growth rates. It should be stated that there is no temperature regime that will increase the proportion of females in the European sea bass. Instead, proper management of temperature during early development will avoid induced masculinization by applying high water temperatures before the thermosensitive period is over. What a proper thermal regime does is allowing the production of as many females as the polygenic system will allow, bearing in mind the sex ratio of a given brood, even reared at the optimal thermal regime, will ultimately depend on the genetic constitution of both parents.

With this information at hand, different strategies have been devised by European sea bass industry to improve growth rate. On one hand, classical breeding programs may incorporate sex ratio as additional phenotypic information to get female-biased progenies. This strategy would be much more efficient if sex-related QTL screening were performed and sex-associated genetic markers explaining an important proportion of trait variance were incorporated following MAS selection programs. On the other hand, controlling sex-ratio through larval rearing temperature protocols could increase female proportions by avoiding masculinization due to elevated water temperature. Currently, efforts are underway aimed at selecting broodstock that will produce the highest number of females and investigating the possibility that, among these fish, those that are more resistant to the masculinizing temperature can also be selected. An additional step would be further investigation of the epigenetic mechanisms responsible for the inheritance of the high-temperature masculinization, as already done in the half-smooth tongue sole (Chen et al., 2014). In the case of the European sea bass, the goal would be the opposite, i.e., to select as future broodstock those fish that despite being reared at high (masculinizing temperature) they do not become masculinized. In this way, the production of the highest number of females across different generations perhaps could be achieved.

SALMONIDS

The Salmonidae family (11 genera, 70 species; salmon, trout, char, whitefishes, and graylings) includes several of the most economically important species for aquaculture and fisheries industry (the third largest world fish production). Salmonids are worldwide distributed and some species have played a vital role in the life and culture of the North hemisphere societies for thousands of years.

Atlantic salmon (*Salmo salar*) is the leading intensively farmed marine fish. In 1998 global farmed production exceeded the world's total wild salmon captures, and in 2010 around 1.2 million tons were produced worldwide (Food and Agriculture Organization [FAO], 2014). Atlantic salmon breeding companies have achieved more than 100% growth increase in around six generations of selection, and significant improvement in disease resistance and delay at the onset of sexual maturation. The vast majority of farmed Atlantic salmon eggs and smolts are now sourced by such breeding companies (Bostock et al., 2010). Another important salmonid species, the rainbow-trout (*O. mykiss*), is the most-widely cultivated cold fresh water fish species in the world.

In salmonids, early maturation occurs differentially in males and females and is responsible for some problems related to intensive culture (Felip et al., 2006) including reducing growth, increasing disease susceptibility and changing of organoleptic properties. Thus, in this group, females are preferred for production. In all salmonid species investigated so far, SD is strictly genetic with a male heterogametic sex-determination system (Davidson et al., 2009), although no heteromorphisms or only slight morphological differentiation have been reported associated to sex chromosomes in some species. In rainbow trout, all-female production is generalized in Europe since females are still immature at harvest. Neomales (hormone sex-reversed genotypic females) are used to produce all-female progenies from XX neomale crossed to XX females. Triploids are also produced using chromosome set manipulation techniques to avoid sexual maturation for the production of individuals of bigger size (more than 3 kg; Piferrer et al., 2009).

Both, endocrine and genetic technologies have been implemented for sex control on a production scale in salmonids. Nevertheless, the phenotypic differentiation of males and females is still problematic until the fish become sexually mature and, thus, in some species it is still necessary a genetic/molecular test for sexing. Sex-specific markers, including a linked sequence to the growth hormone pseudogene, have been developed in the last decade for Pacific species of the *Oncorhynchus* genus, rainbow trout (*O. mykiss*), Chinook salmon (*O. tshawytscha*), coho salmon (*O. kisutch*), chum salmon (*O. keta*), and pink salmon (*O. gorbuscha*; Brunelli and Thorgaard, 2004).

Recently, the *sdY* (sexually dimorphic on the Y chromosome) was identified as a male-specific linked gene on the Y chromosome in most salmonids (Salmoninae, Coregoninae and Thymallinae subfamilies), strongly suggesting that *sdY* may be the conserved master sex-determining gene of this group (Yano et al., 2012, 2013). However, because this gene is not located at homologous genomic positions among the different salmonid species, it has been suggested its jumping associated to mobile elements (Yano et al., 2013). Irrespective of its location, sequences of this gene may represent a useful tool for sexing. However, some exceptions were observed to this general rule, and in the Coregoninae subfamily, while *Stenodus leucichthys* showed *sdY* as a male specific gene, *Coregonus lavaretus* and *Coregonus clupeaformis*, both males and females contain the *sdY* gene, so a different SD mechanism appears to be involved.

The analysis of sex-associated markers in several families of the SALTAS Tasmanian Atlantic salmon program from different male lineages allowed the discovery of three sex-associated markers (Ssa03, Ssa06, and Ssa2) mapping at three different linkage groups. Ssa2 is the same sex-associated marker previously reported in Atlantic European populations (Eisbrenner et al., 2014), but Ssa03 and Ssa06 represent new genomic positions. These three loci showed positive amplification for *sdY* gene and most individuals analyzed showed a good concordance between phenotypic sex and *sdY* PCR amplification, suggesting the movement of *sdY* to new positions within this species. However, some inconsistencies were detected among the sex marker associated genotypes, the presence of *sdY* gene and the phenotypic sex. Based on these findings, and in the fact that the *sdY* protein lacks a DNA binding

domain, these authors suggested the existence of another sex determining gene in Atlantic salmon upstream to *sdY*. Also, they emphasized the importance of using many families to identify sex-associated markers in salmonid species (Eisbrenner et al., 2014). Furthermore, the recent characterization of the male-specific region on the Y chromosome of rainbow trout which contains the *sdY* gene and the male specific marker (OmY1), revealed several male specific SNPs associated with 12 single-copy protein coding sequences whose role in the SD should be further analyzed (Phillips, 2013). Data show that even in a species with an apparent well-established SD mechanism variation is observed. Thus, caution should be taken when applying sex-associated markers for establishing associations and also to for their application to precocious sex determination.

TILAPIA

The tribe Tilapiini includes more than 80 species of cichlid fish (family Cichlidae, order Perciformes). Tilapias are endemic to Africa and the Middle East, but they have been introduced into most tropical and subtropical countries for aquatic weed control and aquaculture. Tilapia culture was considered a resource to improve protein supply in developing countries, but nowadays there is also an important market for tilapias in Japan, United States, European Union as well as in other developed countries. Tilapia is one of the fastest growing fish farming sector, being China the production leader, and it constitutes the second most important group of farmed fish after carps and the most widely grown of any farmed fish (85 producer countries). The main aquaculture species is the Nile tilapia (*O. niloticus*) with a production exceeding 3.2 million metric tons in 2012 (Food and Agriculture Organization [FAO], 2014).

Tilapias exhibit an important sexual growth dimorphism in favor of males. Additionally, tilapias show early maturation (e.g., 4–5 months old in Nile tilapia), which determines successive spawning during the growing period, leading to stunting of growth (Beardmore et al., 2001). All these circumstances make it difficult to establish a uniform product, and all male fry are preferred.

Monosex culture can be obtained by different approaches: manual separation of males and females; hybridization between species to produce all-male offspring; or artificial sex reversal using hormones. The most frequent method for producing all-male populations in tilapia was the treatment with 17 α -methyltestosterone included in the diet of sexually undifferentiated fry. If properly applied, farms can produce male populations with 98 to 100% effectiveness. However, marketing of hormonally treated fish can also be a problem for health and the direct use of hormones is usually forbidden by food safety regulations in European Community, although in other countries it may be allowed. One way to get through this problem is to combine a method for sex reversal with a breeding scheme aimed at obtaining broodstock that produces monosex fry following the reverse procedure as outlined before for turbot (Figure 3). In the case of Nile tilapia, it is necessary to produce YY supermales by crossing neofemales (XY) with regular males (XY). As in turbot, the use of sex-linked DNA markers could shorten the process by distinguishing XX, XY, or YY individuals, thus avoiding the identification of individual genotypes by progeny testing.

Sex chromosomes are not identifiable in tilapias using standard cytogenetic techniques. Most species show 22 chromosome pairs but there is no a heteromorphic sexual chromosome pair. Association studies suggested that sex is determined in tilapias by the existence of major genes located at linkage groups 1, 3, and 23 in the different species (Cnaani et al., 2008; Cnaani, 2013). Additionally, the heterogametic sex can be either the male or the female, depending upon the species, and ZZ/ZW (LG3) and XX/XY (LG1) systems have been reported within the same genus, i.e., in *O. niloticus* and *Tilapia zillii* (XX/XY), and in *Tilapia mariae*, *Oreochromis aureus*, *Oreochromis karongae*, *Oreochromis tanganyicae* (ZZ/ZW). Some families of blue tilapia have been found segregating for both loci, and in these cases the ZW locus appears to be epistatic over the XY (ZW/XY individuals are female; Lee et al., 2004). In Mozambique tilapia (*Oreochromis mossambicus*), the SD locus was found at LG1 (Liu et al., 2013). However, Cnaani et al. (2008) found sex-associated markers on LG1 and LG3 in three families of this species. These discrepancies may be determined by the different genetic background of families and strains used in these studies (Liu et al., 2013). The differences in SD within species show that minor genetic factors segregate and interact with major genes in addition to the influence of environment factors (Baroiller et al., 2009), suggesting that SD should be treated as a quantitative trait and its dissection approached using QTL screening (Eshel et al., 2012).

The main cultured species is the Nile tilapia (GIFT project; Lozano et al., 2013). Most data pointed to LG1 as the sex chromosome in this species (Cnaani et al., 2008) and recently a small candidate region was narrowed by RAD (restriction site associated DNA) sequencing on a 1–2 Mb region on LG1 (Palaiokostas et al., 2013a). However, other sex-linked markers have been identified in *O. niloticus* and its hybrids (*O. niloticus* × *O. aureus*) mapping on LG3 and LG23 (Eshel et al., 2012). Also, it has been shown by linkage analysis that genetic factors are involved in the sensitivity of SD to temperature and that these factors are located depending on families on the same chromosomal regions as the major QTL at LG1, LG3, and LG23 (Lühmann et al., 2012).

The closely sex-associated genomic region identified on LG1 includes 10 genes not previously related to SD in other species, and two SNPs probed to be very useful for sexing individuals, thus being worthy for production all male stocks by industry (Palaiokostas et al., 2013a). However, some of these markers were not associated to sex in different strains or families, so checking several markers on different linkage groups should be done before its application. Most studies on SD in *O. niloticus* have been carried out on fish derived from Lake Manzala in Egypt population, but today this species show a worldwide production and new information is arising which probably will confirm the necessity for checking a set of markers previous to sexing specific strains.

CONCLUSION

Major genetic factors can explain a high proportion of the SD variance in fish in accordance with the hierarchical gonad development of vertebrates and with the models proposed to explain its origin and evolution. However, several other minor genetic and environmental factors also influence sex following a complex

interactive pattern. Thus, the currently available information supports the idea that sex can be regarded as a complex trait in fish, with the influence of one or more genetic factors in addition to possible environmental influences, depending upon the species. The presence of genetic factors regardless of whether SD is under the control of a master gene, a polygenic system or driven by an environmental factor enables their application in MAS programs to exploit the benefits of a particular sex.

ACKNOWLEDGMENTS

This research work was supported by the Spanish Government (Consolider Ingenio Aquagenomics: CSD2007-00002 project) and Spanish Ministerio de Ciencia e Innovación (AGL2009-13273 and AGL2010-15939) projects to Paulino Martínez and Francesc Piferrer.

REFERENCES

- Anderson, J. L., Mari, A. R., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., et al. (2012). Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS ONE* 7:e40701. doi: 10.1371/journal.pone.0040701
- Avise, J. C., and Mank, J. E. (2009). Evolutionary perspectives on hermaphroditism in fishes. *Sex. Dev.* 3, 152–163. doi: 10.1159/000223079
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., et al. (2014). Sex determination: why so many ways of doing it?. *PLoS Biol.* 12:e1001899. doi: 10.1371/journal.pbio.1001899
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376. doi: 10.1371/journal.pone.0003376
- Baroiller, J. F., D'Cotta, H., and Saillant, E. (2009). Environmental effects on fish sex determination and differentiation. *Sex. Dev.* 3, 118–135. doi: 10.1159/000223077
- Beardmore, J. A., Mair, G. C., and Lewis, R. I. (2001). Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* 197, 283–301. doi: 10.1016/S0044-8486(01)00590-7
- Berbejillo, J., Martínez-Bengochea, A., Bedo, G., Brunet, F., Volff, J. N., and Vizziano-Cantonnet, D. (2012). Expression and phylogeny of candidate genes for sex differentiation in a primitive fish species, the Siberian sturgeon, *Acipenser baerii*. *Mol. Reprod. Dev.* 79, 504–516. doi: 10.1002/mrd.22053
- Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24, 94–102. doi: 10.1016/j.tree.2008.09.010
- Blázquez, M., Carrillo, M., Zanuy, S., and Piferrer, F. (1999). Sex ratios in offspring of sex-reversed sea bass and the relationship between growth and phenotypic sex differentiation. *J. Fish Biol.* 55, 916–930. doi: 10.1111/j.1095-8649.1999.tb00730.x
- Blázquez, M., and Piferrer, F. (2004). Cloning, sequence analysis, tissue distribution, and sex-specific expression of the neural form of P450 aromatase in juvenile sea bass (*Dicentrarchus labrax*). *Mol. Cell Endocrinol.* 219, 83–94. doi: 10.1016/j.mce.2004.01.006
- Böhne, A., Heule, C., Boileau, N., and Salzburger, W. (2013). Expression and sequence evolution of aromatase cyp19a1 and other sexual development genes in East African cichlid fishes. *Mol. Biol. Evol.* 30, 2268–2285. doi: 10.1093/molbev/mst124
- Bostock, J., McAndrew, B., Richards, R., Jauncey, K., Telfer, T., Lorenzen, K., et al. (2010). Aquaculture: global status and trends. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2897–2912. doi: 10.1098/rstb.2010.0170
- Bouza, C., Hermida, M., Pardo, B. G., Fernández, C., Fortes, G., Castro, J., et al. (2007). A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics* 177, 2457–2467. doi: 10.1534/genetics.107.075416
- Bouza, C., Hermida, M., Pardo, B. G., Vera, M., Fernández, C., de la Herrán, R., et al. (2012). An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful framework for comparative genomics across model and farmed teleosts. *BMC Genet.* 13:54. doi: 10.1186/1471-2156-13-54
- Bouza, C., Sánchez, L., and Martínez, P. (1994). Karyotypic characterization of turbot (*Scophthalmus maximus*) with conventional, fluorochrome, and restriction endonuclease banding techniques. *Mar. Biol.* 120, 609–613. doi: 10.1007/BF00350082

- Bradley, K. M., Breyer, J. P., Melville, D. B., Broman, K. W., Knapik, E. W., and Smith, J. R. (2011). An SNP-based linkage map for zebrafish reveals sex determination loci. *G3 (Bethesda)* 1, 3–9. doi: 10.1534/g3.111.000190
- Breder, C. M., and Rosen, D. E. (1966). *Modes of Reproduction in Fishes*. New York: Natural History Press.
- Brunelli, J. P., and Thorgaard, G. H. (2004). A new Y-chromosome-specific marker for Pacific salmon. *Trans. Am. Fish Soc.* 133, 1247–1253. doi: 10.1577/T03-049.1
- Bruslé, J., and Roblin, C. (1984). *Sexualité du loup Dicentrarchus Labrax en Condition d'élevage Contrôlé*. Paris: INRA.
- Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)* 95, 118–128. doi: 10.1038/sj.hdy.6800697
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46, 253–260. doi: 10.1038/ng.2890
- Cnaani, A. (2013). The tilapias' chromosomes influencing sex determination. *Cytogenet. Genome Res.* 141, 195–205. doi: 10.1159/000355304
- Cnaani, A., Lee, B. Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., et al. (2008). Genetics of sex determination in tilapia species. *Sex. Dev.* 2, 43–54. doi: 10.1159/000117718
- Cnaani, A., and Levavi-Sivan, B. (2009). Sexual development in fish, practical applications for aquaculture. *Sex. Dev.* 3, 164–175. doi: 10.1159/000223080
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745
- Conover, D. O., and Heins, W. W. (1987). Adaptive variation in environmental and genetic sex determination in a fish. *Nature* 326, 496–498. doi: 10.1038/326496a0
- Conover, D. O., and Kynard, B. E. (1981). Environmental sex determination, interaction of temperature and genotype in a fish. *Science* 213, 577–579. doi: 10.1126/science.213.4507.577
- Cuñado, N., Terrones, J., Sánchez, L., Martínez, P., and Santos, J. L. (2002). Sex-dependent synaptic behaviour in triploid turbot, *Scophthalmus maximus* (Pisces, *Scophthalmidae*). *Heredity* 89, 460–464. doi: 10.1038/sj.hdy.6800165
- Dalla Valle, L., Lunardi, L., Colombo, L., and Belvedere, P. (2002). European sea bass (*Dicentrarchus labrax* L.) cytochrome P450arom: cDNA cloning, expression and genomic organization. *J. Steroid Biochem. Mol. Biol.* 80, 25–34. doi: 10.1016/S0960-0760(01)00170-4
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012
- Davidson, W. S., Kuang, T. K., Fujiki, K., von Schalburg, K. R., and Koop, B. F. (2009). The sex-determining loci and sex chromosomes in the family *salmonidae*. *Sex. Dev.* 3, 78–87. doi: 10.1159/000223073
- Devlin, R. H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological and environmental influences. *Aquaculture* 208, 191–364. doi: 10.1016/S0044-8486(02)00057-1
- Díaz, N., Ribas, L., and Piferrer, F. (2013). The relationship between growth and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture* 408–409, 191–202. doi: 10.1016/j.aquaculture.2013.05.032
- Dou, S. Z., Masuda, R., Tanaka, M., and Tsukamoto, K. (2004). Size hierarchies affecting the social interactions and growth of juvenile Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* 233, 237–249. doi: 10.1016/j.aquaculture.2003.09.054
- Eisbrenner, W. D., Botwright, N., Cook, M., Davidson, E. A., Dominik, S., Elliott, N. G., et al. (2014). Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (*Salmo salar*). *Heredity (Edinb)* 113, 86–92. doi: 10.1038/hdy.2013.55
- Eshel, O., Shirak, A., Weller, J. I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3 (Bethesda)* 2, 35–42. doi: 10.1534/g3.111.001545
- Food and Agriculture Organization [FAO]. (2014). *Cultured Aquatic Species Information Programme. FAO Fisheries and Aquaculture Department*. Available at: <http://www.fao.org/fishery/culturedspecies>
- Felip, A., Piferrer, F., Carrillo, M., and Zanuy, S. (2002). Growth, gonadal development and sex ratios of meiogonogentic diploid sea bass. *J. Fish Biol.* 61, 347–359. doi: 10.1111/j.1095-8649.2002.tb01570.x
- Felip, A., Zanuy, S., and Carrillo, M. (2006). Comparative analysis of growth performance and sperm motility between precocious and non-precocious males in the European sea bass, *Dicentrarchus labrax*, L. *Aquaculture* 256, 570–578. doi: 10.1016/j.aquaculture.2006.02.014
- Fernandino, J. I., Hattori, R. S., Moreno Acosta, O. D., Strüssmann, C. A., and Somoza, G. M. (2013). Environmental stress-induced testis differentiation: androgen as a by-product of cortisol inactivation. *Gen. Comp. Endocrinol.* 192, 36–44. doi: 10.1016/j.ygcen.2013.05.024
- Fujimoto, T., Nishimura, T., Goto-Kazeto, R., Kawakami, Y., Yamaha, E., and Arai, K. (2010). Sexual dimorphism of gonadal structure and gene expression in germ cell-deficient loach, a teleost fish. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17211–17216. doi: 10.1073/pnas.1007032107
- Gardner, L. D., Jayasundara, N., Castilho, P. C., and Block, B. (2012). Microarray gene expression profiles from mature gonad tissues of Atlantic bluefin tuna, *Thunnus thynnus* in the Gulf of Mexico. *BMC Genomics* 13:530. doi: 10.1186/1471-2164-13-530
- Godwin, J. (2009). Social determination of sex in reef fishes. *Semin. Cell Dev. Biol.* 20, 264–270. doi: 10.1016/j.semcdb.2008.12.003
- Goto, R., Saito, T., Takeda, T., Fujimoto, T., Takagi, M., Arai, K., et al. (2012). Germ cells are not the primary factor for sexual fate determination in goldfish. *Dev. Biol.* 370, 98–109. doi: 10.1016/j.ydbio.2012.07.010
- Graves, J. A. M., and Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Gen. Biol.* 11:205. doi: 10.1186/gb-2010-11-4-205
- Haffray, P., Lebègue, E., Jeu, S., Guennoc, M., Guiguen, Y., Baroiller, J. F., et al. (2009). Genetic determination and temperature effects on turbot *Scophthalmus maximus* sex differentiation: an investigation using steroid sex-inverted males and females. *Aquaculture* 294, 30–36. doi: 10.1016/j.aquaculture.2009.05.004
- Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., et al. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2955–2959. doi: 10.1073/pnas.1018392109
- Hayashi, Y., Kobira, H., Yamaguchi, T., Shiraishi, E., Yazawa, T., Hirai, T., et al. (2010). High temperature causes masculinization of genetically female medaka by elevation of cortisol. *Mol. Reprod. Dev.* 77, 679–686. doi: 10.1002/mrd.21203
- Henning, F., Moyses, C. B., Calcagnotto, D., Meyer, A., and de Almeida-Toledo, L. F. (2011). Independent fusions and recent origins of sex chromosomes in the evolution and diversification of glass knife fishes (*Eigenmannia*). *Heredity (Edinb)* 106, 391–400. doi: 10.1038/hdy.2010.82
- Hermida, M., Bouza, C., Fernández, C., Sciarra, A. A., Rodríguez-Ramilo, S. T., Fernández, J., et al. (2013). Compilation of mapping resources in turbot (*Scophthalmus maximus*): a new integrated consensus genetic map. *Aquaculture* 414–415, 19–25. doi: 10.1016/j.aquaculture.2013.07.040
- Herpin, A., Adolff, M. C., Nicol, B., Hinzmann, M., Schmidt, C., Klughammer, J., et al. (2013). Divergent expression regulation of gonad development genes in medaka shows incomplete conservation of the downstream regulatory network of vertebrate sex determination. *Mol. Biol. Evol.* 30, 2328–2346. doi: 10.1093/molbev/mst130
- Heule, C., Salzburger, W., and Böhne, A. (2014). Genetics of sexual development: an evolutionary playground for fish. *Genetics* 196, 579–591. doi: 10.1534/genetics.114.161158
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genet.* 6:e1000862. doi: 10.1371/journal.pgen.1000862
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–503. doi: 10.1038/nature12111
- Imsland, A., Folkvord, A., Stefansson, O., and Taranger, L. (1997). Sexual dimorphism in growth and maturation of turbot, *Scophthalmus maximus* (Rafinesque, 1810). *Aquacult. Res.* 28, 101–114. doi: 10.1111/j.1365-2109.1997.tb01022.x
- Kai, W., Kikuchi, K., Tohari, S., Chew, A. K., Tai, A., Fujiwara, A., et al. (2011). Integration of the genetic map and genome assembly of fugu facilitates insights into distinct features of genome evolution in Teleosts and mammals. *Genome Biol. Evol.* 3, 424–442. doi: 10.1093/gbe/evr041
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., et al. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genet.* 8:e1002798. doi: 10.1371/journal.pgen.1002798

- Kobayashi, T., and Nagahama, Y. (2009). Molecular aspects of GD in a teleost fish, the Nile tilapia. *Sex. Dev.* 3, 108–117. doi: 10.1159/000223076
- Kobayashi, Y., Alam, M. A., Horiguchi, R., Shimizu, A., and Nakamura, M. (2010). Sexually dimorphic expression of gonadotropin subunits in the pituitary of protogynous honeycomb grouper (*Epinephelus merra*): evidence that follicle-stimulating hormone (FSH) induces gonadal sex change. *Biol. Reprod.* 82, 1030–1036. doi: 10.1095/biolreprod.109.080986
- Kondo, M., Nanda, I., Schmid, M., and Scharl, M. (2009). Sex determination and sex chromosome evolution: insights from medaka. *Sex. Dev.* 3, 88–98. doi: 10.1159/000223074
- Koopman, P. (2009). Sex determination: the power of DMRT1. *Trends Genet.* 25, 479–481. doi: 10.1016/j.tig.2009.09.009
- Koumoundouros, G., Divanach, P., Anezaki, L., and Kentouri, M. (2001). Temperature-induced ontogenetic plasticity in sea bass (*Dicentrarchus labrax*). *Mar. Biol.* 139, 817–830. doi: 10.1007/s002270100635
- Koumoundouros, G., Pavlidis, M., Anezaki, L., Kokkari, C., Sterioli, K., Divanach, P., et al. (2002). Temperature sex determination in the European sea bass, *Dicentrarchus labrax* (L., 1758) (*Teleostei, Perciformes, Moronidae*): critical sensitive ontogenetic phase. *J. Exp. Zool.* 292, 573–579. doi: 10.1002/jez.10095
- Lee, B. Y., Hulata, G., and Kocher, T. D. (2004). Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity (Edinb)* 92, 543–549. doi: 10.1038/sj.hdy.6800453
- Lee, K. H., Yamaguchi, A., Rashid, H., Kadomura, K., Yasumoto, S., and Matsuyama, M. (2009). Germ cell degeneration in high-temperature treated pufferfish, *Takifugu rubripes*. *Sex. Dev.* 3, 225–232. doi: 10.1159/000228723
- Lewis, Z. R., McClellan, M. C., Postlethwait, J. H., Cresko, W., and Kaplan, R. (2008). Female-specific increase in primordial germ cells marks sex differentiation in threespine stickleback, *Gasterosteus aculeatus*. *J. Morphol.* 269, 909–921. doi: 10.1002/jmor.10608
- Liew, W. C., Bartfai, R., Lim, Z. J., Sreenivasan, R., Siegfried, K. R., and Orban, L. (2012). Polygenic sex determination system in zebrafish. *PLoS ONE* 7:e34397. doi: 10.1371/journal.pone.0034397
- Liu, F., Sun, F., Li, J., Xia, J. H., Lin, G., Tu, R. J., et al. (2013). A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x *Oreochromis* spp.) and mapping of sex-determining loci. *BMC Genomics* 14:58. doi: 10.1186/1471-2164-14-58
- Lozano, C. A., Gjerde, B., Odegard, J., and Bentsen, H. B. (2013). Heritability estimates for male proportion in the GIFT Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture* 372, 137–148. doi: 10.1016/j.aquaculture.2012.10.029
- Lühmann, L. M., Knorr, C., Hörstgen-Schwark, G., and Wessels, S. (2012). First evidence for family-specific QTL for temperature-dependent sex reversal in Nile tilapia (*Oreochromis niloticus*). *Sex. Dev.* 6, 247–256. doi: 10.1159/000339705
- Mank, J. E., and Avise, J. C. (2009). Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex. Dev.* 3, 60–67. doi: 10.1159/000223071
- Mankiewicz, J. L., Godwin, J., Holler, B. L., Turner, P. M., Murashige, R., Shamey, R., et al. (2013). Masculinizing effect of background color and cortisol in a flat-fish with environmental sex-determination. *Int. Comp. Biol.* 53, 755–765. doi: 10.1093/icb/ict093
- Marín, I., and Baker, B. S. (1998). The evolutionary dynamics of sex determination. *Science* 281, 1990–1994. doi: 10.1126/science.281.5385.1990
- Martínez, P., Bouza, C., Hermida, M., Fernández, J., Toro, M. A., Vera, M., et al. (2009). Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). *Genetics* 183, 1443–1452. doi: 10.1534/genetics.109.107979
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., et al. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 30, 559–563. doi: 10.1038/nature751
- McKay, T. F. (2001). The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 35, 303–339. doi: 10.1146/annurev.genet.35.102401.090633
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., et al. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* 191, 163–170. doi: 10.1534/genetics.111.137497
- Nakamura, S., Kurokawa, H., Asakawa, S., Shimizu, N., and Tanaka, M. (2009). Two distinct types of theca cells in the medaka gonad: germ cell-dependent maintenance of cyp19a1-expressing theca cells. *Dev. Dyn.* 238, 2652–2657. doi: 10.1002/dvdy.22068
- Navarro-Martín, L., Galay-Burgos, M., Sweeney, G., and Piferrer, F. (2009a). Different sox17 transcripts during sex differentiation in sea bass, *Dicentrarchus labrax*. *Mol. Cell. Endocrinol.* 299, 240–251. doi: 10.1016/j.mce.2008.11.013
- Navarro-Martín, L., Blázquez, M., Viñas, J., and Piferrer, F. (2009b). Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*): limitations and opportunities for the production of highly female-biased stocks. *Aquaculture* 296, 347–358. doi: 10.1016/j.aquaculture.2009.07.022
- Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L., et al. (2011). DNA methylation of the gonadal aromatase (cyp19a) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genet.* 7:e1002447. doi: 10.1371/journal.pgen.1002447
- Nef, S., and Vassalli, J. D. (2009). Complementary pathways in mammalian female sex determination. *J. Biol.* 8:74. doi: 10.1186/jbiol173
- Nelson, J. S. (2006). *Fishes of the World*, 4th edn. New York: John Wiley and Sons, Inc., 601.
- Noletto, R. B., Vicari, M. R., Cestari, M. M., and Artoni, R. F. (2012). Variable B chromosome frequencies between males and females of two species of pufferfishes (*Tetraodontiformes*). *Rev. Fish Biol. Fisher.* 22, 343–349. doi: 10.1007/s11160-011-9231-9
- Oliveira, C., Foresti, F., and Hilsdorf, A. W. (2009). Genetics of neotropical fish: from chromosomes to populations. *Fish Physiol. Biochem.* 35, 81–100. doi: 10.1007/s10695-008-9250-1
- Orban, L., Sreenivasan, R., and Olsson, P. E. (2009). Long and winding roads: testis differentiation in zebrafish. *Mol. Cell. Endocrinol.* 312, 35–41. doi: 10.1016/j.mce.2009.04.014
- Ospina-Álvarez, N., and Piferrer, F. (2008). Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* 3:e2837. doi: 10.1371/journal.pone.0002837
- Palaiokostas, C., Bekaert, M., Khan, M. G. Q., Taggart, J. B., Gharbi, K., McAndrew, B. J., et al. (2013a). Mapping and validation of the major sex-determining region in Nile Tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE* 8:e68389. doi: 10.1371/journal.pone.0068389
- Palaiokostas, C., Bekaert, M., Khan, M. G. Q., Taggart, J. B., Gharbi, K., McAndrew, B. J., et al. (2013b). Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics* 14:566. doi: 10.1186/1471-2164-14-566
- Parise-Maltempi, P. P., da Silva, E. L., Rens, W., Dearden, F., O'Brien, P. C., Trifonov, V., et al. (2013). Comparative analysis of sex chromosomes in *Leporinus* species (*Teleostei, Characiformes*) using chromosome painting. *BMC Genet.* 14:60. doi: 10.1186/1471-2156-14-60
- Parker, G. A. (1992). The evolution of sexual size dimorphism in fish. *J. Fish Biol.* 41, 1–20. doi: 10.1111/j.1095-8649.1992.tb03864.x
- Parnell, N. F., and Strelman, J. T. (2013). Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity (Edinb)* 110, 239–246. doi: 10.1038/hdy.2012.73
- Pavlidis, M., Koumoundouros, G., Sterioli, A., Somarakis, S., Divanach, P., and Kentouri, M. (2000). Evidence of temperature-dependent sex determination in the European sea bass (*Dicentrarchus labrax* L.). *J. Exp. Zool.* 287, 225–232. doi: 10.1002/1097-010X(20000801)287:3<225::AID-JEZA>3.0.CO;2-D
- Penman, D. J., and Piferrer, F. (2008). Fish gonadogenesis. Part I: genetic and environmental mechanisms of sex determination. *Rev. Fish. Sci.* 16(Suppl. 1), 16–34. doi: 10.1080/10641260802324610
- Pereiro, P., Balseiro, P., Romero, A., Dios, S., Forn-Cuni, G., Fusté, B., et al. (2012). High-throughput sequence analysis of turbot (*Scophthalmus maximus*) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes. *PLoS ONE* 7:e35369. doi: 10.1371/journal.pone.0035369
- Person-Le Ruyet, J., Mahé, K., Le Bayon, N., and Le Delliou, H. (2004). Effects of temperature on growth and metabolism in a Mediterranean population of European sea bass, *Dicentrarchus labrax*. *Aquaculture* 237, 269–280. doi: 10.1016/j.aquaculture.2004.04.021
- Phillips, R. B. (2013). Evolution of the sex chromosomes in salmonid fishes. *Cytogenet. Genome Res.* 141, 177–185. doi: 10.1159/000355149
- Phillips, R. B., Konkol, N. R., Reed, K. M., and Stein, J. D. (2001). Chromosome painting supports lack of homology among sex chromosomes in *Oncorhynchus*, *Salmo*, and *Salvelinus* (*Salmonidae*). *Genetica* 111, 119–123. doi: 10.1023/A:1013743431738
- Piferrer, F. (2001). Endocrine sex control strategies for the feminization of teleost fish. *Aquaculture* 197, 229–281. doi: 10.1016/S0044-8486(01)00589-0
- Piferrer, F. (2013). Epigenetics of sex determination and gonadogenesis. *Dev. Dyn.* 242, 360–370. doi: 10.1002/dvdy.23924

- Piferrer, F., Beaumont, A., Falguière, J. C., Flajshans, M., Haffray, P., and Colombo, L. (2009). The use of induced polyploidy in the aquaculture of bivalves and fish for performance improvement and genetic containment. *Aquaculture* 293, 125–156. doi: 10.1016/j.aquaculture.2009.04.036
- Piferrer, F., Blázquez, M., Navarro, L., and González, A. (2005). Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass, *Dicentrarchus labrax* L. *Gen. Comp. Endocrinol.* 142, 102–110. doi: 10.1016/j.ygcen.2005.02.011
- Piferrer, F., Ribas, L., and Díaz, N. (2012). Genomic approaches to study genetic and environmental influences on fish sex determination and differentiation. *Mar. Biotechnol.* 14, 591–604. doi: 10.1007/s10126-012-9445-4
- Ribas, L., Pardo, B. G., Fernandez, C., Alvarez-Dios, J. A., Gómez-Tato, A., Quiroga, M. I., et al. (2013). A combined strategy involving Sanger and 454 pyrosequencing increases genomic resources to aid in the management of reproduction, disease control and genetic selection in the turbot (*Scophthalmus maximus*). *BMC Genomics* 14:180. doi: 10.1186/1471-2164-14-180
- Rice, W. R. (1986). On the instability of polygenic sex determination: the effect of sex-specific selection. *Evolution* 40, 633–639. doi: 10.2307/2408584
- Roberts, R. B., Ser, J. R., and Kocher, T. D. (2009). Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* 326, 998–1001. doi: 10.1126/science.1174705
- Ross, J. A., and Peichel, C. L. (2008). Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics* 179, 2173–2182. doi: 10.1534/genetics.108.088559
- Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., and Peichel, C. L. (2009). Turnover of sex chromosomes in the stickleback fishes (*Gasterosteidae*). *PLOS Genet.* 5:e1000391. doi: 10.1371/journal.pgen.1000391
- Saillant, E., Fostier, A., Haffray, P., Menu, B., Thimonier, J., and Chatain, B. (2002). Temperature effects and genotype-temperature interactions on sex determination in the European sea bass (*Dicentrarchus labrax* L.). *J. Exp. Zool.* 292, 494–505. doi: 10.1002/jez.10071
- Saito, D., and Tanaka, M. (2009). Comparative aspects of gonadal sex differentiation in medaka: a conserved role of developing oocytes in sexual canalization. *Sex. Dev.* 3, 99–107. doi: 10.1159/000223075
- Schlueter, P. J., Sang, X., Duan, C., and Wood, A. W. (2007). Insulin-like growth factor receptor 1b is required for zebrafish primordial germ cell migration and survival. *Dev. Biol.* 305, 377–387. doi: 10.1016/j.ydbio.2007.02.015
- Schultheis, C., Böhne, A., Scharlt, M., Volff, J. N., and Galiana-Arnoux, D. (2009). Sex determination diversity and sex chromosome evolution in poeciliid fish. *Sex. Dev.* 3, 68–77. doi: 10.1159/000223072
- Ser, J. R., Roberts, R. B., and Kocher, T. D. (2010). Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution* 64, 486–501. doi: 10.1111/j.1558-5646.2009.00871.x
- Shikano, T., Natri, H. M., Shimada, Y., and Merilä, Y. (2011). High degree of sex chromosome differentiation in stickleback fishes. *BMC Genomics* 12:474. doi: 10.1186/1471-2164-12-474
- Shinomiya, A., Otake, H., Hamaguchi, S., and Sakaizumi, M. (2010). Inherited XX sex reversal originating from wild medaka populations. *Heredity (Edinb)* 105, 443–448. doi: 10.1038/hdy.2010.51
- Siegfried, K. R. (2010). In search of determinants: gene expression during gonadal sex differentiation. *J. Fish Biol.* 76, 1879–1902. doi: 10.1111/j.1095-8649.2010.02594.x
- Siegfried, K. R., and Nusslein-Volhard, C. (2008). Germ line control of female sex determination in zebrafish. *Dev. Biol.* 324, 277–287. doi: 10.1016/j.ydbio.2008.09.025
- Smith, C. A., and Sinclair, A. H. (2004). Sex determination: insights from the chicken. *Bioessays* 26, 120–132. doi: 10.1002/bies.10400
- Sreenivasan, R., Jiang, J., Wang, X., Bárfai, R., Kwan, H. Y., Christoffels, A., et al. (2014). Gonad differentiation in zebrafish is regulated by the canonical Wnt signaling pathway. *Biol. Reprod.* 90:45. doi: 10.1095/biolreprod.113.110874
- Sun, F., Liu, S., Gao, X., Jiang, Y., Perera, D., Wang, X., et al. (2013). Male-biased genes in catfish as revealed by RNA-Seq analysis of the testis transcriptome. *PLoS ONE* 8:e68452. doi: 10.1371/journal.pone.0068452
- Taboada, X., Hermida, M., Pardo, B. G., Vera, M., Piferrer, F., Viñas, A., et al. (2014). Fine mapping and evolution of the major sex determining region in turbot (*Scophthalmus maximus*). *G3 (Bethesda)* doi: 10.1534/g3.114.012328 [Epub ahead of print].
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2007). Evidence for different origins of sex chromosomes in closely related *Oryzias* fishes: substitution of the master sex-determining gene. *Genetics* 177, 2075–2081. doi: 10.1534/genetics.107.075598
- Tao, W., Yuan, J., Zhou, L. Y., Sun, L. N., Sun, Y. L., Yang, S. J., et al. (2013). Characterization of gonadal transcriptomes from Nile Tilapia (*Oreochromis niloticus*) reveals differentially expressed genes. *PLoS ONE* 8:e63604. doi: 10.1371/journal.pone.0063604
- Tripathi, N., Hoffmann, M., Weigel, D., and Dreyer, C. (2009). Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy, *Poecilia reticulata*. *Genetics* 182, 365–374. doi: 10.1534/genetics.108.098541
- Uller, T., and Helanterä, H. (2011). From the origin of sex-determining factors to the evolution of sex-determining systems. *Q. Rev. Biol.* 86, 163–180. doi: 10.1086/661118
- Valenzuela, N., and Lance, V. A. (2004). *Temperature Dependent Sex Determination in Vertebrates*. Washington: Smithsonian Books.
- Vandeputte, M., Dupont-Nivet, M., Chavanne, H., and Chatain, B. (2007). A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics* 176, 1049–1057. doi: 10.1534/genetics.107.072140
- Vicari, M. R., Artoni, R. F., Moreira-Filho, O., and Bertollo, L. A. C. (2008). Diversification of a ZZ/ZW sex chromosome system in *Characidium* fish (*Crenuchidae*, *Characiformes*). *Genetica* 134, 311–317. doi: 10.1007/s10709-007-9238-2
- Vicente, V. E., Moreira-Filho, O., and Camacho, J. P. (1996). Sex-ratio distortion associated with the presence of a B chromosome in *Astyanax scabripinnis* (*Teleostei*, *Characidae*). *Cytogenet. Cell Genet.* 74, 70–75. doi: 10.1159/000134385
- Wu, G. C., Tomy, S., Lee, M. F., Lee, Y. H., Yueh, W. S., Lin, C. J., et al. (2010). Sex differentiation and sex change in the protandrous black porgy, *Acanthopagrus schlegelii*. *Gen. Comp. Endocrinol.* 167, 417–421. doi: 10.1016/j.ygcen.2009.11.003
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., et al. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Curr. Biol.* 22, 1423–1428. doi: 10.1016/j.cub.2012.05.045
- Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., et al. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* 6, 486–496. doi: 10.1111/eva.12032
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., et al. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2469–2474. doi: 10.1073/pnas.0712244105
- Zanuy, S., Carrillo, M., Felip, A., Rodríguez, L., Blázquez, M., Ramos, J., et al. (2001). Genetic, hormonal and environmental approaches for the control of reproduction in the European sea bass (*Dicentrarchus labrax* L.). *Aquaculture* 202, 187–203. doi: 10.1016/S0044-8486(01)00771-2
- Zhou, L., and Gui, J. F. (2010). Molecular mechanisms underlying sex change in hermaphroditic groupers. *Fish Physiol. Biochem.* 36, 181–193. doi: 10.1007/s10695-008-9219-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 June 2014; accepted: 10 September 2014; published online: 29 September 2014.

Citation: Martínez P, Viñas AM, Sánchez L, Díaz N, Ribas L and Piferrer F (2014) Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front. Genet.* 5:340. doi: 10.3389/fgenet.2014.00340

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Martínez, Viñas, Sánchez, Díaz, Ribas and Piferrer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic considerations for mollusk production in aquaculture: current state of knowledge

Marcela P. Astorga *

Instituto de Acuicultura, Universidad Austral de Chile, Puerto Montt, Chile

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Ross Houston, University of Edinburgh, UK

Jesús Fernández, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Spain

***Correspondence:**

Marcela P. Astorga, Instituto de Acuicultura, Universidad Austral de Chile, Los Pinos s/n. Pelluco, Casilla 1327, Puerto Montt, Chile
e-mail: marcelaastorga@spm.uach.cl

In 2012, world mollusk production in aquaculture reached a volume of 15,171,000 tons, representing 23% of total aquaculture production and positioning mollusks as the second most important category of aquaculture products (fishes are the first). Clams and oysters are the mollusk species with the highest production levels, followed in descending order by mussels, scallops, and abalones. In view of the increasing importance attached to genetic information on aquaculture, which can help with good maintenance and thus the sustainability of production, the present work offers a review of the state of knowledge on genetic and genomic information about mollusks produced in aquaculture. The analysis was applied to mollusks which are of importance for aquaculture, with emphasis on the 5 species with the highest production levels. According to FAO, these are: Japanese clam *Ruditapes philippinarum*; Pacific oyster *Crassostrea gigas*; Chilean mussel *Mytilus chilensis*; Blood clam *Anadara granosa* and Chinese clam *Sinonovacula constricta*. To date, the genomes of 5 species of mollusks have been sequenced, only one of which, *Crassostrea gigas*, coincides with the species with the greatest production in aquaculture. Another important species whose genome has been sequenced is *Mytilus galloprovincialis*, which is the second most important mussel in aquaculture production, after *M. chilensis*. Few genetic improvement programs have been reported in comparison with the number reported in fish species. The most commonly investigated species are oysters, with at least 5 genetic improvement programs reported, followed by abalones with 2 programs and mussels with one. The results of this work will establish the current situation with respect to the genetics of mollusks which are of importance for aquaculture production, in order to assist future decisions to ensure the sustainability of these resources.

Keywords: mollusk, genomic, aquaculture, genetic improvement, genetic, bivalves

AQUACULTURE

The aquaculture industry consists of 3 main groups with different production volumes: fish, with a volume of 44.15 million tons in 2012, representing 66.3% of production; mollusks with 15.17 million tons equivalent to 22.8%; crustaceans, 6.44 million tons, 9.7%. The production of other types was 0.86 million tons, or 1.3% (FAO, 2014). Based on these values, mollusks form the second largest group; however the levels of biotechnological research and the technologies applied are very low compared to the high level of development in cultivated fish species.

Biotechnology and genetic research have developed strongly in the last decade, and a large database now exists for fish species of importance for aquaculture (NCBI, Genbank, EMBL). These new research lines have permitted great progress and increased efficiency. Research of this kind has led to the development of many applications for fish production, with an increase in the number of techniques for genetic improvement, and a wider range of genetic variability of the resources for application in studies of traceability, pedigree, degrees of relatedness, and the search for molecular markers for Marked Assisted Selection. These developments are well reviewed in Hulata (2001), Liu and

Cordes (2004), and Martínez (2007); genetic improvement programs are reviewed in Gjedrem (2012) and Gjedrem et al. (2012). In mollusk production, however, the development of such knowledge has been minimal (for Latin American mollusks see Astorga, 2008).

The present work therefore reviews the current state of scientific knowledge on the genetics of mollusk species which are important for aquaculture. The amount of existing information on the genomes of these resources will be assessed—this is needed to establish a baseline for knowledge-based decision-making. Such knowledge would also enable us to define brood-stock selection programs, generate the bases for genetic improvement programs and finally assess the state of a given resource in order to ensure its long term conservation and sustainability.

MOLLUSK PRODUCTION

World mollusk production in 2012 was 15.17 million tons. Total aquaculture production of mollusks is represented by 5 main groups: clams, oysters, mussels, scallops, and abalones. Clam production during 2011 was 4.929 million tons; this was followed by oysters with 4.519 million tons; mussels with

1.802 million tons; scallops with 1.52 million tons; and finally abalones with aquaculture production of 395,000 tons (FAO, 2013).

There are 5 principal mollusk species produced in aquaculture (i.e., with production of over 250,000 tons), all bivalves, which correspond to only 5 species of mollusks reported in the FAO statistics (2013). Three of these are clams: *Ruditapes philippinarum* or Japanese clam; *Anadara granosa* (= *Tegillarca granosa*), Blood clam or Blood cockle from Malaysia; and *Sinonovacula constricta* or Chinese clam. The other two are the oyster *Crassostrea gigas* and the Chilean mussel *Mytilus chilensis*. The detailed production for each of these is shown in **Table 1**, including the economic value generated by the production of these 5 resources with the highest volume in shellfish aquaculture. In this review the state of knowledge on the genetics of each of these 5 species will be evaluated.

GENETIC RESOURCES

Genetic resources is the name given to all the genetic or genome information for a group of representative individuals of a species which constitutes a natural resource (CBD Nagoya Protocol, 2010). Knowledge of the genetic resources of species has become increasingly important with the discovery of the importance of genetic indicators for species conservation, obtaining sufficient raw material to generate successful improvement programs, the sustainability of natural banks and the appropriate management of genetic resources (Kenchington et al., 2003).

Correct use of genetic resources allows the species to be developed sustainably, maintaining it for future generations; this information is also the basis for new genetic improvement programs in species in early or advanced stages of aquaculture (Kenchington et al., 2003; Guo, 2009; Lind et al., 2012). To start mollusk production in a hatchery, Gaffney (2006) proposes that high levels of genetic diversity must be ensured among the brood-stock. Proper levels of variability must then be maintained over time, by monitoring genetic variability to avoid excessive inbreeding (Gaffney, 2006). These factors—which require the application of biotechnologies or the design of genetic improvement plans—have been identified as important for starting aquaculture production.

Genetic improvement programs require reproductive management throughout the whole cycle of the species. Mollusk production is still carried out in extensive aquaculture facilities,

where the initial production stages depend completely on the natural environment and on seed obtained from natural banks for the fattening stage, which makes it difficult to design breeding programs. On the other hand, the costs of obtaining seed from the natural environment are drastically lower than those of seed production in a hatchery, threatening the profitability of mollusk production companies and their ability to invest in genetic improvement programs. These programs may become very profitable, but only if the sale value of the product is very high, as occurs with the oyster, or if the genetic gain is directly related with a reduction in production costs.

GENETIC INFORMATION ON THE MAIN AQUACULTURE SPECIES

(Sort by economic value of production according to **Table 1**).

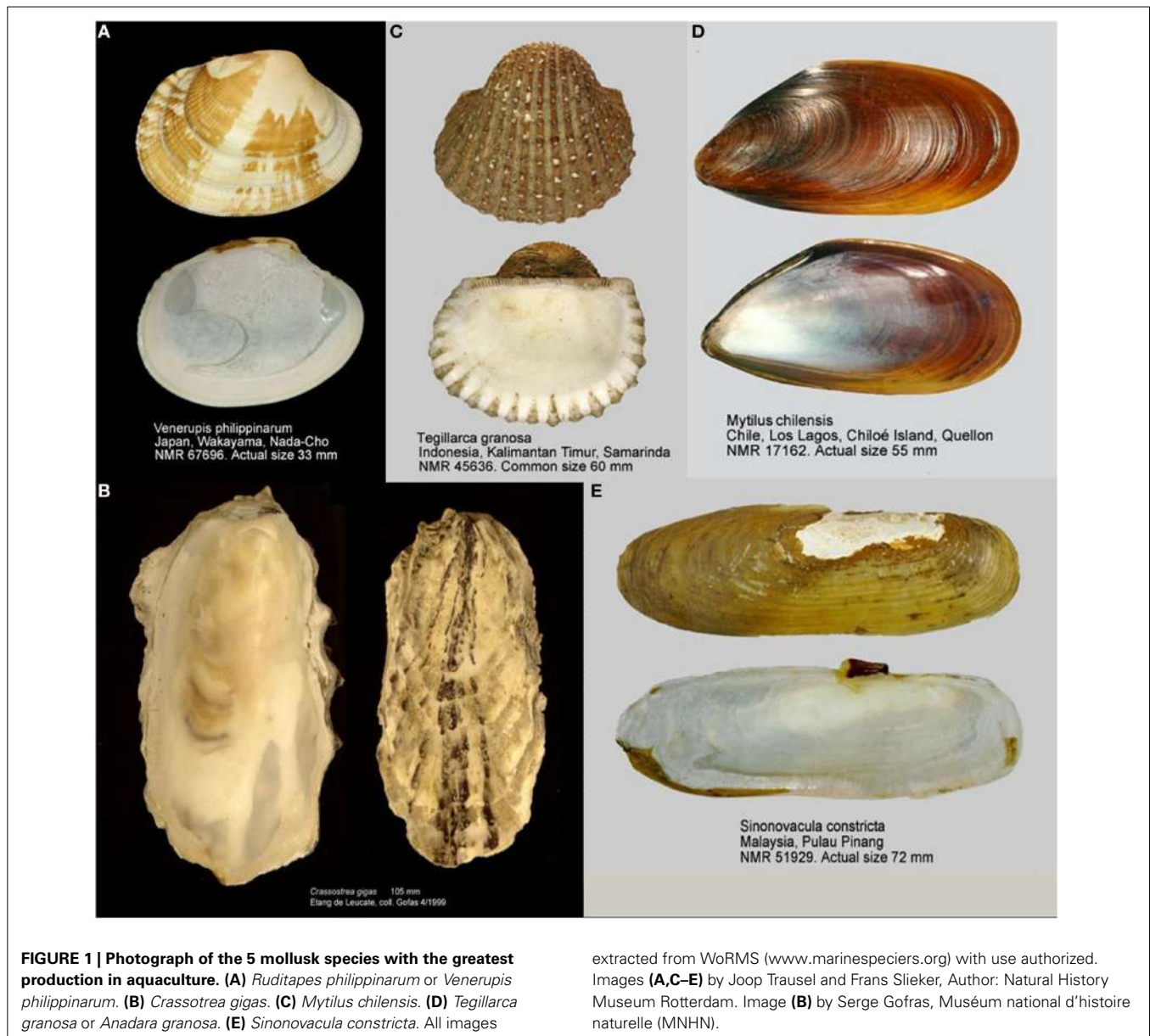
RUDITAPES PHILIPPINARUM

Genetic information on the Malaysian clam *Ruditapes philippinarum* (**Figure 1A**), also known as *Venerupis philippinarum*, was searched using the ISI Web of Science and the National Center of Biotechnology Information (NCBI) search engines and the PubMed database. This allowed us to establish the low level of existing genetic information, despite the fact that it is one of the 5 mollusk species with the highest production in the world and is located in the first place of economic value production. Only around 30 papers discussing the genetics of this species were found (see database at Annex 1), although we consider that the number is not always representative of the information available. The research focused mainly on the genetic characterization of populations and the evaluation of genetic differentiation and population structure (42%), followed by searches for molecular markers (17%), and evaluation of hybrids with congeneric species (13%). A few papers describe research related to the genetic response associated with the immune system, gene expression, and transcriptome (total 21%), and finally 2 works address heritability and selection measurement (8%). However, although the main focus of research is on molecular genetics and molecular biology, a large number of works was found associated with study of gene expression as it relates to the immune response or to exposure to contaminants; there were at least 19 papers from 2012 to date, showing a strong focus on the transcriptome of this species. On searching for genome information in the National Center for Biotechnology Information (NCBI), data was found on 30,442

Table 1 | Aquaculture production and economic value of the 5 mollusk species with production greater than 250,000 tons (FAO, 2013), with review of genetic and genome information.

Species	Common name	Production '000 Ton.	Total value MUS\$	WebScience publications	PubMed publications	Sequences in GenBank	EST in NCBI	Protein in NCBI
<i>Ruditapes philippinarum</i>	Japanese clam	3681	3438	25	273	30,442	6,469	1046
<i>Crassostrea gigas</i>	Pacific oyster	637	1323	200	1821	11,775	206,647	29,004
<i>Mytilus chilensis</i>	Chilean mussel	288	1148	15	35	119	7	27
<i>Anadara granosa</i>	Blood clam	393	468	6	48	641	2,278	158
<i>Sinonovacula constricta</i>	Chinese clam	744	670	5	23	16,383	5,296	204

The genetic information is based on the number of genetics-related publications in Web of Science databases, NCBI, PubMed, databases in GenBank, EST (Expressed Sequences Tag) regions and coded proteins.



DNA, and RNA sequences, 6469 EST (Expressed Sequence Tag) sequences, and 1046 protein sequences (Table 1). Little genetic information is available on this species, despite the fact that it has been cultivated in hatcheries since the mid 1970s (Zhang and Yan, 2006). The information is insufficient to generate a basis for developing genetic improvement programs. The heritability of shell length has been calculated in the natural and cultured population at the larval and juvenile stages (Yan et al., 2014); the values found were 0.22 for larvae and 0.39 for juveniles in natural populations, and 0.17 and 0.87 for cultured populations. These results suggest that selection should be highly effective in this species, and that selecting either a natural or a cultured population to obtain faster growth should be successful. Zhao et al. (2012) established that divergent selection is effective and promotes shell length changes in the Manila clam, finding relatively moderate heritability values of 0.165 and 0.260, but quite high

genetic gain values, with an increase in growth of 3.55% over the control line. Zhao et al. (2012) found that a base population with high genetic variability should be established as the basis for a divergent selection program. This would allow stable production of high quality seed to ensure the sustainability of aquaculture of this resource.

CRASSOSTREA GIGAS

Unlike the species described above, a large quantity of genetic research has been done in the oyster *Crassostrea gigas* (Figure 1B), associated with the search for molecular markers and the genetic characterization of populations. In the area with potential for application in aquaculture, works were found associated with selection response, heritability measurement, and searches for QTL and chromosome manipulation, highlighting the work of Guo et al. (2012), Ge et al. (2014) and Zhong et al. (2014). There

were also many works associated with gene expression in relation to environmental stressors and transcriptome analysis (citations from 2010 to date in Annex 2). This agrees with the large amount of information found in the NCBI database on EST (206,647) as compared to the other species with high aquaculture production (Table 1). In general databases of scientific publications, at least 450 citations from the last 10 years were found relating to the genetics of this species. It is therefore one of the bivalve mollusks with the most highly developed genetic and biotechnological research. Furthermore, the nuclear and mitochondrial genomes of this species have been sequenced, which is the case with only 5 species of mollusk and 2 species of bivalve mollusk. There is a specialized database on this oyster, called Gigabase, run by the IFREMER group in France. At least five genetic improvement programs have been developed for this species (Rye et al., 2010) although Gjedrem et al. (2012) mention only 3 selective breeding programs. There is a large volume of information for the implementation of new genetic improvement programs in the species. For example, we found a work assessing the heritability of shell pigmentation (Evans et al., 2009) which reported high values for narrow sense heritability ($h^2 = 0.59$). This suggests that total left-shell pigmentation in *C. gigas* is strongly influenced by additive genetic variation and therefore amenable to selection. Another characteristic which has been extensively studied in oysters is the genetic response to the high summer mortality observed in hatcheries (Dégremont et al., 2010). High realized heritabilities (h^2r) values have been estimated for high survival groups (above 0.88), higher than the values for low survival groups (0.55–0.68), and the authors conclude that these values would permit good results in a selective breeding program to improve survival (Dégremont et al., 2010). We also found estimates of growth heritability, with high values reported by Li et al. (2011) with 0.40; 0.33; and 0.15 and Wang et al. (2012)—who also report high heritability values for growth rate (0.457; 0.312; and 0.332), which has allowed a solid basis to be established for the further development of genetic improvement programs for this species, ensuring greater genetic gain in the short term.

MYTILUS CHILENSIS

There is less scientific information published on the mussel *Mytilus chilensis* (Figure 1C) than the two previous species; in August 2014, only 15 papers on the genetics of this species could be found in the Web of Science (Annex 3). Aquaculture production is higher for this species than for other mussels, yet the amount of scientific information published is smaller—for example, around 100 scientific publications were found in the same database for the mussel *Mytilus galloprovincialis*. The largest proportion (50%) of the research published on *M. chilensis* is associated with the genetic characterization and population structure of this species distributed along the Chilean coast. Works can be found on the identification of molecular markers, heritability, selection response and gene expression, all in much the same proportions and small quantities. This is corroborated by a search for genome information in the NCBI databases, where only 119 data were found on sequenced genes, and only 7 EST were identified (Table 1). Of the 5 top species in aquaculture production,

M. chilensis presents the least amount of genome information and its genome has not been sequenced. However, it should be noted that the genome of the Mytilid species *Mytilus galloprovincialis*, has been sequenced; this species, cultivated principally around the coast of Spain, presents the next highest aquaculture production among mussels.

There is one record of a genetic improvement program for mussels (Rye et al., 2010; Gjedrem et al., 2012); however, it is not being applied in production because cultivation is based on seed obtained from the natural environment and not from hatcheries. Nevertheless, works exist which estimate heritability and response to selection in this species. High heritability values have been found for a variety of characteristics such as growth in shell height and live weight (0.2–0.9) (Toro et al., 2004a). In larvae these parameters present a heritability range between 0.38 and 0.84 (Toro et al., 2004b). Finally, Alcapan et al. (2007) assessed the effects of environment and ageing on the heritability of body size in *M. chilensis*; they observed great variability in heritability values, conditioned by the site and the age of the individuals. These data would allow good results to be obtained in improvement programs, however the high production costs of seed obtained from hatcheries as opposed to the natural environment remains a limiting factor for this species. This makes the implementation of these improvement programs uncompetitive for producers in the short term. Nevertheless, the possibility cannot be ruled out that it may become necessary in future to make seed of high genetic quality available to this production sector.

ANADARA GRANOSA

In the case of the Blood cockle *Anadara granosa* or *Tegillarca granosa* (Figure 1D), at least 4 publications were found on genetic characterization of populations, and two more relating to the search for molecular markers (Annex 4). There is therefore very little genetic information published for this species, despite its high production value. However, when we searched for genome information, we found more research into EST regions in this species (2278) (Table 1) than in *Mytilus chilensis* (only 7 EST). No records were found of current genetic improvement programs in this species, nor any information on baseline data which would allow the development potential of such programs to be assessed. On the other hand, the natural population structure has been studied (Ni et al., 2012; Wang et al., 2013) and this is a pre-requirement for selecting brood-stock and forming families for an improvement program. Wang et al. (2013) reported high genetic variability values, meaning that high diversity exists for brood-stock selection. However, they found genetic divergence in various sites on the coast of China presenting lower genetic variability, possibly caused by the admixture of artificially produced seed. This type of information suggests that genetic improvement programs need to include cross-breeding in their design in order to prevent genetic variability from being reduced, perhaps by using a large number of breeding individuals or a group which is representative of the genetic variability of the population. Ni et al. (2012) observe high genetic structuring in populations, indicating a high risk of erosion and possible local extinction. Populations need to be protected to prevent loss of genetic variability in the species.

SINONOVACULA CONSTRICTA

Of the 5 top species in aquaculture production, the one with the least scientific information published on genetics is the Chinese clam *Sinonovacula constricta* (Figure 1E). In the Web of Science databases to August 2014, only three publications were found on searches for molecular markers plus two on the genetic evaluation of populations (Annex 5) (Table 1). However, there is a large number of sequenced genes in the NCBI databases, many more than for some other species analyzed. In this respect it ranks third among the top 5 species (Table 1), with the highest number of genes sequenced among the main 5 species produced in aquaculture (16,383 genes). There are no records of genetic improvement programs in this species, nor any publications of estimators for the implementation of such programs. Very little is known about its population structure. Niu et al. (2012) observed high genetic variability in 10 populations on the Chinese coast; however, they found genetic differentiation between sites, and propose the possible presence of cryptic species. This aspect should be studied since it is important to identify the cultivated species correctly and establish different management methods if more than one species exists.

CONCLUDING REMARKS

On the basis of this review it is apparent that there is no obvious relationship between the economic importance of a species, in terms of the volume produced in aquaculture, and the amount of scientific information available about genetics/genomic resources, when restricted to the five most produced species. In this case our analysis was carried out for research into genetics, which is the basis for genetic improvement programs in aquaculture species. The scientific information found is summarized, with the databases used, in Table 1. From this we see that the most studied species, using all the genetic information indicators, is *Crassostrea gigas*. There is great interest in this species and it has been introduced into many parts of the world for cultivation. Among mussels, the small amount of genetic information available on *Mytilus chilensis* is striking. At the same time the congeneric species *Mytilus galloprovincialis* is the most studied mussel; furthermore the latter is one of only two mussel species whose genome has been sequenced, and there are 41,294 records of DNA sequences, 19,756 of EST and 2,424 coded proteins. The species with the largest number of DNA sequences is the oyster *Crassostrea angulata*, with 81,015 DNA sequences, even more than *C. gigas*. On the other hand, *C. gigas* is the bivalve mollusk with the most EST sequences identified.

The quantity of genetic information available on bivalve mollusks varies widely, with great differences between species, even though their production volumes in aquaculture are similar. There is also still little information with which to generate solid bases for genetic improvement programs such as are observed in fish or other cultivated species. The only exception is *C. gigas*, for which adequate base information exists and genetic improvement programs are currently being implemented. Sufficient information exists on the mussel *M. chilensis* to provide a basis for more improvement programs, but the low use of reproduction in controlled environments is a limiting factor in this species. For the other three species there is virtually no basic genetic information

on which predicting their response to selection programs. Further research is required to generate a sufficient basis for selective reproduction, and also to clarify information on natural populations, identifying possible cryptic species or strong differentiation between populations. The lack of knowledge about species with high production in aquaculture could lead to population bottlenecks, since if the brood-stock management plan is unsuitable, or seed extraction from local populations is not handled correctly, the populations which sustain these resources could be drastically reduced. This could result in a severe diminution of genetic diversity, to the point where it might be difficult for the species to recover, meaning that in future these resources which are so important for aquaculture would not be sustainable.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00435/abstract>

REFERENCES

- Alcapan, A. C., Nespolo, R. F., and Toro, J. E. (2007). Heritability of body size in the Chilean blue mussel (*Mytilus chilensis* Hupe 1854): effects of environment and ageing. *Aquacult. Res.* 38, 313–320. doi: 10.1111/j.1365-2109.2007.01678.x
- Astorga, M. P. (2008). “Estado actual del uso de marcadores moleculares en moluscos bivalvos de importancia para la acuicultura,” in *Estado Actual Del Cultivo y Manejo de Moluscos Bivalvos y su Proyección Futura: Factores Que Afectan Su Sustentabilidad en America Latina*, eds A. Lovatelli, A. Farias, and I. Uriarte Actas de Pesca de la FAO. N 12 (Roma: FAO), 277–287.
- CBD Nagoya Protocol. (2010). *Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization*. Nagoya: CBD. Available online at: www.cbd.int/decision/cop/?id=12267
- Dégremont, L., Bédier, E., and Boudry, P. (2010). Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). II. Response to selection for survival and its influence on growth and yield. *Aquaculture* 299, 21–29. doi: 10.1016/j.aquaculture.2009.11.017
- Evans, S., Camara, M. D., and Langdon, C. J. (2009). Heritability of shell pigmentation in the Pacific oyster, *Crassostrea gigas*. *Aquaculture* 286, 211–216. doi: 10.1016/j.aquaculture.2008.09.022
- FAO. (2013). *FAO Yearbook. Fishery and Aquaculture Statistics 2011*. Rome: Food and Agriculture Organization of the United Nations, 76.
- FAO. (2014). *The State of World Fisheries and Aquaculture 2014. Opportunities and challenges*. Rome: Food and Agriculture Organization of the United Nations, 223.
- Gaffney, P. M. (2006). The role of genetics in shellfish restoration. *Aquat. Living Resour.* 19, 277–282. doi: 10.1051/alr:2006028
- Ge, J., Li, Q., Yu, H., and Kong, L. (2014). Identification and mapping of a SCAR marker linked to a locus involved in shell pigmentation of the Pacific oyster (*Crassostrea gigas*). *Aquaculture* 434, 249–253 doi: 10.1016/j.aquaculture.2014.08.027
- Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture* 344, 12–22. doi: 10.1016/j.aquaculture.2012.03.003
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Guo, X. (2009). Use and exchange of genetic resources in molluscan aquaculture. *Rev. Aquacult.* 1, 251–259. doi: 10.1111/j.1753-5131.2009.01014.x
- Guo, X., Li, Q., Wang, Q. Z., and Kong, L. F. (2012). Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Mar. Biotechnol.* 14, 218–226. doi: 10.1007/s10126-011-9405-4
- Hulata, G. (2001). Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica* 111, 155–173. doi: 10.1023/A:1013776931796
- Kenchington, E., Heino, M., and Nielsen, E. E. (2003). Managing marine genetic diversity: time for action? *ICES J. Mar. Sci.* 60, 1172–1176. doi: 10.1016/S1054-3139(03)00136-X

- Li, Q., Wang, Q., Liu, S., and Kong, L. (2011). Selection response and realized heritability for growth in three stocks of the Pacific oyster *Crassostrea gigas*. *Fish. Sci.* 77, 643–648. doi: 10.1007/s12562-011-0369-0
- Lind, C. E., Ponzoni, R. W., Nguyen, N. H., and Khaw, H. L. (2012). Selective breeding in fish and conservation of genetic resources for aquaculture. *Reprod. Dom. Anim.* 47, 255–263. doi: 10.1111/j.1439-0531.2012.02084.x
- Liu, Z. J., and Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37. doi: 10.1016/j.aquaculture.2004.05.027
- Martínez, V. (2007). “Marker-assisted selection in fish and shellfish breeding schemes,” in *Marker-Assisted Selection—Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*, ed E. P. Guimarães (Rome: FAO), 329–362.
- Ni, G., Li, Q., Kong, L., and Zheng, X. (2012). Phylogeography of the bivalve *Tegillarca granosa* in coastal China: implications for management and conservation. *Mar. Ecol. Prog. Ser.* 452, 119–130. doi: 10.3354/meps09624
- Niu, D.-H., Feng, B.-B., Liu, D.-B., Zhong, Y.-M., Shen, H.-D., and Li, J.-L. (2012). Significant genetic differentiation among ten populations of the razor clam *Sinonovacula constricta* along the coast of china revealed by a microsatellite analysis. *Zool. Stud.* 51, 406–414.
- Rye, M., Gjerde, B., and Gjedrem, T. (2010). “Genetic improvement programs for aquaculture species in developed countries,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*. Available online at: <http://www.kongressband.de/wcgalp2010/assets/pdf/0963.pdf>. (Accessed January 12, 2012).
- Toro, J. E., Alcapán, A. C., Vergara, A. M., and Ojeda, J. A. (2004b). Heritability estimates of larval and spat height in the Chilean blue mussel (*Mytilus chilensis* Hupe 1854) produced under controlled laboratory conditions. *Aquacult. Res.* 35, 56–61. doi: 10.1111/j.1365-2109.2004.00985.x
- Toro, J. E., Alcapán, A. C., Ojeda, J. A., and Vergara, A. M. (2004a). Selection response for growth rate (shell height and live weight) in the Chilean blue mussel (*Mytilus chilensis* hupe 1854). *J. Shellfish Res.* 23, 753–757.
- Wang, Q. Z., Li, Q., Kong, L. F., and Yu, R. H. (2012). Response to selection for fast growth in the second generation of Pacific Oyster (*Crassostrea gigas*). *J. Ocean Univ. China* 11, 413–418. doi: 10.1007/s11802-012-1909-7
- Wang, Y. J., Zeng, Q. G., and Xu, L. N. (2013). Population structure of the blood clam (*Tegillarca granosa*) in China based on microsatellite markers. *Genet. Mol. Res.* 12, 892–900. doi: 10.4238/2013.April.2.6
- Yan, X., Huo, Z., Yang, F., and Zhang, G. (2014). Heritability of larval and juvenile growth for two stocks of Manila clam *Ruditapes philippinarum*. *Aquacult. Res.* 45, 484–490. doi: 10.1111/j.1365-2109.2012.03250.x
- Zhang, G., and Yan, X. (2006). A new three-phase culture method for Manila clam, *Ruditapes philippinarum*, farming in northern China. *Aquaculture* 258, 452–461. doi: 10.1016/j.aquaculture.2006.04.046
- Zhao, L., Yan, X., Huo, Z., and Yang, F. (2012). Divergent selection for shell length in the Manila Clam, *Ruditapes philippinarum*. *J. World Aquacult. Soc.* 43, 878–884. doi: 10.1111/j.1749-7345.2012.00612.x
- Zhong, X., Li, Q., Guo, X., Yu, H., and Kong, L. (2014). QTL mapping for glycogen content and shell pigmentation in the Pacific oyster *Crassostrea gigas* using microsatellites and SNPs. *Aquacult. Int.* 22, 1877–1889. doi: 10.1007/s10499-014-9789-z

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 September 2014; accepted: 24 November 2014; published online: 10 December 2014.

Citation: Astorga MP (2014) Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Front. Genet.* 5:435. doi: 10.3389/fgene.2014.00435 This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Astorga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



RNA-seq as a powerful tool for penaeid shrimp genetic progress

Camilla A. Santos*, Danielly V. Blanck and Patrícia D. de Freitas

Laboratory of Molecular Biodiversity and Conservation, Department of Genetics and Evolution, Federal University of São Carlos, São Carlos, Brazil

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Hsiao-Ching Liu, North Carolina State University, USA

Gonzalo Rincon, University of California, Davis, USA

*Correspondence:

Camilla A. Santos, Laboratory of Molecular Biodiversity and Conservation, Department of Genetics and Evolution, Federal University of São Carlos, Washington Luis Road (SP 310), Km 235, Caixa Postal 676, CEP: 13565-905, São Carlos, São Paulo, Brazil
e-mail: camilla.alves@yahoo.com.br

The sequences of all different RNA transcripts present in a cell or tissue that are related to the gene expression and its functional control represent what it is called a transcriptome. The transcripts vary between cells, tissues, ontogenetic and environmental conditions, and the knowledge that can be gained through them is of a solid relevance for genetic applications in aquaculture. Some of the techniques used in transcriptome studies, such as microarrays, are being replaced for next-generation sequencing approaches. RNA-seq emerges as a new possibility for the transcriptome complexity analysis as well as for the candidate genes and polymorphisms identification of penaeid species. Thus, it may also help to understand the determination of complex traits mechanisms and genetic improvement of stocks. In this review, it is first introduced an overview of transcriptome analysis by RNA-seq, followed by a discussion of how this approach may be applied in genetic progress within penaeid stocks.

Keywords: differential expression analysis, differential phenotype, high-throughput transcriptome, penaeidae transcriptome, candidate genes

INTRODUCTION

The term RNA-seq has been used to make reference to a transcriptome produced by methods of next generation sequencing (NGS), which ensure a good coverage of transcripts detection, due to the sequencing of millions of reads ranging from 25 to 300 bp, depending on the platform used (Wang et al., 2009; Oshlack et al., 2010). The full set of transcripts in a cell is known as transcriptome. They involve all types of ribonucleic acids (RNAs), including the protein coding messenger ribonucleic acid (mRNA) and the non-coding ribonucleic acid (ncRNA) such as ribosomal RNAs (rRNA), transfer RNAs (tRNA), and the small nuclear RNAs (snRNA). These RNAs may be differentially expressed according to the tissue, the stage of development and the physiological condition being accessed (Wang et al., 2009; Anders and Huber, 2010).

Transcriptome studies have been widely conducted in order to identify new genes, prospect simple sequence repeats (SSR) and single nucleotide polymorphisms (SNP) markers and to analyze differentially expressed genes. Such approaches have been helping to understand different mechanisms related to cellular control and describe important metabolic pathways, what enables a better understanding of the genotype–phenotype relationship (Marguerat and Bähler, 2010; Khatri et al., 2012; Qian et al., 2014).

Small and large scale transcriptome analyses and differential expression studies, such as Expressed Sequence Tags (ESTs) and microarrays, have been carried out in some penaeid shrimp species (Rojtinnakorn et al., 2002; La Vega et al., 2007; James et al., 2010; Brady et al., 2013). However, RNA-seq approaches are still incipient in shrimp (Table 1; Li et al., 2012, 2013; Guo et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014; Yu et al., 2014). Therefore, the method herein has emerged as a new possibility for the transcriptome complexity

analyses in face of varied production and/or experimental conditions. Consequently, such approach aims to develop genetically improved strains, focusing mainly on the resistance factor.

In order to obtain a transcriptome via RNA-seq, some steps should be followed: (i) selection of tissue of interest and isolation of RNA molecules; (ii) construction of cDNA libraries; (iii) utilization of a NGS platform; and (iv) the reads analysis in order to establish unigenes and the transcriptome assemble through bioinformatics tools.

The tissue choice should be based on the study aim and/or the genes nature to be analyzed. As a parallel, a transcriptome consists in taking a photograph from a specific time in a cell, highlighting only the condition at that short period of time. In this manner, tissue selection and the suitable time to perform a transcriptome requires preparation, otherwise the experiment as a whole may be biased (Wang et al., 2009). Libraries establishment is crucial for the final result in face of the many laboratorial procedures that are conducted, leading to some biases in the obtained results (Wang et al., 2009). Bioinformatics analysis is also an important step and includes the use of computational tools that guarantee the processing of large volumes of data generated by next-gen (Gavery and Roberts, 2012; Guo et al., 2013).

Within this review, it is presented a brief overview of the RNA-seq method, including its main advantages and limitations. Following that, it will be discussed how such technique may be applied to obtain genetic progress in penaeid shrimp farming.

RNA-seq: ADVANTAGES AND LIMITATIONS

The transcriptome assembly may be based on a reference genome available (Wang et al., 2009), which allows to quickly locate similar

Table 1 | Summary of studies on penaeid shrimp using the RNA-seq approach.

Reference/species	NGS platform	Application	Annotation of genomic information
Baranski et al. (2014)/ <i>Penaeus monodon</i>	Illumina	SNPs and microsatellite prospecting	136,223 contigs; 473,620 SNPs
Li et al. (2013)/ <i>Fenneropenaeus chinensis</i>	Illumina	Immune response to WSSV	46,676 unigenes; 805 differentially expressed genes
Yu et al. (2014)/ <i>Litopenaeus vannamei</i>	Illumina	SNPs prospecting	96,040 SNPs
Guo et al. (2013)/ <i>L. vannamei</i>	Illumina	Stress; immune response; apoptosis	42,336 unigenes; 2,445 differentially expressed genes
Sookruksawong et al. (2013)/ <i>L. vannamei</i>	Illumina	Immune response to TSV	61,937 contigs; 1,374 differentially expressed contigs
Zeng et al. (2013)/ <i>L. vannamei</i>	454	Immune response to TSV; microsatellites prospecting	15,004 unigenes; 770 microsatellites; 1,311 differentially expressed genes
Xue et al. (2013)/ <i>L. vannamei</i>	Illumina	Immune response to WSSV	52,073 unigenes; 1,179 differentially expressed genes
Li et al. (2012)/ <i>L. vannamei</i>	Illumina	Genomic annotation	109,169 unigenes

regions using local alignment algorithms and presents higher reliability due to the large volume of small sized reads coming from alternative splicing. It equally provides a more even coverage of the genome (Anders and Huber, 2010; Qian et al., 2014). On the other hand, even when there is no reference genome available, *de novo* transcriptome assembly may be carried out using specific algorithms, which stands as a solid advantage for some species that have not been widely studied yet (Howe et al., 2013; O'Neil and Emrich, 2013).

When performing transcriptome via RNA-seq, a high coverage is obtained, which allows the discovery of new genes and polymorphisms (Marguerat and Bähler, 2010; Yu et al., 2014). Li et al. (2012) evaluated the abundance and coverage of transcriptomes obtained by RNA-seq in *Litopenaeus vannamei*. By comparing such data to the ESTs available on GenBank, it was found that only 14.2% (15,519 out of 109,169) of unigenes obtained by RNA-seq were also found in the EST libraries, generating a lot of new informative data. In addition to that, the wide coverage associated with high resolution provided by this technique ensured high accuracy in SNP discovery in coding genes (Yu et al., 2014).

That taken, RNA-seq allows the detection of variations in a single nucleotide, enabling the detection of the expression of protein isoforms and their respective allelic variants, characterizing SNPs (Baranski et al., 2014; Yu et al., 2014). Polymorphic microsatellites or SSRs have equally been identified through RNA-seq analysis (Mohd-Shamsudin et al., 2013; Zeng et al., 2013; Baranski et al., 2014). On those cases, though, a wider coverage of the reference genome is suggested (Qian et al., 2014) once the presence of highly repetitive regions could stand as a limiting factor by compromising the transcriptome assembly.

As could be observed, RNA-seq has been considered a solid method for the large-scale gene expression analysis due to the fact it does not require prior genome knowledge (Wang et al., 2009) and enables the detection of isoforms arising from alternative splicing (Ghosh and Qin, 2010). Even when involving several samples, such technique is accessible with moderate costs. In this case multiplex

runs containing up to 10 samples per sequencing lane can be performed in some platforms and the costs are no longer a limiting factor.

Another RNA-seq advantage is its wide dynamic range (ratio between the minimum and maximum expression level). This feature makes it suitable for measuring low, medium and high expression levels of the genes, not requiring very sophisticated normalization. By contrast, DNA microarrays show reliable results only for medium expression levels and therefore have a much smaller dynamic range. Thus, RNA-seq provides much more informative data, requiring less biological material and lower costs, becoming this technique popular for measuring gene expression on a large scale (Sharov et al., 2004; Wang et al., 2009).

RNA-seq APPLICATION WITHIN PENAID SHRIMP AQUACULTURE

The use of RNA-seq in species of penaeid shrimp can be focused on transcriptome characterization, functional annotation, gene expression profiles analysis, and gene-associated markers identification. In this section, the emphasis is given to the analysis of differential expression, identification of molecular markers, and its potential to promote genetic gain and development of improved penaeid strains. Similar studies have allowed the identification of candidate genes or quantitative trait loci (QTLs), which could be related to traits of interest for aquaculture, such as reproduction, sex determination, growth, immunity, and tolerance against environmental stress. Data involving pathways are also relevant in order to obtain more details about the interaction mechanism between the expressed products and their importance and applicability.

IDENTIFYING CANDIDATE GENES THROUGH DIFFERENTIAL EXPRESSION ANALYSIS

Although the applicability of RNA-seq in transcriptome and differential expression in aquatic organisms have increased in the past 3 years, the results found in literature and in the Sequence

Read Archive Database of the National Center for Biotechnology Information (SRA-NCBI) indicate that such approach still is incipient for penaeid shrimp. SRA databank, per instance, presents only 28 deposits of data generated by NGS in what regards the species *L. vannamei*, *L. stylirostris*, and *Penaeus monodon* (<http://www.ncbi.nlm.nih.gov/sra/?term=penaeidae>). In literature, it has been found next-gen data only for *P. monodon* (Baranski et al., 2014) *Fenneropenaeus chinensis* (Li et al., 2013) and *L. vannamei* (Li et al., 2012; Guo et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Yu et al., 2014). Mostly, the research in this field has been covering the identification of genes connected to immunity, mainly concerning the white spot syndrome virus (WSSV) and the taura syndrome virus (TSV; Li et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014). Both syndromes have caused great economic losses for the shrimp industry throughout the past few decades.

Despite the fact that crustaceans do not own an immune system, some candidate genes have been obtained from hemolymph and hepatopancreas tissues. It is clearly seen in some differential expression studies concerning *L. vannamei* species, which represents the biggest portion of the marine shrimp worldwide production (Gucic et al., 2013). Among the main genes studied, it is possible to find those related to toll-like and signaling receptors, apoptosis, *Vibrio cholerae* infection and other immune proteins (e.g., phagosome, hemocyanin, crustacyanin, antiviral), antioxidant enzymes (the peroxidases and glutathione ones), and lectins (Figure 1; Li et al., 2012, 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014; Yu et al., 2014).

Data related to toll-like and lectin proteins demonstrate that those may act as signaling molecules, what causes the increase of peptides expression responsible for controlling the immune response (Wang et al., 2014). On the other hand, genes associated with apoptosis may indicate an attempt to prevent proliferation of viruses and possible damages to genetic material, through death of infected cells. Specifically in what regards large number of proteins related to infection by *V. cholerae* response is due to the recurring presence of this group of bacteria within shrimp farming tanks (Banerjee et al., 2012).

Information regarding the main metabolic pathways and the quantity of most frequent genes in each pathway were also collected, as part of the data obtained via functional annotation for RNA-seq. In penaeid, the most commonly described pathways were those involving the general metabolism, spliceosome, RNA transport, *V. cholerae* infection, phagosome and the antioxidant ones, which include peroxidase enzymes (Li et al., 2012, 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Yu et al., 2014). Spliceosome and RNA transport pathway supposedly act in new transcripts formation, providing genetic variants that may contribute to resistance (Yang et al., 2007).

Regarding Gene Ontology (GO) categories, all studies in penaeid have mainly reported the same data. Considering the biological processes, per instance, the most frequent were metabolism and biological regulation. In what regards cellular components, genes are mostly expressed at the cell and some unspecific organelles. Finally, concerning the molecular function, the most

common ones were catabolic activity and binding (Li et al., 2012, 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014; Yu et al., 2014). Overall results such as these were expected, since the penaeid species previously mentioned herein have too little information about their genomes available. In the case of *L. vannamei*, only approximately 12,000 gene products were described, which may be useful in a comparative approach concerning a *de novo* assembly (<http://www.ncbi.nlm.nih.gov/protein/?term=Litopenaeus+vannamei>).

IDENTIFYING GENE ASSOCIATED MARKERS

In what regards RNA-seq technology, it has also been proving to be an extremely useful tool for identifying SNPs, which may be also used to develop high density SNPs chips for studies concerning the genome wide association (GWAS) and to build high density linkage maps (Baranski et al., 2014; Yu et al., 2014). Furthermore, SNPs can be used as markers in order to distinguish allelic transcripts whilst studying the allele-specific expression (Bell and Beck, 2009).

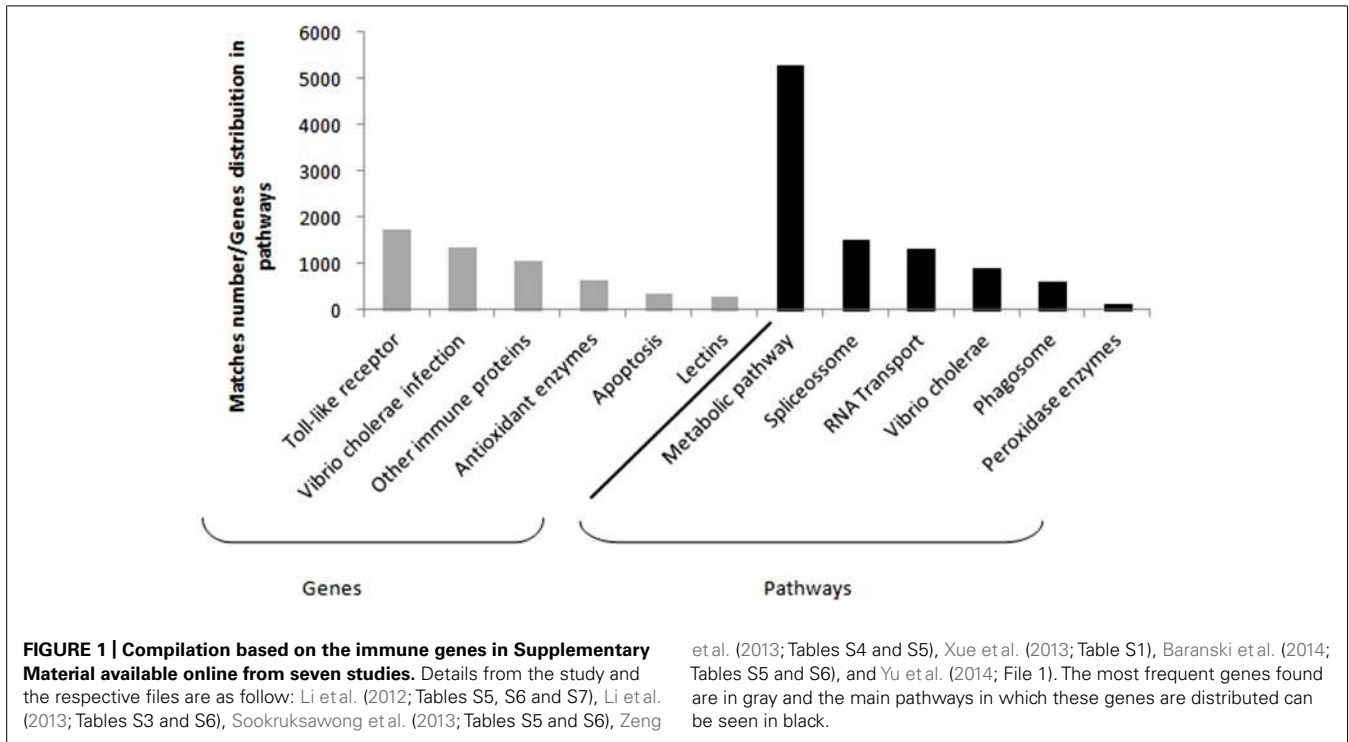
In a recent study, Yu et al. (2014) prospected SNPs in *L. vannamei*. A total of 58,717 unigenes and 36,277 high quality SNPs were predicted by transcriptomes “M” (produced by the authors themselves) and “P” (downloaded from SRA database, session number SRR346404, which was published by Li et al., 2012), respectively. Those SNPs were spread out among 25,071 unigenes and allocated to 254 pathways at the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. The main pathways containing high number of SNPs were metabolic pathways, amoebiasis, *V. cholerae* infection, RNA transport, and actin cytoskeleton regulation.

Baranski et al. (2014) used the approach to build a high density linkage map in *P. monodon*. A total of 6,000 out of 473,620 SNPs/indels putative were genotyped by using the Illumina iSelectCerca genotyping matrix. Out of those SNPs, 3,959 were mapped in 44 linking groups and out of those 2,340 were functionally annotated according to the GO database (see dataset S5 and S6 from Baranski et al., 2014). According to the authors, these polymorphisms may be causal or closely related to other mutations that affect important traits, such as resistance to diseases and reproductive performance.

The identification and functional annotation of SNPs identified by Baranski et al. (2014) and Yu et al. (2014) studies represent a useful resource to comprehend mechanisms determining complex traits and, consequently, to develop programs aiming the genetic improvement of these characters in penaeid shrimp strains. That taken, those SNPs can be applied both in marker assisted selection (MAS), using SNPs closely associated with QTL, and in genomic selection, through complete set of identified SNPs. That increases the rate of genetic gain per generation in traits of great interest to the shrimp industry, such as growth and resistance to disease.

FINAL CONSIDERATIONS

One of the possible challenges that arise within genetic gain is the development of penaeid strains that may simultaneously present high growth development and pathogens resistance.



Genetic correlation studies have shown that there is a negative phenotypic correlation between the resistance to diseases and the weight gained by the animals (Argue et al., 2002; Gitterle et al., 2005; Cock et al., 2009). Cock et al. (2009), reinforce the fact specimen potentially resistant to WSSV also present low reproduction efficiency. Such remarks suggest that genes with pleiotropic effects may be responsible for the trade-off observed between these traits in penaeid shrimp. From this perspective, the RNA-seq technique can be used for discovering such genes, since the overlapping of differentially expressed genes in both strains resistant to pathogens and in large weight gain strains can also be verified. Therefore, up-regulated or down-regulated genes expressed in these two strains could indicate a possible pleiotropic effect. Besides, mRNA studies allied with RNA-seq method could also be used for micro RNAs (miRNA) analyses. This approach was applied in aquaculture species, such as freshwater prawn *Macrobrachium rosenbergii* (Tan et al., 2013), and tilapia (Huang et al., 2012). As a result, it has been shown that miRNAs are critical regulators of generalized cellular functions such as differentiation, proliferation, and cell growth.

Another challenge within aquaculture is the difficulty in achieving sexual maturity and spawning of penaeid species (except for *L. vannamei*), under the farming conditions (Lo et al., 2007; Brady et al., 2013). As an attempt to overcome such problem, ablation of the eyestalk has been conducted for many years. Nevertheless, such practice is associated to high mortality rates, and low spawning and survival rates (Huberman, 2000). Considering this, the transcriptome analysis obtained from reproductive organs of native and captive specimens of penaeid shrimp via RNA-seq may significantly contribute to the

identification of the underlying causes of reproductive dysfunction observed in farmed animals. Furthermore, the discovery of genes involved in gonadal maturation and reproductive performance may assist in gametogenesis, handling studies involving these species.

Finally, transcriptome and differential expression analysis by RNA-seq may be a powerful approach to optimize the penaeid diet composition (nutrigenomics), especially for those species that do not count on a specific availability of diet. The approach may be used to identify specific changes in molecular level (Chávez-Calvillo et al., 2010), which in turn also cause metabolic and physiological changes in shrimp treated with different diets (e.g., levels of crude protein, levels of plant protein inclusion and of antioxidants, vitamins, and polyunsaturated fatty acids). Thus, nutrigenomics can be used to produce healthy animals and safe and high quality products for the consumer, emerging as a promising area of research for sustainability and profitability in aquaculture (Cerdà and Machado, 2013).

Although NGS technologies are showing their efficiency in works related to gene expression, other methodologies such as third-generation sequencing, also referred to as single-molecule sequencing (Single-Molecule Real-Time, SMRT), are being developed, but already showing limitations. More advanced techniques of sequencing are also on the way, such as “next-next-generation,” which it is capable of handling millions of DNA molecules simultaneously, including cDNAs from the RNAs.

Considering the many technologies that are already available or emerging, researchers can only venture in this world of possible and promising technologies. Various research groups

should seek to unite efforts in order to overcome the difficult and challenging task of applying the enormous potential of these new methods to advance and progress in penaeid shrimp aquaculture.

ACKNOWLEDGMENTS

The authors would like to thank the Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP) for the financial support (2012/17322-8).

REFERENCES

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106
- Argue, B. J., Arce, S. M., Lotz, J. M., and Moss, S. M. (2002). Selective breeding of Pacific White Shrimp (*Litopenaeus vannamei*) for growth and resistance to Taura Syndrome Virus. *Aquaculture* 204, 447–460. doi: 10.1016/S0044-8486(01)00830-4
- Banerjee, S., Ooi, M. C., Shariff, M., and Khatoon, H. (2012). Antibiotic resistant Salmonella and Vibrio associated with farmed *Litopenaeus vannamei*. *Sci. World J.* 2012, 6. doi: 10.1100/2012/130136
- Baranski, M., Gopikrishna, G., Robinson, N. A., Katneni, V. K., Shekhar, M. S., Shanmugakarthik, J., et al. (2014). The development of a high density linkage map for Black Tiger Shrimp (*Penaeus monodon*) based on cSNPs. *PLoS ONE* 9:e85413. doi: 10.1371/journal.pone.0085413
- Bell, C. G., and Beck, S. (2009). Advances in the identification and analysis of allele-specific expression. *Genome Med.* 1, 56. doi: 10.1186/gm56
- Brady, P., Elizur, A., Cummins, S. F., Ngyuen, N. H., Williams, R., and Knibb, W. (2013). Differential expression microarrays reveal candidate genes potentially associated with reproductive dysfunction of captive-reared prawn *Penaeus monodon*. *Aquaculture* 400–401, 14–28. doi: 10.1016/j.aquaculture.2013.02.038
- Cerdà, J., and Machado, M. (2013). Advances in genomics for flatfish aquaculture. *Genes Nutrition* 8, 5–17. doi: 10.1007/s12263-012-0312-318
- Chávez-Calvillo, G., Perez-Rueda, E., Lizama, G., Zúñiga Aguilar, J. J., Gaxiola, G., Cuzon, G., et al. (2010). Differential gene expression in *Litopenaeus vannamei* shrimp in response to diet changes. *Aquaculture* 300, 137–141. doi: 10.1016/j.aquaculture.2009.11.027
- Cock, J., Gitterle, T., Salazar, M., and Rye, M. (2009). Breeding for disease resistance of penaeid shrimps. *Aquaculture* 286, 1–11. doi: 10.1016/j.aquaculture.2008.09.011
- Gavry, M. R., and Roberts, S. B. (2012). Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*. *Comp. Biochem. Physiol. Part D* 7, 94–99. doi: 10.1016/j.cbd.2011.12.003
- Ghosh, D., and Qin, Z. S. (2010). Statistical issues in the analysis of ChIP-Seq and RNA-Seq data. *Genes (Basel)* 1, 317–334. doi: 10.3390/genes1020317
- Gitterle, T., Salte, R., Gjerde, B., Cock, J., Johansen, H., Salazar, M., et al. (2005). Genetic (co)variation in resistance to White Spot Syndrome Virus (WSSV) and harvest weight in *Penaeus (Litopenaeus) vannamei*. *Aquaculture* 246, 139–149. doi: 10.1016/j.aquaculture.2005.02.011
- Gucic, M., Cortés-Jacinto, E., Civera-Cerecedo, R., Ricque-Marie, D., and Martínez-Córdova, L. R. (2013). Apparent carbohydrate and lipid digestibility of feeds for whiteleg shrimp, *Litopenaeus vannamei* (Decapoda: Penaeidae), cultivated at different salinities. *Rev. Biol. Trop.* 61, 1201–1213. doi: 10.15517/rbt.v61i3.11935
- Guo, H., Ye, C.-X., Wang, A.-L., Xian, J.-A., Liao, S.-A., Miao, Y.-T., et al. (2013). Transcriptome analysis of the Pacific White Shrimp *Litopenaeus vannamei* exposed to nitrite by RNA-Seq. *Fish Shellfish Immunol.* 35, 2008–2016. doi: 10.1016/j.fsi.2013.09.019
- Howe, G. T., Yu, J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., et al. (2013). A SNP resource for douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics* 14:137. doi: 10.1186/1471-2164-14-137
- Huang, C. W., Li, Y. H., Hu, S. Y., Chi, J. R., Lin, G. H., Lin, C. C., et al. (2012). Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (*Oreochromis niloticus*). *J. Anim. Sci.* 90, 4266–4279. doi: 10.2527/jas.2012-5142
- Huberman, A. (2000). Shrimp endocrinology. *Rev. Aquac.* 191, 191–208. doi: 10.1016/S0044-8486(00)00428-2
- James, R., Thampuran, N., Lalitha, K. V., Rajan, L. A., and Joseph, T. C. (2010). Differential gene expression profile of the hepatopancreas of White Spot Syndrome Virus infected Fenneropenaeus indicus by Suppression Subtractive Hybridization. *Fish Shellfish Immunol.* 29, 884–889. doi: 10.1016/j.fsi.2010.07.029
- Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comp. Biol.* 8:e1002375. doi: 10.1371/journal.pcbi.1002375
- La Vega, E., Degnan, B. M., Hall, M. R., and Wilson, K. J. (2007). Differential expression of immune-related genes and transposable elements in Black Tiger Shrimp (*Penaeus monodon*) exposed to a range of environmental stressors. *Fish Shellfish Immunol.* 23, 1072–1088. doi: 10.1016/j.fsi.2007.05.001
- Li, C., Weng, S., Chen, Y., Yu, X., Lü, L., Zhang, H., et al. (2012). Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. *PLoS ONE* 7:e47442. doi: 10.1371/journal.pone.0047442
- Li, S., Zhang, X., Sun, Z., Li, F., and Xiang, J. (2013). Transcriptome analysis on Chinese Shrimp Fenneropenaeus chinensis during WSSV acute infection. *PLoS ONE* 8:e58627. doi: 10.1371/journal.pone.0058627
- Lo, T. S., Cui, Z., Mong, J. L. Y., Wong, Q. W. L., Chan, S. -M., Kwan, H. S., et al. (2007). Molecular coordinated regulation of gene expression during ovarian development in the penaeid shrimp. *Mar. Biotechnol.* 9, 459–468. doi: 10.1007/s10126-007-9006-4
- Marguerat, S., and Bähler, J. (2010). RNA-Seq: from technology to biology. *Cell. Mol. Life Sci.* 67, 569–579. doi: 10.1007/s00018-009-0180-6
- Mohd-Shamsudin, M. I., Kang, Y., Lili, Z., Tan, T. T., Kwong, Q. B., Liu, H., et al. (2013). In-depth transcriptomic analysis on giant freshwater prawns. *PLoS ONE* 8:e60839. doi: 10.1371/journal.pone.0060839
- O'Neil, S. T., and Emrich, S. J. (2013). Assessing de novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* 14:465. doi: 10.1186/1471-2164-14-465
- Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-Seq reads to differential expression results. *Genome Biol.* 11, 220. doi: 10.1186/gb-2010-11-12-220
- Qian, X., Ba, Y., Zhuang, Q., and Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *OMICS* 18, 1–13. doi: 10.1089/omi.2013.0110
- Rojtinnakorn, J., Hirono, I., Itami, T., Takahashi, Y., and Aoki, T. (2002). Gene expression in haemocytes of Kuruma Prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. *Fish Shellfish Immunol.* 13, 69–83. doi: 10.1006/fsim.2001.0382
- Sharov, V., Kwong, K. Y., Frank, B., Chen, E., Hasseman, J., Gaspard, R., et al. (2004). The Limits of Log-Ratios. *BMC Biotechnol.* 4:3. doi: 10.1186/1472-6750-4-3
- Sookruksawong, S., Sun, F., Liu, Z., and Tassanakajon, A. (2013). RNA-Seq analysis reveals genes associated with resistance to Taura Syndrome Virus (TSV) in the Pacific White Shrimp *Litopenaeus vannamei*. *Dev. Comp. Immunol.* 41, 523–533. doi: 10.1016/j.dci.2013.07.020
- Tan, T. T., Chen, M., Harikrishna, J. A., Khairuddin, N., Shamsudin, M., Maizatul, I., et al. (2013). Deep parallel sequencing reveals conserved and novel miRNAs in gill and hepatopancreas of Giant Freshwater Prawn. *Fish Shellfish Immunol.* 35, 1061–1069. doi: 10.1016/j.fsi.2013.06.017
- Wang, X.-W., Xu, J.-D., Zhao, X.-E., Vasta, G. R., and Wang, J.-X. (2014). A shrimp C-type lectin inhibits proliferation of the hemolymph microbiota by maintaining the expression of antimicrobial peptides. *J. Biol. Chem.* 289, 11779–11790. doi: 10.1074/jbc.M114.552307
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Xue, S., Liu, Y., Zhang, Y., Sun, Y., Geng, X., and Sun, J. (2013). Sequencing and de novo analysis of the hemocytes transcriptome in *Litopenaeus vannamei* response to White Spot Syndrome Virus infection. *PLoS ONE* 8:e76718. doi: 10.1371/journal.pone.0076718
- Yang, L., Park, J., and Graveley, B. R. (2007). Splicing from the outside in. *Mol. Cell.* 27, 861–862. doi: 10.1016/j.molcel.2007.09.003

- Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., et al. (2014). SNP discovery in the transcriptome of White Pacific Shrimp *Litopenaeus vannamei* by next generation sequencing. *PLoS ONE* 9:e87218. doi: 10.1371/journal.pone.0087218
- Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., et al. (2013). Transcriptome analysis of Pacific White Shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura Syndrome Virus (TSV) experimental infection. *PLoS ONE* 8:e57515. doi: 10.1371/journal.pone.0057515

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 June 2014; accepted: 11 August 2014; published online: 28 August 2014.

Citation: Santos CA, Blanck DV and de Freitas PD (2014) RNA-seq as a powerful tool for penaeid shrimp genetic progress. *Front. Genet.* 5:298. doi: 10.3389/fgene.2014.00298

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Santos, Blanck and de Freitas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Appearance traits in fish farming: progress from classical genetics to genomics, providing insight into current and potential genetic improvement

Nelson Colihueque^{1*} and Cristian Araneda²

¹ Laboratorio de Biología Molecular y Citogenética, Departamento de Ciencias Biológicas y Biodiversidad, Universidad de Los Lagos, Osorno, Chile

² Laboratorio de Biotecnología y Genética Aplicada a la Acuicultura, Departamento de Producción Animal, Facultad de Ciencias Agronómicas, Universidad de Chile, Santiago, Chile

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Michelle Martinez-Montemayor, Universidad Central del Caribe, Puerto Rico
Antti Kause, MTT Agrifood Research Finland, Finland

*Correspondence:

Nelson Colihueque, Laboratorio de Biología Molecular y Citogenética, Departamento de Ciencias Biológicas y Biodiversidad, Universidad de Los Lagos, Avenida Alcalde Fuchslocher 1305, Casilla 933, Osorno, Chile
e-mail: ncolih@ulagos.cl

Appearance traits in fish, those external body characteristics that influence consumer acceptance at point of sale, have come to the forefront of commercial fish farming, as culture profitability is closely linked to management of these traits. Appearance traits comprise mainly body shape and skin pigmentation. Analysis of the genetic basis of these traits in different fish reveals significant genetic variation within populations, indicating potential for their genetic improvement. Work into ascertaining the minor or major genes underlying appearance traits for commercial fish is emerging, with substantial progress in model fish in terms of identifying genes that control body shape and skin colors. In this review, we describe research progress to date, especially with regard to commercial fish, and discuss genomic findings in model fish in order to better address the genetic basis of the traits. Given that appearance traits are important in commercial fish, the genomic information related to this issue promises to accelerate the selection process in coming years.

Keywords: skin pigmentation, body shape, appearance traits, fish farming, quantitative trait loci

INTRODUCTION

Over the past few decades, body shape and skin pigmentation have become valuable appearance traits in commercial fish (Table 1). Due to increasing market sophistication, fish size, meat quality, and other traditional traits are not the only attributes that influence consumer choice at point of sale, especially when fish are sold whole.

On the subject of skin pigmentation, previous work conducted in other livestock species has demonstrated that proper handling of pigmentation traits allows for response to consumer demands for various food products, such as skin color in pigs and egg shell, yolk, and skin color in chickens (see review by Hudon, 1994). This topic is relevant for producers because the color of a food product is a quality attribute for the consumer. For example, consumers perceive redder salmon filets as being fresher, better-tasting, and higher quality as compared with paler salmon, and, therefore, they are willing to pay significantly more for the product (Anderson, 2000; Alfnes et al., 2006). Therefore, to satisfy modern market demands and increase profitability, producers are forced to manage external traits more intensively on an industrial scale, in particular body shape and skin color.

However, this is not an easy task, because body shape and skin color in fish are complex traits, involving numerous genetic and environmental factors. Thus, progress in this field will depend in part on dissecting the underlying genetics of these traits for future implementation of modern selection strategies, such as marker-assisted selection based on molecular data.

In commercial fish, such as the common carp, tilapia, sea bream, and salmonids (Pillay and Kutty, 2005), this strategy will complement progress made to date based solely on breeding values estimated with phenotypic and genealogical information or classical genetics which, for example, has enabled the development of new strains (Figure 1).

Further understanding of this issue may be gained from progress achieved in model and ornamental fish, where characterization of the inheritance mode of mutation, genes, and quantitative trait loci (QTL) for external traits is more advanced.

In this review, we describe efforts made to improve the external traits in commercial fish based on classical genetic approaches, as well as recent progress in genomics, the latter initially aimed at identifying the specific region that harbors genes controlling quantitative traits. This information, together with data available on this issue in model fish, will enhance progress in this field, an objective of tremendous importance for producers who need to increase the competitiveness of their cultures by managing external characteristics that give added value to cultured fish.

FISH BODY SHAPE

In common carp, one of the first domesticated fish in the world (Balon, 2004), the long process of domestication has produced a domesticated phenotype very different from the wild-type phenotype (Ankorion et al., 1992; Zhang et al., 2013). Many of these changes have arisen due to intentional selection of traits, but it is equally true that many traits are the result of an unintentional selection process. This phenomenon emerged as a result of the

Table 1 | Examples of appearance traits for body shape and skin pigmentation used in fish farming.

Category of trait	Species	Strains	Trait characteristic	Practical application	Reference
Body shape	Common carp (<i>Cyprinus carpio</i>)	Aischgrund and Galician var. <i>wuyuanensis</i> (also called Purse red carp)	High-backed Broadly elliptical body (red skin); standard length to body height ratio: 2.3	To increase esthetics of body shape To increase esthetics, creating a desirable “wallet” shape and all-red body color	Ankorion et al. (1992) Zhang et al. (2013)
		var. <i>haematopterus</i> (also called Amur wild carp)	Spindle-shaped body (steel-gray skin); standard length to body height ratio: 3.5	To increase desirability for sport fishing (its body is elongated, making it an excellent sport fish)	Zhang et al. (2013)
	Rainbow trout (<i>Oncorhynchus mykiss</i>)	Finnish national breeding program	Slender body (i.e., low body height to length ratio; silvery skin with fewer spots)	To produce fish more visually appealing desirable for the whole carcass market	Kause et al. (2003)
Skin color	Tilapia (<i>Oreochromis niloticus</i>)	Red strains (red Stirling, red Yumbo, etc)	Red body color without signs of normal black pigmentation of wild-type fish	To add value to the final product	McAndrew et al. (1988), Moreira et al. (2005)
		Rainbow trout (<i>Oncorhynchus mykiss</i>)	Blue Back	Intense bluish back; whitish belly; reduced number of dark spots, both on the back and below the lateral line	To satisfy market demands

adaptation of fish to captive conditions, quite different than the natural environment inhabited by wild-type fish.

The domesticated phenotype phenomenon arises because the body shape of an organism results from the integration of morphological, behavioral, and physiological traits (Reid and Peichel, 2010), where different genetic and environmental pressures can lead to functional trade-offs (Reznick and Ghalambor, 2001; Ghalambor et al., 2003; Walker, 2010). This creates functional constraints, where those changes with the greatest positive and fewest negative effects on fitness will be selected (Reid and Peichel, 2010). For example, in natural populations, there is a relationship between body shape and swimming performance, but body shape is also influenced by foraging behavior, the risk of predation, and stream velocity (Webb, 1984; Walker, 1997). The trade-off for body shape also operates in captive populations. For instance, cultured populations of rainbow trout selected for rapid growth result in more rotund fish, given the existence of a positive genetic correlation of body mass with body shape and condition factor (Gjerde and Schaeffer, 1989; Kause et al., 2003); that is, mass gain in fish achieved by increasing body width and height rather than by increasing body length.

It has been shown that other factors, such as water velocity (Pakkasmaa and Piironen, 2001), rearing environment (Swain et al., 1991), fish density, and diet (Higgs et al., 1992; Einen et al., 1998; Jenkins et al., 1999) may also modify body shape in fish. This phenomenon occurs given that many morphological growth-related traits show phenotypic and genetic correlations in fish (Kause et al., 2003; Martyniuk et al., 2003), whose origins

are related to the genetic architecture of traits such as covariation in QTL location or conservation of chromosomal regions homologous across species (Reid et al., 2005).

In common carp, this process has produced various phenotypes of commercial value that are currently used in fish farming, such as the high-backed and elliptical body shape morphs, typical of the Galician and Wuyuanensis strains, respectively (Ankorion et al., 1992; Zhang et al., 2013). Even this process of domestication can reach a high level of body shape modification, such as it has been observed in the ornamental goldfish (*Carassius auratus*), where various morphological traits have been modified (e.g., body shape, fins, and eyes). Several of these modified traits can be found in the same individual, giving rise to popular strains called “monstrosities” (Balon, 1990).

Alterations of morphology characteristics, mainly body length and fins (Haffter et al., 1996) can also be seen in mutants of zebrafish (*Danio rerio*). In the guppy (*Poecilia reticulata*), variation in male body shape occurs in association with mating success (Tripathi et al., 2009). Therefore, available evidence in fish indicates that these organisms are highly amenable to morphological modification, already widely explored in ornamental as well as in model fish.

The underlying genetics of phenotypic variation is beginning to be understood in various commercial fish (Massault et al., 2009; Loukovitis et al., 2013; Zhang et al., 2013). These studies are focusing on finding significant QTL for morphometric traits based mostly on geometric morphometry; in these analyses, different types of molecular markers have been used. These investigations

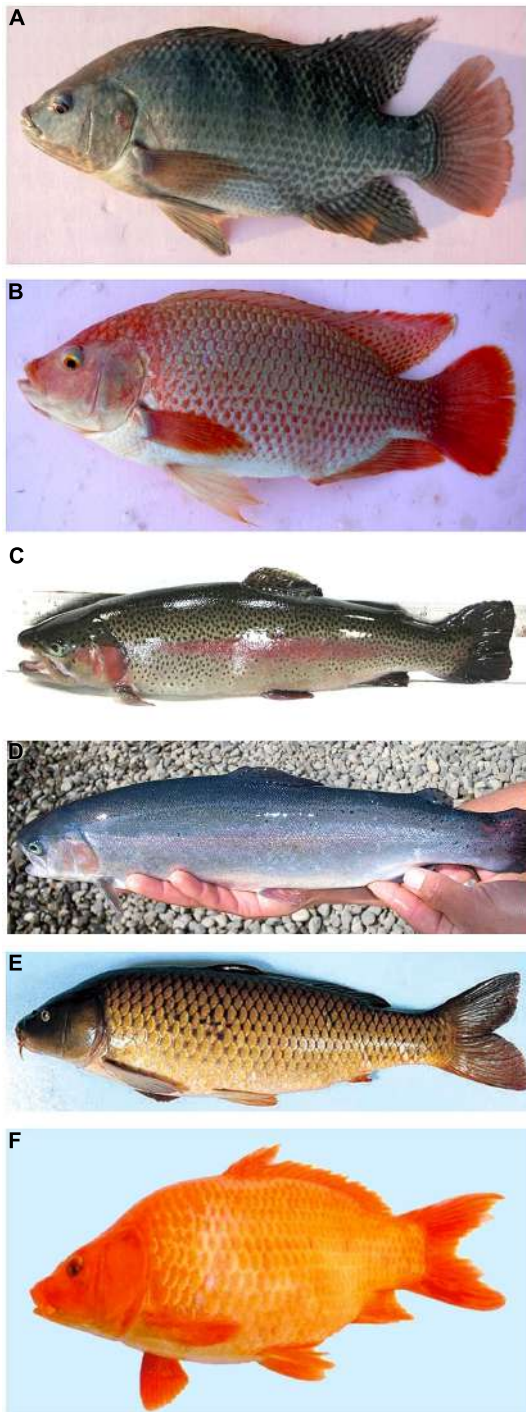


FIGURE 1 | Examples of commercial fish strains with improved skin pigmentation and body shape. (A) A wild-type tilapia (*Oreochromis niloticus*) with normal black pigmentation; **(B)** a red strain tilapia (red Yumbo) with improved red skin pigmentation; **(C)** a wild-type rainbow trout (*Oncorhynchus mykiss*) with normal pigmentation and **(D)** a Blue Back rainbow trout with improved intense bluish back, whitish belly, and a reduced number of dark spots; **(E)** a common carp (*Cyprinus carpio*) of var. *haematopterus* (or Amur wild carp) with a spindle-shaped body and steel-gray skin color; and **(F)** a common carp of var. *wuyuanensis* with improved broadly elliptical body (red skin color).

are contributing to our understanding of the genetic architecture of divergence in body shape, by means determining the number of genes or QTLs that contribute to a particular trait, or the number of traits that a particular gene or QTL affects, i.e., the pleiotropic effect, and the location of genes or QTLs within the genome that affect body shape, along with their interaction.

Progress in model fish should be mentioned here, particularly in zebrafish (Haffter et al., 1996) where a set of dominant Mendelian loci affecting body shape and fins in induced mutants have been identified. For example, loci that affect body shape may cause a reduction of overall body length in the adult fish, due to a reduction either in the length of vertebrae (*stöpsel* mutant) or number of vertebrae (*däumling* mutant). Interestingly, mechanisms of body shape variation involving axial length modification also occur naturally across several fish species (Ward and Mehta, 2010), which indicates that this mechanism has been of evolutionary significance for body form differentiation in fish.

Recent work on QTL searching in commercial fish clearly supports the existence of major genes underlying the quantitative genetic variation of morphological and body shape-related traits (**Table 2**). In Gilthead seabream (*Sparus aurata*), Loukovitis et al. (2013), using half-sib regression analysis, found significant morphology QTLs, e.g., distances from pectoral fin to dorsal fin or from pectoral fin to anal fin (see **Table 2**), in three linkage groups (9, 21, and 25) identified at genome-wide level that explain 18.5 to 27.1% of trait variation. This result suggests the existence of one locus in each linkage group affecting several traits in this fish. Moreover, given that QTLs affecting body weight were located at the same positions for the linkage groups 9 and 21 (Loukovitis et al., 2011), the authors conclude that there might be only one pleiotropic QTL in each LG affecting overall body size. This is in accordance with the high genetic correlations ($r_G > 90\%$) observed between all traits analyzed (see **Table 2**), including body weight. These results, combined with those obtained from previous studies (Boulton et al., 2011), underline highly significant loci affecting overall morphology in *S. aurata*.

On the other hand, using half-sib regression analysis and variance component analysis at the genome-wide level in sea bass (*Dicentrarchus labrax*), six significant QTLs for a combination of morphometric traits (standard length, head length, body length, pre-anal length, abdominal length, post-anal length, head depth, body depth; see **Table 2**) on linkage groups 1B, 4, 6, 7, 15, and 24 were reported by Massault et al. (2009). These QTLs explain between 9.4 and 16% of phenotypic variance. In this study, a body weight QTL was discovered at the same linkage groups (linkage groups 4 and 6) and at similar positions as morphology QTLs, which might explain the high correlation observed between body weight and all morphometric traits studied in this fish.

Moreover, in common carp (*Cyprinus carpio*), in a primary genome-wide scan using single nucleotide polymorphisms (SNPs) and microsatellite markers, Zhang et al. (2013) found five significant QTLs for body-shape related traits (body height, body width and standard length) located at linkage groups 1, 12, and 20, which explain 20.4 to 20.7%, 18.9 to 21.1%, and 19.5% of phenotypic

Table 2 | Examples of QTLs for body shape-related traits mapped on different commercial fish genomes.

Species	Trait	Linkage group (LG) where QTLs were detected. Name of QTL (bold), linked markers, and respective phenotypic variation noted in parentheses	Reference
Gilthead seabream (<i>Sparus aurata</i>)	Standard length, body length, pectoral dorsal 1, pectoral dorsal 2, body length 2, pectoral anal 2, pectoral anal 1, dorsal fin length, belly length, body depth 2, body depth 3, head length	LG-9 (SL, BL, PecDor1, PecDors2, BL2, PectAnal1, PectAnal2 , Bd20, 18.5–23%) LG-21 (DFL, BellyL, BD2, BD3 , ELD36 and SAGT1, 18.6–27.1%) LG-25 (HL , 22.6%)	Loukovitis et al. (2013)
Sea bass (<i>Dicentrarchus labrax</i>)	Combination of morphometric traits (standard length, head length, body length, pre-anal length, abdominal length, post-anal length, head depth, body depth)	LG-1B (MORPH , 14%) LG-4 (MORPH , 13%) LG-6 (MORPH , 9.4%) LG-7 (MORPH , 16%) LG-15 (MORPH , 12%) LG-24 (MORPH , 13%)	Massault et al. (2009)
Common carp (<i>Cyprinus carpio</i> var. <i>haematopterus</i>)	Body height Body width Standard length	LG-1 (qbh1 , snp0163, 20.4%; qbw1 , snp0163, 20.7%) LG-12 (qsl12 , snp0315, 21.1%; qbh12 , snp1133, 18.9%) LG-20 (qbh20 , hlj1717, 19.5%)	Zhang et al. (2013)

QTLs were detected at genome-wide level using permutation tests at a significance threshold value of $P < 0.05$ for *S. aurata* and *D. labrax*, and of $P < 0.01$ for *C. carpio*.

variance, respectively. Given that QTLs of linkage group 1 were located in the same interval, it was concluded that only one QTL produced pleiotropic effects on these traits, which was not the case for QTLs found in linkage group 12, indicating that different factors control the traits. Importantly, this study provides strong evidence that the marked body shape differences of *Cyprinus carpio* populations, in particular between *Cyprinus carpio* var. *wuyuanensis* and *Cyprinus carpio* var. *haematopterus*, depend on quantitative genetic variations that control different body shape-related traits that may have originated through the process of selective breeding that has occurred for decades in this species.

In salmonids, another important group of commercial fish (Pillay and Kutty, 2005), progress has been made in this field through QTLs searching, mainly for growth-related traits, including fork length, body weight, and Fulton's condition factor, and also for meristic traits (for review see Araneda et al., 2008). For example, in Atlantic salmon (*Salmo salar*), four QTLs for condition factor and two for body weight were detected in comparative studies with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). One strong QTL explaining 20.1% of variation in body weight was found on linkage group AS-8, while another QTL with a strong effect on condition factor accounting for 24.9% of trait variation was found on linkage group AS-14. This result suggests that a significant portion of quantitative variation in body weight and condition factor in this species is under the control of a few QTLs with relatively large effects.

However, to date no study has been specifically undertaken to search QTLs in salmonids based on the geometric morphology method. It is noteworthy that in some species of this group, such as rainbow trout, a marked intra- and inter-population differentiation in body shape has been observed (Kause et al., 2003; Hecht et al., 2012; Pulcini et al., 2013). For example, Pulcini et al. (2013), in a common-garden experiment, found a marked morphological variation in body shape traits such as body profile, head length, dorsal and anal fin length, and caudal peduncle size, using geometric morphometry among wild, semi-wild and domestic lines of this species. Domestic lines have a deeper body profile, with longer dorsal and anal fins and shorter and deeper caudal peduncles than wild lines. This differentiation, attributed to exposure of domestic lines to captive conditions, suggests that the variations may result from fixed genetic differences among lines due to the existence of QTL. Therefore, further QTL analysis in rainbow trout would be useful in clarifying the underlying genetics of this striking differentiation in body shape. To achieve this goal the use of SNPs it is possible given that these markers are considered to be the most desirable molecular markers for developing high-density genome scan to discover and locate target genes underlying the quantitative traits (Wang et al., 1998). This approach has been demonstrated to be efficient to discover several QTLs in guppy that control the complex patterns of skin pigmentation of males (Tripathi et al., 2009). However, it needs to use next-generation sequencing analysis to discover

thousands of such SNPs (Miller et al., 2007; Davey et al., 2011) and in several cases developing SNP chips to perform genome-wide scans.

SKIN PIGMENTATION

The cellular basis of skin pigmentation in fish is well known. Skin color depends on five types of pigment cells (or chromatophores) known as melanophores, xanthophores, erythrophores, iridophores, and leucophores, each producing a different color (black or brown, yellow or orange, red, iridescent, blue, silver or gold, and white, respectively; Fujii, 1969, 1993). The underlying genetics of skin pigmentation phenotype, however, has been explored mostly for qualitative traits by means of large-scale analyses of natural or induced color mutants, mainly in model fishes (see review by Colihueque, 2010). Evidence from studies in these and in other fish species indicates that the inheritance mode of qualitative traits for skin pigmentation has a simple genetic

basis (Tave, 1986), i.e., a monogenetic control, which may be recessive, completely/incompletely dominant, and co-dominant or sex-linked. Moreover, these studies indicate that several genes participate in producing a specific skin color or color pattern that may be involved in chromatophore development, pigment synthesis, and pigment expression. For example, about 90 and 40 genes of this type have been identified in zebrafish and medaka (*Oryzias latipes*), respectively, that control specification, proliferation, survival, differentiation, and distribution of chromatophores, among other processes.

However, recent studies emphasize that skin pigmentation in fish can also possess a more complex genetic architecture, characterized by specific genome regions that harbor genes controlling quantitative traits (Table 3). For example, in the threespine sticklebacks (*Gasterosteus aculeatus*), two significant QTLs on linkage groups 1 and 6 that control the degree of barring and explain 26.6% of the variance of the trait were found

Table 3 | Examples of QTLs for skin pigmentation traits mapped on different fish model genomes.

Species	Trait	Linkage group (LG) where QTLs were detected. Name of QTL (bold), linked markers, and respective phenotypic variation noted in parentheses	Reference
Threespine sticklebacks (<i>Gasterosteus aculeatus</i>)	Degree of barring	LG-1 (barring , chrI:3310077, 6.6%)	Greenwood et al. (2011)
	Degree of melanization	LG-1 (melanization , chrI:4816374, 11.7%) LG-6 (barring ,	
	Number of dorsal melanophores	chrVI:15780594, 20%)	
	Number of ventral melanophores	LG-1 (ventral melanophores , chrI:21909727, 8.9%) LG-7 (dorsal melanophores , chrVII:1728753, 11.6%)	
Guppy (<i>Poecilia reticulata</i>)	Dorsal fin black area	LG-1, LG-2, LG-4, LG-9, LG-12 (Dorsal fin black area ,	Tripathi et al. (2009)
	Dorsal fin orange area	22.9%)	
	Central blue white spot	LG-1, LG-2, LG-4, LG-6, LG-7, LG-16 (Dorsal fin orange	
	Anterior main black stripe	area , 26.8%)	
	Anterior orange spot	LG-2, LG-12, LG-17, LG-21, LG-23 (Central blue white	
	Black spot by gonopodium	spot , 26.5%)	
	Central orange spot	LG-8, LG-19 (Anterior main black stripe , 9.4%)	
	Posterior main black stripe	LG-6, LG-7, LG-9, LG-12, LG-20 (Anterior orange spot ,	
	Posterior ventral black stripe	23%)	
	Posterior orange spot	LG-01, LG-04, LG-07, LG-12, LG-16 (Black spot by	
	Hind fin lower orange area	gonopodium , 19.1%) LG-05, LG-08, LG-20, LG-22 (Central orange spot ,	

QTLs were detected at genome-wide level using permutation tests at a significance threshold value of $P < 0.05$ for *G. aculeatus* and *P. reticulata*.

(Greenwood et al., 2011). Given that these QTLs were associated with spatial variation in melanophore number (linkage group 6) and degree of melanization of melanophores (linkage group 1), this finding reveals the existence of different loci underlying variation in pigment patterns of this fish, which shows striking diversity of among freshwater (barred) and marine (unbarred) populations. Moreover, the number of dorsal and ventral melanophores is also controlled by different loci in this fish, since they were mapped to linkage group 7 and linkage group 1, respectively.

Through synteny analysis, Greenwood et al. (2011) identified the *Gja5* gene contained in the barring QTL on linkage group 6, which encodes a gap junction protein whose mutation disrupts the normal pigmentation pattern in zebrafish, in which spots form in place of the typical horizontal stripes, caused by an altered melanophore distribution (Watanabe et al., 2006). Moreover, the region on linkage group 1 that mapped QTLs associated with both barring and degree of melanization contains the *tyrosinase* gene, which encodes a key enzyme in melanin synthesis, whose mutation eliminates all pigmentation in zebrafish and medaka (Koga et al., 1995; Iida et al., 2004). Therefore, these results suggest that a few genes with large effects underlie the pigmentation pattern variation in the threespine sticklebacks.

In guppy males, a more complex control of pigmentation pattern has been observed (Tripathi et al., 2009), including a phenotype characterized by multi-colored areas with an ornamental function involved in female choice and in male mating success, and therefore, important for male fitness. In the genome of this fish, using interval mapping and the multiple-QTL model, 49 QTLs for 11 areas of pigmentation traits were found (see **Table 3**), which explain 9.4 to 26.8% of phenotypic variation in these traits. In addition, these QTLs were mapped in 19 out of 24 linkage groups of this species, although mainly on linkage group 12 and 4. QTLs located on linkage group 12, which corresponds to the sex chromosome of the guppy, indicate that loci responsible for polymorphisms in guppy color patterns are clustered on this chromosome. This result coincides with previous knowledge regarding physical linkage of major color pattern loci to sex chromosomes in this species (Winge and Ditlevsen, 1947; Khoo et al., 1999). In summary, the results obtained in this fish strongly suggest that multiple QTLs with minor effects contribute to each color trait in guppy males.

In commercial fish, most color phenotypes of commercial value are qualitative traits known to be under Mendelian control, such as the Red Stirling strain of tilapia (*Oreochromis niloticus*; dominant inheritance, McAndrew et al., 1988), or the iridescent metallic blue variant of rainbow trout (recessive inheritance, Kincaid, 1975); therefore, given its simple inheritance mode, they could be more easily subjected to selective breeding for new stocks with particular colors.

However, there are some rainbow trout skin pigmentation phenotypes of commercial value, such as the Blue Back (Colihueque et al., 2011) and Finnish national breeding program (Kause et al., 2003) traits with complex pigmentation patterns comprising several attributes (skin color, number, size, and position of dark spots) that vary continuously at the intrapopulation level. A substantial quantitative genetic component for the different attributes

that compose these traits has been reported (Kause et al., 2003; Díaz et al., 2011). As it has been seen in model fish, it is possible that these skin pigmentation traits may possess a complex genetic architecture, with the existence of a variable number of quantitative loci with a minor or major effect for the different trait attributes. Further analysis of these traits will clarify their particular genetic architecture.

CONCLUDING REMARKS

In farmed fish, several traits are taken into account in order to obtain a quality fish harvest suitable for marketing. These traits include body shape and skin pigmentation, both of which affect consumer acceptance of marketed fish at the point of sale. A fish with an improved appearance has greater consumer acceptance and, therefore, has a higher sale value than a fish with a normal appearance.

There has been some progress in this area with commercial fish, including traditional and new cultures, mainly through selective breeding or classical genetic analysis. This selection strategy has resulted in new fish stocks whose market participation is constantly increasing, contributing to the improved profitability of fish cultures. This trend is expected to continue over the next few years due to the sophistication of the market in many areas of the world. Therefore, there is interest in fish selection to ensure specimens that are visually appealing, for example, tilapia, rainbow trout, common carp, gilthead sea bream, and sea bass.

However, to meet this challenge, fish farmers must adapt and align their selective breeding goals with market demands. One tool that may be explored to achieve this objective derives from the discovery of QTLs or genes that underlie body shape and skin pigmentation, in which continuous variation of the different attributes that compose these traits is usually observed. This information could be used to implement selective breeding based on molecular markers tightly linked to QTLs that control various appearance traits of commercial interest, that is, marker-assisted selection. This strategy may offer a more rapid response, yielding fish with a specific external appearance to satisfy market demands.

ACKNOWLEDGMENTS

The suggestions and constructive comments of all those who helped to improve the final version of this manuscript, are gratefully acknowledged. The publication fee of this work was supported by the Departamento de Ciencias Biológicas y Biodiversidad of the Universidad de Los Lagos.

REFERENCES

- Alfnes, E., Guttormsen, A. G., Steine, G., and Kolstad, K. (2006). Consumers' willingness to pay for the color of salmon: a choice experiment with real economic incentives. *Am. J. Agric. Econ.* 88, 1050–1061. doi: 10.1111/j.1467-8276.2006.00915.x
- Anderson, S. (2000). Salmon color and the consumer. *Paper Presented at: IIFET 2000 International Institute of Fisheries Economics and Trade. Proceedings of the International Institute of Fisheries Economics and Trade 2000*, Corvallis, Oregon.
- Ankorion, Y., Moav, R., and Wohlfarth, G. W. (1992). Bidirectional mass selection for body shape in common Carp. *Genet. Sel. Evol.* 24, 43–52. doi: 10.1186/1297-9686-24-1-43

- Araneda, C., Neira, R., Lam, N., and Iturra, P. (2008). "Salmonids," in *Genome Mapping and Genomics in Animals*, eds T. D. Kocher and C. Kole (Berlin: Springer-Verlag), 1–43. doi: 10.1007/978-3-540-73837-4_1
- Balon, E. K. (1990). Epigenesis of an epigeneticist: the development of some alternative concepts on the early ontogeny and evolution of fishes. *Guelph Ichthyol. Rev.* 1, 1–42.
- Balon, E. K. (2004). About the oldest domesticates among fishes. *J. Fish Biol.* 65(Suppl. A), 1–27. doi: 10.1111/j.0022-1112.2004.00563.x
- Boulton, K., Massault, C., Houston, R. D., de Koning, D. J., Haley, C. S., Bovenhuis, H., et al. (2011). QTL affecting morphometric traits and stress response in the Gilthead Seabream (*Sparus aurata*). *Aquaculture* 319, 58–66. doi: 10.1016/j.aquaculture.2011.06.044
- Colihueque, N. (2010). Genetics of skin pigmentation: clues and prospects for improving the external appearance of farmed salmonids. *Rev. Fish Biol. Fisher.* 20, 71–86. doi: 10.1007/s11160-009-9121-6
- Colihueque, N., Parraguez, M., Estay, F. J., and Díaz, N. F. (2011). Skin color characterization in rainbow trout by use of computer-based image analysis. *N. Am. J. Aquac.* 73, 249–258. doi: 10.1080/1522055.2011.581578
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012
- Díaz, N. F., Estay, F., Araneda, C., and Colihueque, N. (2011). Heritability of Blue Back skin color and correlations with growth in rainbow trout, *Oncorhynchus mykiss*. Paper Presented at: Australian Society for Fish Biology 2011 Conference. *Proceeding of the Australian Society for Fish Biology 2011 Conference*, Townsville, Australia.
- Einen, O., Waagan, B., and Thomassen, M. S. (1998). Starvation prior to slaughter in Atlantic salmon (*Salmo salar*). I. Effects on weight loss, body shape, slaughter and fillet-yield, proximate and fatty acid composition. *Aquaculture* 166, 85–104. doi: 10.1016/S0044-8486(98)00279-8
- Fujii, R. (1969). "Chromatophores and pigments," in *Fish Physiology*, Vol. III, eds W. S. Hoar and D. J. Randall (New York: Academic Press), 307–354.
- Fujii, R. (1993). Cytophysiology of fish chromatophores. *Int. Rev. Cytol.* 143, 191–255. doi: 10.1016/S0074-7696(08)61876-8
- Ghalambor, C. K., Walker, J. A., and Reznick, D. N. (2003). Multi-trait selection, adaptation, and constraints on the evolution of burst swimming performance. *Integr. Comp. Biol.* 43, 431–438. doi: 10.1093/icb/43.3.431
- Gjerde, B., and Schaeffer, L. R. (1989). Body traits in rainbow trout. II. Estimates of heritabilities and of phenotypic and genetic correlations. *Aquaculture* 80, 25–44. doi: 10.1016/0044-8486(89)90271-8
- Greenwood, A. K., Jones, F. C., Chan, Y. F., Brady, S. D., Absher, D. M., Grimwood, J., et al. (2011). The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks. *Heredity* 107, 155–166. doi: 10.1038/hdy.2011.1
- Haffter, J., Mary, O., Lin, S., Farrell, M. J., Vogelsang, E., Haas, F., et al. (1996). Mutations affecting pigmentation and shape of the adult Zebrafish. *Dev. Genes Evol.* 206, 260–276. doi: 10.1007/s004270050051
- Hecht, B. C., Thrower, F. P., Hale, M. C., Miller, M. R., and Nichols, K. M. (2012). Genetic architecture of migration-related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*. *G3 (Bethesda)* 2, 1113–1127. doi: 10.1534/g3.112.003137
- Higgs, D. A., Dosanjh, B. S., Uin, L. M., Himick, B. A., and Eales, J. G. (1992). Effects of dietary lipid and carbohydrate levels and chronic 3,5,3 α -triiodo-L-thyronine treatment on growth, appetite, food and protein utilization and body composition of immature rainbow trout, *Oncorhynchus mykiss*, at low temperature. *Aquaculture* 105, 175–190. doi: 10.1016/0044-8486(92)90129-9
- Hudon, J. (1994). Biotechnological applications of research on animal pigmentation. *Biotech. Adv.* 2, 49–69. doi: 10.1016/0734-9750(94)90290-9
- Iida, A., Inagaki, H., Suzuki, M., Wakamatsu, Y., Hori, H., and Koga, A. (2004). The *tyrosinase* gene of the β^0 albino mutant of the medaka fish carries a transposable element insertion in the promoter region. *Pigment Cell Res.* 17, 158–164. doi: 10.1046/j.1600-0749.2003.00122.x
- Jenkins, T. M., Diehl, S., Kratzer, K. W., and Cooper, S. D. (1999). Effects of population density on individual growth of brown trout in streams. *Ecology* 80, 941–956. doi: 10.1890/0012-9658(1999)080[0941:EOPDOI]2.0.CO;2
- Kause, A., Ritola, O., Paananen, T., Eskelinen, U., and Mantysaari, E. (2003). Big and beautiful? Quantitative genetic parameters for appearance of large rainbow trout. *J. Fish Biol.* 62, 610–622. doi: 10.1046/j.1095-8649.2003.00051.x
- Khoo, L., Lim, T. M., Chan, W. K., and Phang, V. P. E. (1999). Linkage analysis and mapping of three sex-linked color pattern genes in the guppy, *Poecilia reticulata*. *Zool. Sci.* 16, 893–903. doi: 10.2108/zsj.16.893
- Kincaid, H. L. (1975). Iridescent metallic blue color variant in rainbow trout. *J. Hered.* 66, 100–102.
- Koga, A., Inagaki, H., Bessho, Y., and Hori, H. (1995). Insertion of a novel transposable element in the *tyrosinase* gene is responsible for an albino mutation in the medaka fish, *Oryzias latipes*. *Mol. Gen. Genet.* 249, 400–405. doi: 10.1007/BF00287101
- Loukovitis, D., Batargias, C., Sarropoulou, E., Apostolidis, A. P., Kotoulas, G., Magoulas, A., et al. (2013). Quantitative trait loci affecting morphology traits in gilthead seabream (*Sparus aurata* L.). *Anim. Genet.* 44, 480–483. doi: 10.1111/age.12027
- Loukovitis, D., Sarropoulou, E., Tsigenopoulos, C. S., Batargias, C., Magoulas, A., Apostolidis, A. P., et al. (2011). Quantitative trait loci involved in sex determination and body growth in the gilthead seabream (*Sparus aurata* L.) through targeted genome scan. *PLoS ONE* 6:e16599. doi: 10.1371/journal.pone.0016599
- Martyniuk, C. J., Perry, G. M. L., Moghadam, H. K., Ferguson, M. M., and Danzmann, R. G. (2003). The genetic architecture of correlations among growth-related traits and male age at maturation in rainbow trout (*Oncorhynchus mykiss*). *J. Fish Biol.* 63, 746–764. doi: 10.1046/j.1095-8649.2003.00188.x
- Massault, C., Hellemans, B., Louro, B., Batargias, C., Van Houdt, J. K. J., Canario, A., et al. (2009). QTL for body weight, morphometric traits and stress response in European Sea Bass *Dicentrarchus labrax*. *Anim. Genet.* 41, 337–345. doi: 10.1111/j.1365-2052.2009.02010.x
- McAndrew, B. J., Roubal, F. R., Roberts, R. J., Bullock, A. M., and McEwen, I. M. (1988). The genetics and history of red, blond and associated color variants in *Oreochromis niloticus*. *Genetica* 76, 127–137. doi: 10.1007/BF00058811
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17, 240–248. doi: 10.1101/gr.5681207
- Moreira, A. A., Marques Moreira, H. L., and Silva Hilsdorf, A. W. (2005). Comparative growth performance of two Nile tilapia (Chitralada and Red-Stirling), their crosses and the Israeli tetra hybrid ND-56. *Aquac. Res.* 36, 1049–1055. doi: 10.1111/j.1365-2109.2005.01318.x
- Pakkasmaa, S., and Piironen, J. (2001). Water velocity shapes juvenile salmonids. *Evol. Ecol.* 14, 721–730. doi: 10.1023/A:1011691810801
- Pillay, T. V. R., and Kutty, M. N. (2005). *Aquaculture Principles and Practices*. Oxford: Blackwell Publishing Ltd.
- Pulcini, D., Wheeler, P. A., Cataudella, S., Russo, T., and Thorgaard, G. H. (2013). Domestication shapes morphology in rainbow trout *Oncorhynchus mykiss*. *J. Fish Biol.* 82, 390–407. doi: 10.1111/jfb.12002
- Reid, D. P., Szanto, A., Glebe, B., Danzmann, R. G., and Ferguson, M. M. (2005). QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity* 94, 166–172. doi: 10.1038/sj.hdy.6800590
- Reid, D. T., and Peichel, C. L. (2010). Perspectives on the genetic architecture of divergence in body shape in sticklebacks. *Integr. Comp. Biol.* 50, 1057–1066. doi: 10.1093/icb/icq030
- Reznick, D. N., and Ghalambor, C. K. (2001). The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112–113, 183–198. doi: 10.1023/A:1013352109042
- Swain, D. P., Riddell, B. E., and Murray, C. B. (1991). Morphological differences between hatchery and wild populations of coho salmon (*Oncorhynchus kisutch*): environmental versus genetic origin. *Can. J. Fish. Aquat. Sci.* 48, 1783–1791. doi: 10.1139/f91-210
- Tave, D. (1986). *Genetics for Hatchery Managers*. Westport, CT: AVI Publishing Company Inc.
- Tripathi, N., Hoffmann, M., Willing, E. M., Lanz, C., Weigel, D., and Dreyer, C. (2009). Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. *Proc. Biol. Sci.* 276, 2195–2208. doi: 10.1098/rspb.2008.1930
- Walker, J. A. (1997). Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (Gasterosteidae) body shape. *Biol. J. Linn. Soc.* 61, 3–50. doi: 10.1006/bijl.1996.9999
- Walker, J. A. (2010). An integrative model of evolutionary covariance: a symposium on body shape in fishes. *Integr. Comp. Biol.* 50, 1051–1056. doi: 10.1093/icb/icq014

- Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082. doi: 10.1126/science.280.5366.1077
- Ward, A. B., and Mehta, R. S. (2010). Axial elongation in fishes: using morphological approaches to elucidate developmental mechanisms in studying body shape. *Integr. Comp. Biol.* 50, 1106–1119. doi: 10.1093/icb/icq029
- Watanabe, M., Iwashita, M., Ishii, M., Kurachi, Y., Kawakami, A., Kondo, S., et al. (2006). Spot pattern of leopard *Danio* is caused by mutation in the zebrafish *connexin41.8* gene. *EMBO Rep.* 7, 893–897. doi: 10.1038/sj.embor.7400757
- Webb, P. W. (1984). Body form, locomotion and foraging in aquatic vertebrates. *Am. Zool.* 24, 107–120.
- Winge, O., and Ditlevsen, E. (1947). Colour inheritance and sex determination in *Lebistes*. *Heredity* 1, 65–83. doi: 10.1038/hdy.1947.4
- Zhang, Y., Wang, S., Li, J., Zhang, X., Jiang, L., Xu, P., et al. (2013). Primary genome scan for complex body shape-related traits in the common carp *Cyprinus Carpio*. *J. Fish Biol.* 82, 125–140. doi: 10.1111/j.1095-8649.2012.03469.x
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 April 2014; accepted: 10 July 2014; published online: 04 August 2014.

Citation: Colihueque N and Araneda C (2014) Appearance traits in fish farming: progress from classical genetics to genomics, providing insight into current and potential genetic improvement. *Front. Genet.* 5:251. doi: 10.3389/fgene.2014.00251

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Colihueque and Araneda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Primary analysis of repeat elements of the Asian seabass (*Lates calcarifer*) transcriptome and genome

Inna S. Kuznetsova^{1,2*}, Natascha M. Thevasagayam¹, Prakki S. R. Sridatta¹, Aleksey S. Komissarov^{2,3}, Jolly M. Saju¹, Si Y. Ngoh^{1,4}, Junhui Jiang^{1,5}, Xueyan Shen¹ and László Orbán^{1,6,7*}

¹ Reproductive Genomics Group, Strategic Research Program, Temasek Life Sciences Laboratory, The National University of Singapore, Singapore, Republic of Singapore

² Institute of Cytology of the Russian Academy of Sciences, St-Petersburg, Russia

³ Theodosius Dobzhansky Center for Genome Bioinformatics, St Petersburg State University, St Petersburg, Russia

⁴ School of Biological Sciences, Nanyang Technological University, Singapore, Republic of Singapore

⁵ Agri-Food and Veterinary Authority of Singapore, Singapore, Republic of Singapore

⁶ Department of Animal Sciences and Animal Husbandry, Georgikon Faculty, University of Pannonia, Keszthely, Hungary

⁷ Department of Biological Sciences, National University of Singapore, Singapore, Republic of Singapore

Edited by:

Scott Newman, Genus plc, USA

Reviewed by:

Hendrik-Jan Megens, Wageningen University, Netherlands

Scott Newman, Genus plc, USA

*Correspondence:

Inna S. Kuznetsova and László Orbán, Reproductive Genomics Group, Temasek Life Sciences Laboratory, The National University of Singapore, 1 Research Link, Singapore 117604, Republic of Singapore

e-mail: inna@tll.org.sg;

laszlo@tll.org.sg

As part of our Asian seabass genome project, we are generating an inventory of repeat elements in the genome and transcriptome. The karyotype showed a diploid number of $2n = 24$ chromosomes with a variable number of B-chromosomes. The transcriptome and genome of Asian seabass were searched for repetitive elements with experimental and bioinformatics tools. Six different types of repeats constituting 8–14% of the genome were characterized. Repetitive elements were clustered in the pericentromeric heterochromatin of all chromosomes, but some of them were preferentially accumulated in pretelomeric and pericentromeric regions of several chromosomes pairs and have chromosomes specific arrangement. From the dispersed class of fish-specific non-LTR retrotransposon elements Rex1 and MAUI-like repeats were analyzed. They were wide-spread both in the genome and transcriptome, accumulated on the pericentromeric and peritelomeric areas of all chromosomes. Every analyzed repeat was represented in the Asian seabass transcriptome, some showed differential expression between the gonads. The other group of repeats analyzed belongs to the rRNA multigene family. FISH signal for 5S rDNA was located on a single pair of chromosomes, whereas that for 18S rDNA was found on two pairs. A BAC-derived contig containing rDNA was sequenced and assembled into a scaffold containing incomplete fragments of 18S rDNA. Their assembly and chromosomal position revealed that this part of Asian seabass genome is extremely rich in repeats containing evolutionarily conserved and novel sequences. In summary, transcriptome assemblies and cDNA data are suitable for the identification of repetitive DNA from unknown genomes and for comparative investigation of conserved elements between teleosts and other vertebrates.

Keywords: repeated DNA, Asian seabass, transcriptome, rDNA, chromosomes

INTRODUCTION

Asian seabass (*Lates calcarifer*), also known as barramundi, belongs to the family *Latidae* in the order *Perciformes*, which represents the largest order of vertebrates. Most members of the *Latidae* family are endemic fishes in Africa and the Indian and Pacific Oceans (Luna, 2008). The Asian seabass is a fast-growing, popular food fish found in tropical and sub-tropical fresh and salt waters. Their white flaky meat and the low number of Y-bones have made them a highly rated table fish. This species is a protandrous hermaphrodite: individuals typically mature as males and later reverse their sex to become females (Moore, 1979; Guiguen et al., 1993). The availability of classical aquaculture technologies made the Asian seabass potentially suitable for marker-assisted selection (MAS). Over the past 10 years, together with our two

partners in Singapore, we have been working on a MAS project of the Asian seabass. Information encoded in the genes and in the whole genome is required to increase the resolution of selection from MAS to genomic selection in order to make the approach suitable to identify minor-effect genes that control only a small proportion of a trait (Liu and Cordes, 2004; Gjedrem and Baranski, 2009; Gjedrem, 2010).

Recently, we have started the Asian seabass Genome Project in order to develop new platforms for the eventual improvement of the selection process. A substantial part of nuclear genomes of most eukaryotes is occupied by various types of repetitive DNA sequences (Britten and Kohne, 1968) that make the assembly of genomic sequences difficult (Sutton et al., 1995). On the other hand, repeated DNA elements

have numerous advantages for genomic studies. They have been extensively applied as physical chromosome markers in comparative studies for the identification of chromosomal rearrangements, the identification of sex chromosomes, chromosome evolution analysis and applied genetics (Ferreira and Martins, 2008).

Repetitive DNA sequences are classified based on genomic organization of repetitive units as dispersed and tandem. Dispersed repeats, represented by various classes of transposable elements encode for proteins which facilitate their replication and integration into nuclear genome (Devine et al., 1997). Tandem repeats are organized in tandem arrays that may be large and consist of thousands or even millions of repetitive units arranged in head-to-tail orientation (Willard, 1991). Chromosome loci rich in polymorphic satellite DNA usually show specific banding patterns and this makes them potentially useful as cytogenetic markers to discriminate individual chromosomes (Lee et al., 2005; Han et al., 2008). The presence of repeat-derived chromosome landmarks enables the identification of individual chromosomes and is a prerequisite to study structural changes accompanying evolution and speciation and to follow chromosome behavior and transmission in interspecific hybrids (Lee et al., 2005).

The processes by which satellites arise and evolve are not well understood; unequal crossing over, gene conversion, transposition and formation of extra chromosomal circular DNA were all implicated (Carone et al., 2009; Erukashvily and Ponomartsev, 2013). The Nucleus Organizer Regions (NOR) are described as highly repetitive genome sites related to the rRNA synthesis. These regions present small, active transcription sites (highly conserved during evolution) and non-transcribed spacing segments (highly variable), organized as two distinct multigene families 47S and 5S rRNA genes. Tandemly arrayed rDNA repeats are composed of hundreds to thousands copies, disturbed by dispersed elements and usually belong to separate chromosomes (Cioffi et al., 2011).

Repeat regions can cause serious problems during genome assemblies, especially for those that involve short reads produced by next generation sequencing (NGS) technologies. This is due to the fact that a fragment from a repeat region can have false overlaps with fragments from other repeat regions, resulting in the merging of unrelated regions and incorrect final assemblies. The fragments from repeat regions are identified by the number of potential overlaps each fragment has based on pairwise comparisons (Sutton et al., 1995).

In this study, we focus on the primary analysis of repeat elements of the Asian seabass transcriptome and genome using approaches based on classical cytogenetics, molecular biology and next generation sequencing technology. Through this work we have demonstrated that the transcriptome assembly and cDNA data contains repeats and it is suitable for the identification of repeat DNAs from an unknown genome and for comparative investigation of conserved repeat elements between species. Altogether thirteen repeat elements, including 5S rDNA and 18S rDNA, have been identified from the Asian seabass genome, together they constitute 8–14% of the genome. All of them produce short transcripts. Some repeats are enriched on specific

chromosomes, while others are generally distributed on the karyotype, except for B-chromosomes.

MATERIALS AND METHODS

PRIMARY CULTURE AND CHROMOSOMES PREPARATION

Asian seabass larvae of 1–2 days post-hatching (dph) age from the Marine Aquaculture Center (St John's Island, Singapore) were sacrificed by placing on ice and they were seeded into cell culture dishes in L15 or RPMI medium supplemented with 20% fetal calf serum and antibiotic and antimetabolic solution (Sigma). The primary culture was incubated at 29°C with 5% CO₂. After 1–2 weeks, the cells were incubated with 0.01% of colchicine for 5–6 h. Chromosome spreads were prepared using the method outlined by Pradeep et al. (2011). Chromomycin A3 (CMA3) and DAPI fluorochrome stainings followed the methodologies of Schmid (1980) and Schweizer (1980), respectively.

DNA AND RNA EXTRACTION

Tissue samples from brain and gonad of four female and four male individuals were collected and stored at –80°C until use. Genomic DNA (gDNA) was extracted from 1 to 2 dph larvae, as well as the liver, ovary and testis of 5 years old male and female individuals using the standard phenol-chloroform procedure (Sambrook and Russell, 2001). Total RNA was isolated using the Trizol kit (Invitrogen).

QUANTITATIVE REAL-TIME PCR (qPCR) ANALYSES

One mg gDNA-free total RNA from the ovary, testis as well as female and male brain, respectively, of adult Asian seabass individuals was reverse-transcribed using a cDNA Synthesis Kit (Qiagen) according to the manufacturer's recommendations. The cDNA samples obtained were diluted 1:10 with sterile water before their use as templates in real-time quantitative PCR (qPCR). Levels of mRNA were determined using qPCR and SYBR Green chemistry on a Stratagene Mx3000P (Agilent Technologies, USA) using elongation factor 1-alpha and ribosomal protein L8 as reference genes. The Relative Expression Software Tool was used to calculate the relative expression of target mRNA. Results are reported as mean ± standard error. Significant differences between means were analyzed with the pairwise fixed reallocation randomization test (Pfaffl et al., 2002). In all cases, a value of $p < 0.5$ was used to indicate significant differences. As the expression levels between the male and female brains were similar their values were combined.

S1-NUCLEASE TREATMENT

The procedure was modified from the protocol used to isolate Cot-1 DNA (Zwick et al., 1991; Ferreira and Martins, 2008). DNA samples (300 µL of 100–500 ng/µL genomic DNA in 0.5 M NaCl) were boiled for 30 min and the size range of fragmented DNA was checked by electrophoresis in a 1% agarose gel (0.1–5 kb). Samples of 50 µL of fragmented DNA were then denatured at 95°C for 10 min, placed into ice-cold solution for 10 s, and transferred to a 65°C water bath for re-annealing. Following 1 min of re-annealing, the samples were incubated at 37°C for 20 min with 0.5 U of S1 nuclease to permit digestion of single-stranded sequences.

PCR-BASED ISOLATION OF REPETITIVE DNAs USING SPECIFIC PRIMERS

The sequences of Rex1, 5S, and 18S rDNA were amplified directly from total gDNA by specific primers (Supplementary Table S1). Partial sequences of the MoSat_SB, GGSat_SB, and YRep_SB repeat DNAs were detected in the Asian seabass transcriptome dataset. They were PCR-amplified from total gDNA with specific primers (Supplementary Table S1). Primers were designed with PRIMER3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) software using the transcriptome and Repbase data as reference sequences. The telomere probe was obtained by PCR-amplification with (TTAGG)₅ primer.

SELECTION OF 18S rDNA-CONTAINING CLONES FROM THE BAC LIBRARY, THEIR SEQUENCING AND ASSEMBLY

The Asian seabass BAC library was a replica of that described earlier (Xia et al., 2010), except this version contained only 38,400 of the 49,152 clones reported earlier. With an average insert size of 98 kb (range: 45–200 kb), this provided a 5.4-fold coverage for the haploid genome. From the library described above, clones from one hundred 384-well plates were processed by a 3D pooling method yielding 140 pooled samples. 5S and 18S rDNA primers (Supplementary Table S1) were used for screening the above pools to identify clones containing 18S rDNA.

BAC SEQUENCING AND ASSEMBLY

BAC-end sequences were generated by Sanger sequencing using the identical BAC DNA preparations as templates with the universal primers T7 (5'-TAATACGACTCACTATAGGG-3') and pLBRP (5'-CTCGTATGTTGTGTGGAATTGTGAGCC-3'). The inserts of the two BAC clones containing 18S rDNA were sequenced both by Ion Torrent (2,584,864 reads; AIT Biotech, Singapore) and 150 bp paired-end sequencing on Illumina HiSeq (2,584,864 reads; Yourgene, Taipei). The raw reads were corrected with Quake program with parameter $k = 13$ [<http://genomebiology.com/2010/11/11/R116/abstract>]. Corrected reads from the two NGS platforms were mapped separately with Bowtie program (Langmead et al., 2009) to the human reference genome, to E. coli genome, to the vector sequence (pCC1BA), European seabass (*Dicentrarchus labrax*; http://www.ncbi.nlm.nih.gov/assembly/GCA_000180815.1/); an African cichlid (*Astatotilapia burtoni* http://www.ncbi.nlm.nih.gov/assembly/GCA_000239415.1/), zebrafish (*Danio rerio*; http://www.ncbi.nlm.nih.gov/assembly/GCF_00002035.4/), Nile tilapia (*Oreochromis niloticus*; http://www.ncbi.nlm.nih.gov/assembly/GCA_000188235.2/), Atlantic salmon (*Salmo salar*; http://www.ncbi.nlm.nih.gov/assembly/GCA_00233375.1/) and to the expected rDNA sequences. Reads that mapped only to E. coli, vector, and/or to the human genome were discarded. The remaining reads were assembled into contigs with SPAdes 2.4 assembler with default parameters (Bankevich et al., 2012). Short contigs (<200 bp) with a coverage less than 100x were discarded. The assembled contigs were compared against the nucleotide collection (nr/nt) database with BLAST on NCBI website. To assemble the scaffold drafts, two datasets were used: (1) contigs obtained by single read sequencing (Ion Torrent); and (2) contigs generated from paired-end reads (Illumina HiSeq 2000). The HiSeq-derived data was used for contig assembly,

whereas the Ion Torrent-derived data was used for contig positioning according to their physical map relative to HiSeq data (Supplementary Figure S1). For the overlapping Ion Torrent datasets we computed k-mers with the Jellyfish program (Marcais and Kingsford, 2011) that were mapped to assembled contigs with the Bowtie software (<http://www.bowtie-bio.sourceforge.net/index.shtml>). According to the computed coverage, the Ion Torrent-derived contigs were separated into three groups: those located in the left part of physical map, those located in the middle part (overlap), and the rest located in the right part. Assembled contigs have been deposited at GenBank, under the accession numbers of KF432408–KF432412.

THE PARTIAL ASIAN SEABASS TRANSCRIPTOME ASSEMBLY

The sequences used for the bioinformatic identification of repeats were obtained from the Asian seabass transcriptome assembly (Supplementary Figure S1). The brief description of the sequencing and assembly of the first batch of HiSeq data is described below. Pooled RNA samples from various organs and developmental stages of Asian seabass were sequenced on Illumina HiSeq 2000 with 2 × 100 pair-end reads. Following quality- and adapter-trimming, a total of ~487 million reads were obtained and deposited to NCBI SRA [SRP033113]. Potentially contaminant reads were removed from the dataset if they did not show any homology with seabass or zebrafish mRNA sequences, resulting in ~485 million clean reads (~242 million pairs) which were then assembled using Trinity (version R2012-06-08) to generate 459,979 contigs.

All these contigs were further assembled using CAP3, resulting in 363,785 contigs. This Transcriptome Shotgun Assembly project has been deposited at GenBank, under the accession number of GAQL00000000 (the version described in this paper, is the first version, GAQL01000000). Contigs with an average read coverage ≥ 3 were further clustered using CD-HIT-EST resulting in 182,842 contigs.

SEQUENCE ANALYSIS

All sequence comparisons between large sets of sequences were performed using standard algorithms such as BLAST (Altschul et al., 1990). To check for the presence of repeat elements the sequence sets were searched against the Repbase database (Kohany et al., 2006) using CENSOR (Ver: 4.2.28) with follow parameters: default, nofilter, minsim 0.75, show_simple, bprg blastn, mode norm and the Nile tilapia repeat collection, http://cowry.agri.huji.ac.il/DATA_SET_RM/RepeatLib.html (Shirak et al., 2009). Phylogenetic relations between species have been determined by Interactive Tree Of Life software v2.2.2. (Letunic and Bork, 2011). Comparative transcriptome sequence analysis with Repbase of different species is presented in Supplementary Table S3. For detection of microsatellites, the Tandem Repeat Finder version 2.02 software (Benson, 1999) was used with the following parameters: match: 2, mismatch: 7, delta: 7, PM: 80, PI: 10, minscore: 50, maxperiod: 300 (Supplementary Table S4).

cDNA LIBRARIES

The following five additional cDNA sequence datasets were used in the comparative analysis: Nile tilapia (*Oreochromis niloticus*;

67 Mb), three-spined stickleback (*Gasterosteus aculeatus*; 45 Mb), Japanese medaka (*Oryzias latipes*; 38 Mb), mouse (*Mus musculus*; 159 Mb) and human (*Homo sapiens*; 297 Mb). These datasets were obtained from the Ensembl genome browser (<http://www.ensembl.org/index.html>).

CLONING AND SEQUENCING

The restricted DNA fragments and PCR products were cloned into pGEM-T plasmid vector (Promega) and transfected in DH5a *E. coli* competent cells according to standard protocols (Sambrook and Russell, 2001). Positive recombinant clones were sequenced by Sanger protocol.

SOUTHERN AND DOT-BLOT HYBRIDIZATION

Genomic DNA was dot-blotted onto positively charged nylon membranes (Hybond-N++, Amersham) in a series of dilutions ranging from 50 ng to 2 µg. Filter-immobilized DNA was hybridized with PCR products of repeat DNA labeled with dig-dUTP. The plasmid and PCR product with a cloned fragment was used to produce the calibration curve which was used to estimate the proportion of the sequences in the genome. Hybridization was performed at 42°C in a solution composed of 6xSSC, 0.5% SDS, 5xDenhardt, 50 µg/ml salmon sperm, 50% formamide, 0.01 mol/L EDTA, and 20 ng/ml probe (Cardinali et al., 2000; Sambrook and Russell, 2001). Densitometric quantification from the dot-blot was performed using Gel-Pro Analyzer Version 3.1.00.00.

For Southern hybridization, samples of genomic and BAC DNA (10 µg) were completely or partially digested with various restriction enzymes (Hind III, RsaII, EcoRI, BamHI, or HinfI). The digestion products were subjected to gel electrophoresis on 1–1.5% agarose gels and southern-transferred to a Hybond-N++ nylon membrane. The hybridized DIG-labeled probe was detected with an antidigoxigenin alkaline phosphatase-conjugated antibody (Roche, Germany) according to the manufacturer's protocols.

KARYOTYPES AND FLUORESCENT *IN SITU* HYBRIDIZATION (FISH)

In order to determine the karyotype of the Asian seabass, 20 good metaphase plates were used. The classification of chromosomes followed (Levan et al., 1964). Metacentrics (M) and submetacentrics (SM) were described as two-arm chromosomes, whereas submetacentrics (ST) and acrocentrics (A) as one-arm chromosomes.

All probes were labeled with digoxigenin-11-dUTP (DIG) and biotin-16-dUTP (Roche) for FISH. The labeled nucleotides were incorporated into fragments by PCR, using M13 forward and reverse primers. The slides were denatured in 70% formamide/2xSSC for 3 min at 72°C. For each slide, 50 µl of hybridization solution (containing 1 µg of each labeled probe, 50% of formamide, 2x SSC, 10% dextran sulfate), was denatured for 10 min at 75°C and allowed to prehybridize for 1 h at 37°C. Hybridization took place for 16–18 h at 37°C. Post-hybridization washes were in 4x SSC for 5 min at 73°C and 2x SSC for 5 min at room temperature. Following a wash in PBST (PBS, 0.1% Tween), slides were incubated with a Rofamin Red-avidin (Invitrogen) and FITC-conjugated anti-DIG antibody (Roche). Finally, the slides were counterstained with DAPI

and mounted in an antifade solution (Vectashield, Vector laboratories, Burlingame, CA, USA). Images were captured with a Nikon (CCD) camera on a Zeiss/MetaMorph epifluorescence microscope and were optimized using Adobe Photoshop CS2.

RESULTS

THE KARYOTYPE OF ASIAN SEABASS SHOWED 24 PAIRS OF AUTOSOMES AND A VARIABLE NUMBER OF B-CHROMOSOMES

The karyotype of Asian seabass embryos showed a diploid number of $2n = 48$ and a karyotype formula of $1M + 1SM + 11ST + 11A$ ($FN = 37$). The individual karyotypes also contained a variable number (2–10) of additional microchromosomes, suggesting the occurrence of B-chromosomes (Figures 1A,B). When stained with CMA3, most of chromosomes showed the presence of a GC-rich block near the pericentromeric (periCEN) chromatin (Figure 1C). All chromosome spreads had at least two GC-rich B-chromosomes (Figure 1C), whereas AT-rich B-chromosomes (Figure 1B) were detected only in 60% of the chromosome spreads.

ISOLATION, CLONING AND CHROMOSOMAL LOCALIZATION OF REPETITIVE ELEMENTS FROM THE ASIAN SEABASS TRANSCRIPTOME AND GENOME

We have combined the power of three different approaches for the isolation and characterization of repetitive elements from the Asian seabass genome (Supplementary Figure S1). The first approach was digestion of genomic DNA by S1-nuclease. Out of the 12 sequences cloned, nine were unknown short

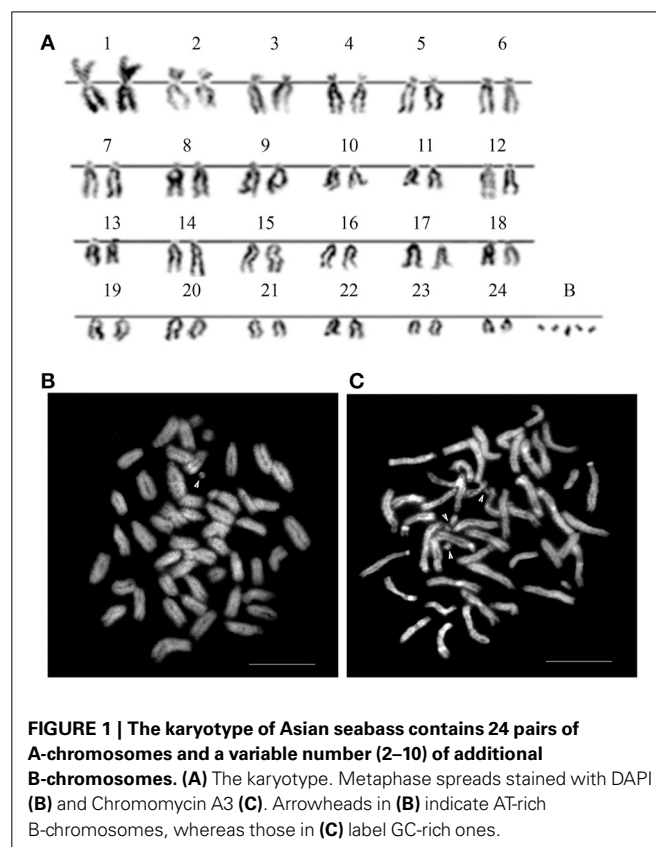


FIGURE 1 | The karyotype of Asian seabass contains 24 pairs of A-chromosomes and a variable number (2–10) of additional B-chromosomes. (A) The karyotype. Metaphase spreads stained with DAPI (B) and Chromomycin A3 (C). Arrowheads in (B) indicate AT-rich B-chromosomes, whereas those in (C) label GC-rich ones.

sequences of 100–300 bp length and not utilized for subsequent analysis. Two of the remaining three have shown similarity with OnSatB (GenBank ID:S57288) periCEN DNA of the Nile tilapia genome (Supplementary Table S1) and they were called OnSat_SB (Table 1). The divergence between the two OnSat_SB sequences was less than 1%. Southern blot hybridization of total DNA with OnSat_SB probe yielded satellite-like character ladders with steps of 200 bp length (Figure 2A). The third sequence was a short fragment of a CR1 non-long-terminal-repeat (non-LTR) retrotransposon MAUI_SB (Table 1), which showed similarity with the ORF1 part of a non-LTR retrotransposon, called MAUI from the green spotted pufferfish (*Takifugu rubripes*) genome (Poulter et al., 1999).

The second approach was based on the analysis of the draft assembly of Asian seabass transcriptome that contained a high proportion of unknown transcripts (57%). Their comparison with Repbase showed only a 10% overlap. In total, 14,283 various repeat-containing transcripts were found, their size ranged from 26 to 3600 bp (Supplementary Tables S3, S4). They could be divided into seven classes of repeats: DNA transposons, LTR retrotransposons, non-LTR retrotransposons, endogenous retroviruses, microsatellites, satellites and unclassified (Supplementary Tables S3–S5). When the relative proportion of repeats was compared among four teleost (Asian seabass, Japanese medaka, Nile tilapia and three-spined stickleback) and two mammalian (mouse and human) species, the repeat profile for the Asian seabass was the closest to that of the Japanese medaka (Figure 3A, Supplementary Table S6). While the overall percentage of repeats in the six vertebrate genomes was similar, the teleost transcriptomes contained a lower percentage of repetitive elements than the mammalian ones (Table 2).

Sequences that showed homology with characterized satDNA-like repeat elements from other fish genomes were selected from

the draft transcriptome assembly (Supplementary Tables S3, S5). One of them aligned with a mosaic satellite from the zebrafish genome (MoSat_DR; GenBank ID: DP000237), while two other fragments showed sequence homology with two different sex chromosome-specific DNAs: the first with repetitive AT-rich DNA sequences from the Y chromosome of the Mediterranean fruit fly, *Ceratitidis capitata* (YREP_CC; GenBank ID: AF115330; Supplementary Table S1) and the second with W-specific satellite DNA (GGSat; GenBank ID: X57344, Supplementary Table S1) from the chicken genome (Kawai et al., 2007). The three repeats were cloned and sequenced (Table 1); their sequences showed more than 70% of homology with the corresponding reference sequences listed above. Southern-blot hybridization of gDNA with MoSat_SB and distribution of PCR fragments yielded satellite-like ladder patterns with about 100 bp increments (Figures 2B,C). On the other hand, the PCR distribution fragments for YRep_SB and GGSat_SB were dispersed (i.e., didn't produce the expected ladder) (Figure 2C).

Among the retrotransposon elements, the Rex group is the most highly represented in the different fish species. We managed to clone two Rex1-related sequences, Rex1_SB and MAUI_SB, from the Asian seabass using specific primers (Table 1). Next, we determined the chromosomal localization of these cloned repetitive sequences. Cytogenetic mapping of these CR1 non-LTR retrotransposons showed stronger signals at the periCEN and telomeric (Tel) regions of all chromosomes and weaker ones on the chromosome arms (Figures 4A,B). The FISH signals of YREP_SB were seen at periCEN regions of all chromosomes and accumulated at 3 pairs of acrocentric chromosomes in the pretelomeric (preTel) and periCEN regions (Figure 4C), whereas those of GGSAT_SB were dispersed through all the chromosomes (Figure 4C). SatDNAs occupied the periCEN regions of all chromosomes (Figures 4D,E). OnSat_SB and MoSat_SB FISH signals

Table 1 | The inventory and characterization of nine repeats isolated from the Asian seabass transcriptome and genome.

Repeat name	NCBI access. No	Size of clones (bp)	AT%	Repeat description		
				Type	Location	Class of repeats
MAUI_SB	KC842206	164		CR1, non-LTR retrotransposon,	periCEN*, dispersed	Dispersed
Rex_SB	KC842207	551	48	CR1, non-LTR retrotransposon	periCEN	
	KC842208	560				
YREP_SB	KC842210	326	73	unknown	periCEN, dispersed	Dispersed
GGSAT_SB	KC842209	302	66	unknown	dispersed	
OnSat_SB	KC842204	233	56	satellite DNA	periCEN	Tandem
	KC842205	217				
MoSat_SB	KC842211	148	59	satellite DNA	periCEN	Tandem
	KC842212	155				
Telomere (TTAAGGG) ₅	N/A**	35	50	Tandem repeat	Telomere	
5S rDNA	KC842203	556	ND*	Ribosomal DNA	periCEN	Dispersed
18S rDNA	KC842201	306	ND*	Ribosomal DNA	periCEN	
	KC842202	756				

*periCEN, pericentromeric.

**N/A, not applicable.

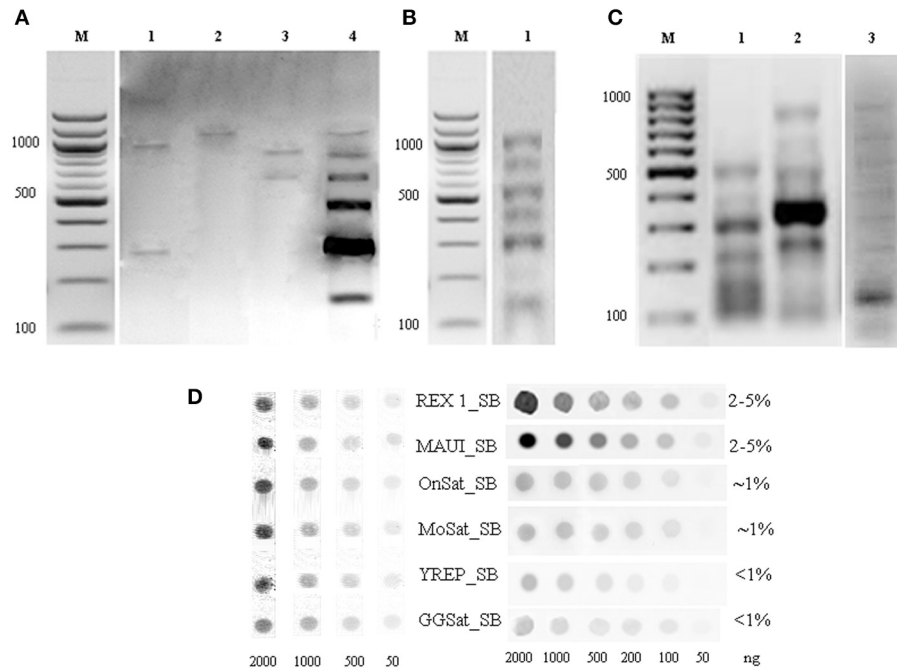


FIGURE 2 | Southern blot hybridization of Asian seabass genomic DNA (gDNA) digested by restriction endonucleases and hybridized with repeat-derived probes. (A) gDNA digested by EcoRI (1), HapI (2), BamHI (3), or RsaI (4) and hybridized with an OnSat_SB probe. **(B)** gDNA digested by HinI and hybridized with a MoSat_SB (1) probe. **(C)** PCR fragment distribution of GGSat_SB (1), YRep_SB (2), and MoSat_SB (3) amplified with sequence-specific primers indicated in

Supplementary Table S1. Probes were labeled with DIG-dUTP. M—molecular mass marker of 100bp, the corresponding molecular masses are indicated on the left. **(D)** Dot-blot hybridization of Asian seabass genomic DNA with DIG-labeled probes for six different repeats allowed for the estimation of their proportion in the genome. Genomic repeat content (%) is indicated in the right column. Samples used for the DNA standard are at the left side.

were both enriched mostly at the pericEN regions of all chromosomes (**Figure 4D**), but the latter were also detected in the preTel and pericEN regions of three pairs of acrocentric chromosomes (**Figure 4E**). All teleosts analyzed so far showed the presence of a standard telomeric repeat (TTAGGG)_n (Chew et al., 2002). In the Asian seabass, telomere signals were detected only at the ends of the chromosomes (**Figure 4F**).

AT LEAST 8% OF THE ASIAN SEABASS GENOME CONTAINS REPEATS AND SOME OF THEM SHOW DIFFERENTIAL EXPRESSION IN THE GONADS

The relative amount of OnSat_SB, MoSat_SB, YREP_SB, GGSat_SB, Rex1_SB, and MAUI_SB sequences in the Asian seabass genome was estimated using dot-blot hybridization (**Figure 2D**). The results showed that around 4–10% of the genome contained CR1-type non-LTR retrotransposons, MAUI_SB and REX1_SB, whereas about 2% consisted of two types of satDNA: OnSat_SB and MoSat_SB (**Figure 2D**). Together, the six repeat types constituted approximately 8–14% of the Asian seabass genome.

Seven repeats belonging to tandem and dispersed classes of repetitive DNA have been discovered. All of them identified matching sequence reads in the transcriptome, indicating that they produce RNA products (Supplementary Table S3) and four of them (MoSat_SB, Rex1_SB, YRep_SB, and GGSat_SB) were subjected for quantitative analysis of expression levels. While the

relative expression level of MoSat_SB and Rex1_SB in the testis was higher than the ovary ($P < 0.05$; pairwise fixed reallocation randomization test), those of YRep_SB and GGSat_SB did not exhibit such differences between the two gonads (**Table 3**). In the brain, the expression level of Rex1_SB, but not MoSat_SB, also showed an increase compared to that of the ovary.

THE GENOMIC ORGANIZATION OF TWO RIBOSOMAL DNAs SHOWED SPECIES-SPECIFIC AND EVOLUTIONARILY CONSERVED FEATURES

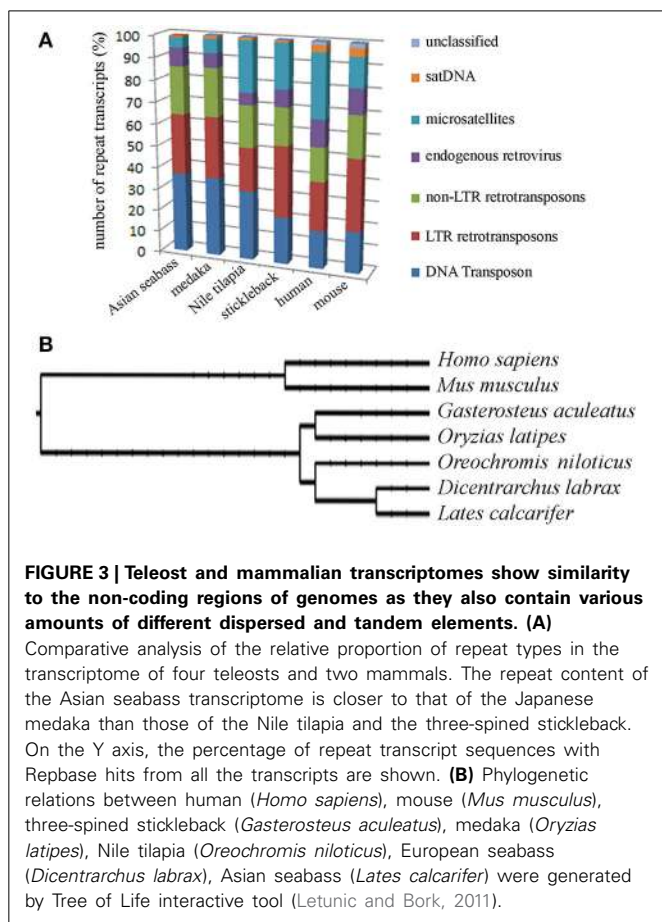
Evolutionarily conserved fragments of 5S rDNA and 18S rDNA (**Table 1**) were obtained through PCR amplification with teleost-specific primers. Two fragments of 18S rDNA 306 bp and 756 bp as well as 556 bp fragments of 5S rDNA (120 bp of these sequences are conserved across the species analyzed) were cloned and sequenced (**Figure 5A**, Lanes 1–3).

Labeled probes were produced for both rDNA types and they were hybridized onto the chromosomes. FISH signals for the two rDNAs appeared on separate chromosomes: the signal for 5S rDNA was located on a single chromosome pair (**Figure 5B**), whereas that for 18S rDNA was found in the pericEN region of two pairs of chromosomes (**Figure 5C**).

The 18S rDNA-specific primers were used for screening a BAC library to identify clones containing this ribosomal DNA. Two BAC clones (A6&N6), which partially overlapped forming a contig of ca. 101 kb length were chosen. According to the Southern data, the contig was expected to contain short fragments of 18S

rDNA (Figure 5A, Lanes 4–5). When the two BAC clones forming the contig were labeled and used as probes for hybridization, the signal was chromosome-specific: it showed accumulation in the periCEN regions of two acrocentric chromosomes, and two periTel regions of two small acrocentric chromosomes, while a slightly dispersed signal was observed on all chromosome arms (Figure 5D).

The two BAC clones were sequenced using the HiSeq paired-end protocols and assembled into contigs. Five non-overlapping contigs were produced, their combined length was 120,780 bp



(Supplementary Table S2), whereas the length of the A6N6 contig formed by the two BAC clones selected for sequencing was 101 kb according the physical map (FPC contig 2979; Xia et al., 2010). Then single-end reads produced by the Ion Torrent together with the HiSeq data from the five contigs were used for scaffolding based on: (1) alignment with known fish genomic sequences; and (2) the coverage for each contig by the Ion Torrent reads. The total length of the resulting A6N6 scaffold was 120 kb, which was longer than the combined length of the two BAC inserts. A dot-matrix analysis of the resulting A6N6 scaffold was performed with chromosome sequences corresponding to Linkage group 1 of the European seabass (*Dicentrarchus labrax*; GenBank: FQ310506.3); the results showed 92% of identity between the two sequences (Figure 6). The A6N6 scaffold did not produce any annotated BLASTN hits. In absence of LG-specific markers in these sequences, we were unable to match A6N6 to a unique chromosome(s). The reference sequence from European seabass genome contains incomplete fragments of the rDNA. Comparison to these indicated that our scaffold contained incomplete, short fragments of 18S rDNA (the polyA tail and 40–80 bp). Masking with Repbase sequences showed that only 6.9% of the scaffold yielded hits with annotated repeats. All repeat fragments found were generally short (14–314 bp; Supplementary Table S7). Around 14% of the scaffold produced hits when compared with BLAST against the Asian seabass transcriptome database.

DISCUSSION

Detailed information about repetitive elements present in a particular genome can provide useful information that can be used to improve de novo genome assemblies. For this reason, we have initiated a repeat inventory of the Asian seabass, as part of our genome project. The transcriptome and genome of this teleost species were searched for repetitive elements with experimental and bioinformatic tools. Studies involving comparative genomics have revealed that most vertebrate lineages contain different populations of dispersed elements, including retrotransposons and DNA transposons, and significant differences could be observed in their proportions among species of the same lineage. Highly repetitive satellite sequences and moderately repetitive transposable elements form constitutive heterochromatin, which is mostly

Table 2 | Comparative analysis of the repeat content of teleost and mammalian transcriptomes and genomes showed distinct differences between groups and among species.

Data	Organism	Species					
		Asian seabass	Nile tilapia	Three-spined stickleback	Japanese medaka	Mouse	Human
Genome	Genome size (Mb)	700	927	447	870	2790	3200
	% of repeats	8–14	14	25	17	50	98
Transcriptome	Assembly size (Mb)	169.6*	67.84	45.68	38.25	159.35	297.08
	% of repeats	9.31	7.8	5.7	5.9	10.28	12.74
	Number of transcripts	182,911*	26,788	21,628	24,675	85,276	192,628

*Incomplete assembly, in progress.

Data for analyses—with the exception of those from Asian seabass—were obtained from NCBI (www.ncbi.nih.gov) and Ensembl (www.ensembl.org) databases.

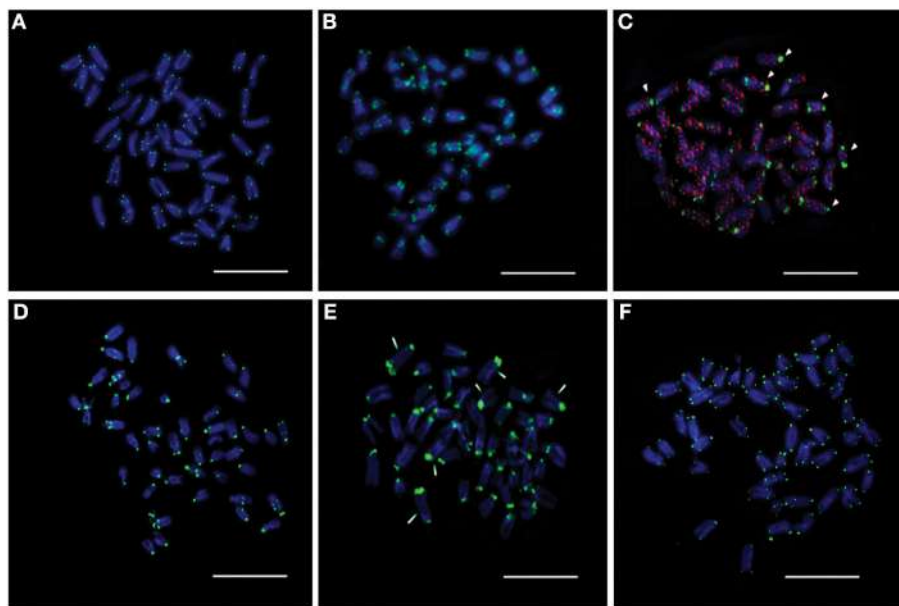


FIGURE 4 | The chromosomal distribution of seven cloned Asian seabass repeats determined by FISH assay. (A) Rex1_SB; (B) MAUI_SB; (C) YRep_SB (green) and GGSat_SB (red); (D) OnSat_SB;

(E) MoSat_SB, and (F) Telomere. Chromosomes were counterstained with DAPI (blue). Arrowheads indicate chromosomes with enriched signals. Bar—5 μ m.

Table 3 | Comparison of the levels (Ct) of repeat-derived transcripts from the gonad and brain of five male and five female adult Asian seabass individuals detected differential expression of MoSat_SB and Rex1_SB between the testis and ovary.

Repeat/Gene name	Ct value		
	Ovary	Testis	Brain
MoSat_SB	20.40 \pm 1	26.42 \pm 1.58*	23.6 \pm 1.18
Rex1_SB	23.92 \pm 1.24	28.80 \pm 1.72*	28.23 \pm 1.4*
YRep_SB	19.60 \pm 0.98	20.72 \pm 1.24	19.37 \pm 0.98
GGSat_SB	19.60 \pm 0.89	22.48 \pm 1.34	18.7 \pm 0.87
Elongation factor 1-alpha	34.00 \pm 1.7	35.40 \pm 2.1	32.00 \pm 1.92
Ribosomal protein L8	17.50 \pm 0.87	16.80 \pm 1	16.00 \pm 0.8

Indicates the Ct values with statistical difference. Statistically significant differences were calculated by the pairwise fixed reallocation randomization test ($P < 0.05$). The Ct values of male and female brain were similar, thus the data were combined. Elongation factor 1-alpha and ribosomal protein L8 were used as reference genes.

located in the preTEL, CEN and periCEN regions of chromosomes (for reviews see Choo, 1997; Kawai et al., 2007; Erukashvili and Ponomartsev, 2013). Our data have demonstrated that transcriptomes show similarity to the non-coding regions of genomes as they also contain various dispersed and tandem elements in different teleost species (Figure 3A; Supplementary Tables S3–S6). The current size of the Asian seabass transcriptome assembly (169.6 Mb) is much larger than those of the other fish species (38–68 Mb; Table 1), presumably due to the following reasons: (i) it is partial and yet to be mapped onto the genome; (ii) while the other teleost transcriptomes were mostly generated

with the aid of Sanger- and/or pyrosequencing of cDNAs, the Asian seabass transcriptome was produced by high throughput next generation sequencing resulting in a much larger amount of sequence data; (iii) Ensembl transcripts were obtained from RNA-seq-based gene models (Flicek et al., 2012), whereas the current version of the Asian seabass transcriptome wasn't; and (iv) there is a possibility of “inflation” due to unique transcripts that have retained introns (i.e., sequencing of immature mRNAs). Although our “in progress” Asian seabass transcriptome is derived from over a dozen different adult tissues and samples collected at several developmental stages, we cannot exclude the possibility that we have missed repeat-derived transcripts that were active, but not expressed in the samples used for sequencing.

Repeat-derived transcript sequences form only ca. 6–13% of the Asian seabass transcriptome, whereas most of the teleost genomes assembled so far contain more than 14% of various repeat types (Table 2). This indicates that the large differences in the repeat content between species are due to the repeat DNAs being present not only in the non-coding regions of genomes (Kramerov and Vassetzky, 2005), but also in the transcriptomes (Table 2, Supplementary Tables S3–S6). Studies involving comparative genomics have revealed that most vertebrate lineages contain different populations of retrotransposable elements and DNA transposons with significant differences frequently observed among species of the same lineage (Ferreira et al., 2011). The repeat-derived transcripts of teleosts are generally short, their average length is 50–400 bp; the largest transcript length is no longer than 3.6 kb for Asian seabass compared to 6 kb for human transcriptome (Supplementary Table S3). The longest repeat-derived transcripts in human and mouse generally belong to

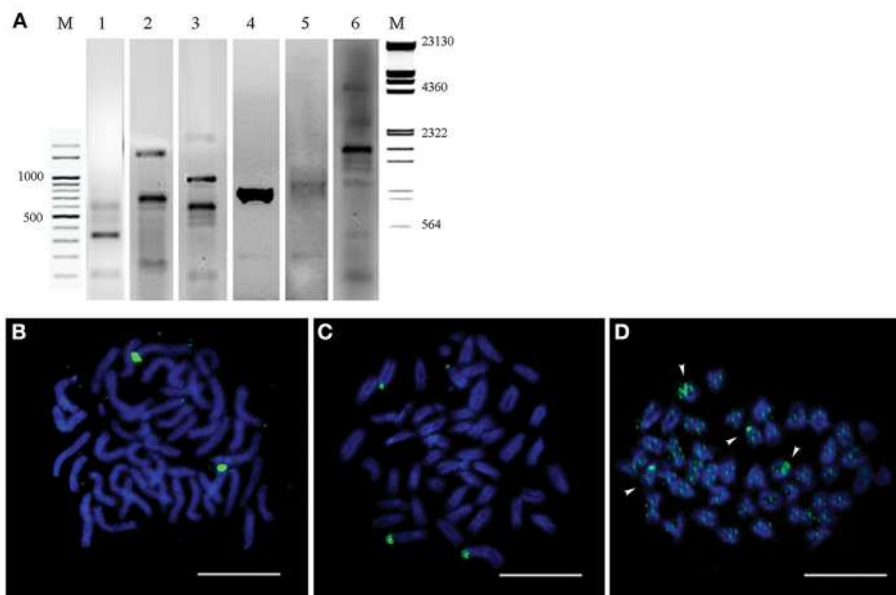


FIGURE 5 | Genomic hybridization and chromosomal location of two ribosomal DNAs (5S and 18s rDNA). (A) 18S rDNA (Lanes 1, 2) and 5S rDNA (Line 3) PCR products of the *Lates calcarifer* genome visualized on 2% agarose gel. Southern blot hybridization of the Asian seabass total DNA digested by restriction endonucleases HinfI and HindIII with the DIG-labeled 18S rDNA (Line 4) and 5S rDNA (Line 5) probes. 18S rDNA PCR product from A6 and N6 BACs was used as a DIG-labeled probe for Southern blot

hybridization of the same two BAC clones digested by HinfI (Line 6). M—molecular mass markers. The corresponding molecular masses are indicated on the left and right of the figure. (B) Fluorescent *in situ* hybridization onto metaphase chromosomes with 5S rDNA probe. (C) Fluorescent *in situ* hybridization onto metaphase chromosomes with 18S rDNA probe. (D) Chromosomal distribution of the A6 and N6 BACs. Arrows indicate chromosomes with enriched signals. Bar—5 μ m.

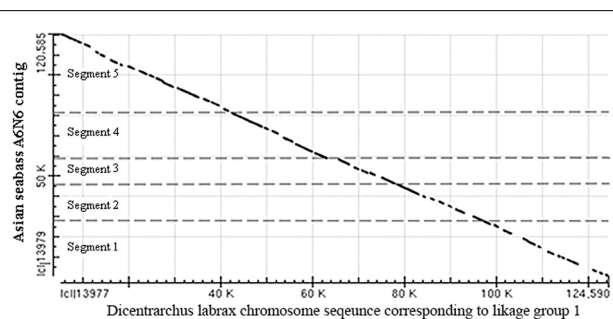


FIGURE 6 | Dot-matrix analysis of the arrangement of five contigs in the A6N6 scaffold showed good alignment with sequences corresponding to LG1 of European seabass. Vertical—18S rDNA scaffold, horizontal—European seabass (*Dicentrarchus labrax*) chromosome sequence corresponding to linkage group 1; (GenBank: FQ310506.3; 13,066,894–13,196,483 bp). For the identity of the five contigs see Supplementary Table S2. These two species are distantly located on the phylogenetic tree of teleosts, but belong to the same order (see **Figure 3B**).

show similarity with each other across different fish genomes (Supplementary Table S3). The maintenance of *D. melanogaster* telomeric DNA is accomplished by repeated retrotransposition of the non-LTR retrotransposon HeT-A/TART specifically to telomeres, in contrast to the maintenance of tandem arrays of species-specific simple DNA repeats at telomeres by telomerase in many eukaryotic organisms (Maxwell, 2004). The result of transcript analysis was consistent with TART producing an array of heterogeneous sense and antisense transcripts (Maxwell, 2004). TART, a non-LTR retrotransposon, was also found in the Asian seabass transcriptome (Supplementary Tables S3, S6). This repeat belongs to the Jockey group (Casacuberta and Pardue, 2003) and its transcript size is longer in teleosts than in mammals (**Figure 7**). The DNA transposon-based transcript content in human and mouse is lower than those of fishes, but that of the endogenous retrovirus group is higher in mammals (**Figure 3**). We demonstrated a fish-specific transcriptome-based repeat profile with a few group of repeats, where the proportion of groups Rex1, SINE, R4, Gypsy, and Jockey is higher in fishes than in human and mouse (**Figure 7**), while that of ERV3 and L1 is lower (**Figure 7**; Supplementary Table S6). The relative amount of primate-specific SINE1/7SL or Alu transcripts is about three magnitudes higher in human than in fishes. The same tendency has been observed in the distribution of R4/Rex6, Rex1, and Gypsy retrotransposons based on comparative analysis of whole genome shotgun sequences of pufferfishes (Japanese fugu and green-spotted pufferfish), zebrafish, human and mouse genomes (Volf et al., 2003; Shirak et al., 2009). The short sequence length

the non-LTR retrotransposon L1, ERV, or Mariner/TC1 groups (Supplementary Table S3), while those in fish transcriptomes are part of a partially overlapping set of repeat types, excluding ERV. The longest repeat transcript for Asian seabass is MAUI_SB with 3575 bp (Supplementary Table S3). Some dispersed, transcript-derived repeats from retrotransposons, such as MAUI (Poulter et al., 1999), Gypsy, Rex1 (Volf et al., 2000, 2003), Bell, and TART (Casacuberta and Pardue, 2003) are longer than 1 kb and

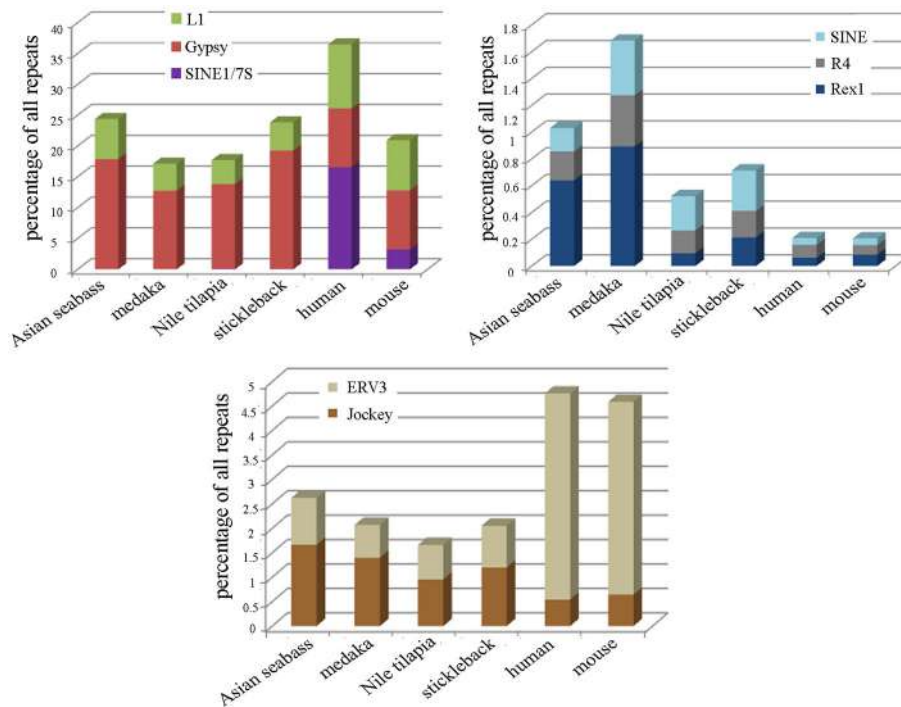


FIGURE 7 | The relative proportion of eight repeats is different among the transcriptomes of teleosts and mammals. The following repeats were analyzed: Jockey, ERV3, Rex1, R4, SINE, SINE/7SL, Gypsy, and L1 in the transcriptomes of Asian seabass,

Japanese medaka, Nile tilapia, three-spined stickleback, human and mouse. Proportions are shown as the percentage of all repeat-containing transcripts per species (Y-axis). The size of transcriptome for the species is indicated in **Table 2**.

of most repeats from different fish genomes and transcriptomes, including satDNA (Vofft et al., 2003; Shirak et al., 2009), indicates that a large part of fish-specific repeat sequences are not characterized.

Moreover, although transcriptomes tested so far tend to have a unique satDNA distribution profile, all investigated fish transcriptomes have similar sets of satellite DNA (Supplementary Table S5). For example, the major satDNA transcript in Asian seabass is OnSat_SB, which has shown similarity with Nile tilapia genome, and is present in the other fish transcriptome datasets, albeit with lower level of similarity. On the other hand, the two main satDNA sequences of the Nile tilapia genome, Satellite A (SATA) and Satellite B (SATB) (Shirak et al., 2009), have not been found in the partial transcriptome data of Asian seabass so far (Supplementary Table S5). The phylogeny based on mitochondrial and nuclear markers suggests independent amplification from the “library.” The existence of a “library of satDNA” has been previously demonstrated experimentally (Mestrovic et al., 1998; Bruvo et al., 2003; Pons et al., 2004) and suggested a common ancestor whose genome harbored all or most of the major satDNA families present in the living species at low copy numbers (Pons et al., 2004). Thus, we can confirm that satDNA is shared by a group of related organisms at variable copy number (Mravinac and Plohl, 2010). The existence of short RNA fragments originating from satDNA has been shown experimentally (Rizzi et al., 2004; Valgardsdottir et al., 2005; Erukashvily et al., 2009; Ting et al., 2011; Kuznetsova et al., 2012). Sequencing

of such transcripts confirmed that they consisted of satellites only. They are polyadenylated (Rizzi et al., 2004; Erukashvily et al., 2009) and exhibit mostly intranuclear spot-like localization (Valgardsdottir et al., 2005; Kuznetsova et al., 2012). The reported length of satDNA transcripts in mammals varied from 20 bp to 5 kb (Valgardsdottir et al., 2005; Ting et al., 2011) probably because the transcripts were revealed at different stages of their processing. In the various cell types and at different stages of development or cell cycle, the transcription of satellites is asymmetrical: either the sense or the antisense strand is transcribed (Rizzi et al., 2004; Valgardsdottir et al., 2005; Erukashvily et al., 2009; Ting et al., 2011; Kuznetsova et al., 2012). Our experiments have demonstrated a different expression level of satDNA (MoSat_SB) and Cr1 non-LTR retrotransposon Rex1_SB, among the ovary and testis—but not between the male and female brain—of adult Asian seabass (**Table 3**).

The distribution pattern of constitutive heterochromatin is a good chromosome marker for some teleosts (Kantek et al., 2009; Vicari et al., 2010). Different strategies have been utilized for isolating repetitive DNA sequences: (1) traditional method genomic DNA restriction (Beridze, 1986); (2) re-association kinetics based on *Cot-1* DNA (Devine et al., 1997; Langmead et al., 2009); (3) the microdissection of chromosomes submitted to C-banding and subsequent amplification of heterochromatic sequences using DOP-PCR or WGS amplification primers (Cioffi et al., 2013); and (4) bioinformatic analysis of WGS and NGS sequences (Komissarov et al., 2011) and transcriptome data

(Jiang et al., 2012). Some of these approaches have been used to investigate the repeats in this study.

Six new repeats belonging to tandem and dispersed groups of repetitive DNA have been identified, they constitute ~8–14% of the Asian seabass genome. The MoSat_SB and YRep_SB repeats are enriched on individual chromosomes, while the other four—Rex1_SB, MAUI_SB, OnSat_SB, and GGSat_SB—are generally distributed on the autosomes throughout the karyotype (Figures 4, 8). The analysis of the chromosomal location of repeated elements demonstrated that they are, in most cases, compartmentalized in heterochromatic regions and the periCEN regions of chromosomes tend to show a unique repeat distribution as has been observed in several other vertebrates (Beridze, 1986; Erukashvily and Ponomartsev, 2013). Our results have demonstrated that the compartmentalization of repeat elements is mostly restricted to heterochromatic AT-rich segments, whereas most of the chromosomes have large GC-rich blocks within their centromeric regions (Figures 1, 8). The AT-rich B-chromosomes of the Asian seabass contain chromatin which differs from heterochromatin identified in periCEN regions of the autosomal chromosomes (Figure 8). The GC-rich B-chromosomes, as chromosomes with potentially higher recombination rate could be useful for future characterization *Lates* populations throughout the Indian, South-East Asian and Australian regions.

Transposable elements (TEs) can be organized in clusters or dispersed throughout the genome. CR1 elements regulate gene activity and located within genes, as MAUI from the *fugu* genome (Poulter et al., 1999). Among the CR1 clade of LINE element family of non-LTR retrotransposons, Rex was characterized for the first time in the genome of the *Xiphophorus* and it was found to be widely present in different fish genomes (Volff et al., 2000). Phylogenetic analysis revealed that Rex1 retrotransposons were frequently active during fish evolution. They formed multiple ancient lineages, which underwent several independent and recent bursts of retrotransposition and invaded fish genomes with

variable success rate (Volff et al., 2000; Ferreira et al., 2011). The divergence between two cloned sequences of Rex1 in the Asian seabass was ~10%, at the same time about 20% of divergence was found among Rex1 elements isolated from three different teleost species (Volff et al., 2000, 2003). The physical mapping of different Rex elements showed that they were primarily compartmentalized in the periCEN heterochromatic regions, although dispersed or clustered signals in euchromatic regions were also observed (Figure 4A, Valente et al., 2011). The presence of TEs in heterochromatin can be correlated with their role in structure and organization of heterochromatic areas (such as centromeres) or with the lower selective pressure that act on these gene-poor regions. Rex elements were also concentrated in the largest chromosome pair of the Nile tilapia, *Oreochromis niloticus*. This chromosome pair is supposed to have originated by fusions, demonstrating the possible involvement of TEs with chromosome rearrangements (Valente et al., 2011). Rex1, 3, and 6 are non-LTR elements that have been active during the evolution of different fish lineages. In the family of *Cichlidae* (*Perciformes*), Rex elements are organized in clusters within the genome of the majority of the species (Martins et al., 2004). Rex elements showed a wide distribution among fishes and could be observed both in the periCEN region and euchromatic regions of chromosomes. However, at present, due to the lack of an assembled genome, we are unable to resolve their exact location in relation to the periCEN region. Rex elements can also be associated with genes involved with sex determination (Volff et al., 2000). In this study, we have demonstrated the existence of Rex1-derived transcripts in Asian seabass; moreover, their expression levels were significantly lower in the adult ovary than in the testis (Table 3).

The Nucleolus Organizer Regions (NORs) are highly repetitive genome sites related to rRNA synthesis (Preuss and Pikaard, 2007; Pinhal et al., 2008). NORs present small, active transcription sites (evolutionally conserved 120 bp for 5S rDNA and 300–2800 bp for 18S rDNA) and highly variable non-transcribed spacing segments with their own structural dynamics, in which the presence of transposons located close to the genome regions has been identified (Martins et al., 2004). rDNA sequences and their chromosomal location have proven to be valuable as genetic markers to distinguish closely related species and also in the understanding of the dynamic of repetitive sequences in the genomes. In this study we have shown that the chromosomal position of 5S rDNA and 18S rDNA probes in Asian seabass is similar to those observed in different fish species (Martins et al., 2004; Mantovani et al., 2005; Cioffi et al., 2010; Merlo et al., 2010; Nakajima et al., 2012). At the cytogenetic level, the 5S rDNA probes hybridized mostly near to the centromere region; while the 18S rRNA gene probe has identified variable positions across different fish species (Martins et al., 2004; Mantovani et al., 2005; Cioffi et al., 2010; Merlo et al., 2010; Nakajima et al., 2012).

In order to improve our understanding of their organization, rDNA sequences were analyzed in the Asian seabass genome. Two clones were identified through hybridization with 18SrDNA-derived probes from a BAC library; they formed a single contig (A6N6). They were sequenced with two NGS platforms (Illumina HiSeq and Ion Torrent) and assembled. The A6N6 contig is extremely rich in repeats, containing conserved (Supplementary

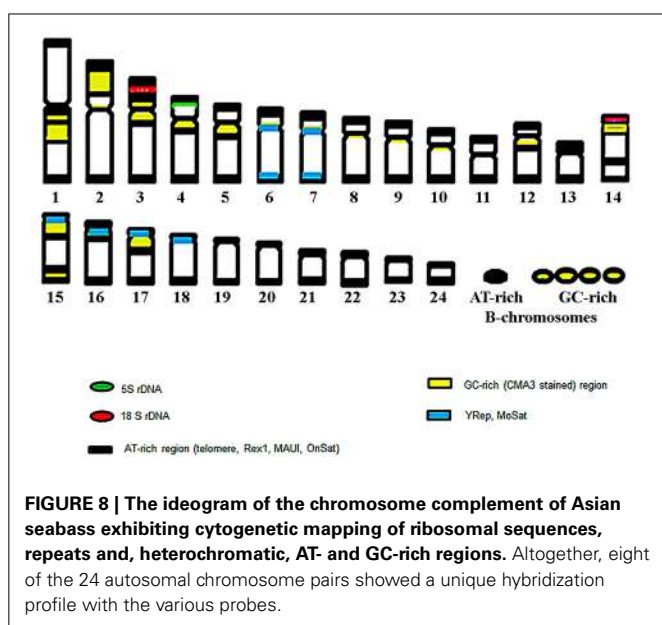


Table S7) and novel sequences. Some of them are chromosome-specific, whereas others are dispersed throughout the chromosomes (Figure 5D). Due to the shortness of reads, the assembled repeats tend to be highly fragmented. In teleosts, the spacer portion of the major rDNA is generally GC-rich (Schmid and Guttenbach, 1988), whereas our A6N6 scaffold shows a high (59%) AT content. It was shown in the red wolf fish (*Erythrinus erythrinus*) that rDNA sequences were co-localized with the Rex3 retrotransposable element in the centromeric heterochromatin (Valente et al., 2011) and they have also shown strong association with Tc1/Mariner and Rex elements (Schmid and Guttenbach, 1988; Cioffi et al., 2010; Merlo et al., 2010; Nakajima et al., 2012). The presence of Tc1/Mariner and Rex1_SB repeats was observed in the A6N6 scaffold (Supplementary Table S7). Transcripts derived from the A6N6 scaffold showed sequence homology with a predicted gene (PDZ domain containing ring finger 3) from fish and amphibian genomes. The synteny of Asian seabass A6N6 scaffold with a European seabass linkage group (Figure 6) and their phylogenetic association (Figure 3B) strengthened the idea that the chromosomes have undergone rearrangements during evolution. These rearrangements were likely mediated by retrotransposon activity in which the insertion of the retrotransposable element into rDNA sequences created an rDNA-transposon complex that moved and dispersed in the karyotype (Cioffi et al., 2010; Nakajima et al., 2012).

The transcriptome and genome of Asian seabass were searched for repetitive elements with experimental and bioinformatics tools. In summary, our data indicate that the sequence and structure of most Asian seabass repeat DNAs are likely to be unknown. Detailed analysis of the completed genome assembly will provide the final proof for this suggestion. The approach used for the analysis of repeats in this study has also yielded useful knowledge only about evolutionarily conserved repeats. All analyzed sequences, except Tel, belong to the periCEN part of chromosomes (Figure 8), they could form chromosome-specific periCEN patterns of compartmentalization and the transcripts of some show differential expression between the adult gonads of Asian seabass sex.

ACKNOWLEDGMENTS

The authors thank Jun Hong Xia and Gen Hua Yue for access to their BAC library. This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Competitive Research Programme (Award No: NRF-CRP7-2010-001) and the Strategic Research Program of Temasek Life Science Laboratory for László Orbán, as well as the Russian Ministry of Science (Mega-grant no.11.G34.31.0068) and a grant from presidium RAS (MCB, N01200955639).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00223/abstract>

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Beridze, T. (1986). *Satellite DNA*. Berlin: Springer-Verlag.
- Britten, R. J., and Kohne, D. E. (1968). Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161, 529–540. doi: 10.1126/science.161.3841.529
- Bruvo, B., Pons, J., Ugarković, D., Juan, C., Petitpierre, E., and Plohl, M. (2003). Evolution of low-copy number and major satellite DNA sequences coexisting in two *Pimelia* species-groups (*Coleoptera*). *Gene* 312, 85–94. doi: 10.1016/S0378-1119(03)00603-6
- Cardinali, G., Liti, G., and Martini, A. (2000). Non-radioactive dot-blot DNA reassociation for unequivocal yeast identification. *Int. J. Syst. Evol. Microbiol.* 50, 931–936. doi: 10.1099/00207713-50-2-931
- Carone, D. M., Longo, M. S., Ferreri, G. C., Hall, L., Harris, M., Shook, N., et al. (2009). A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma* 118, 113–125. doi: 10.1007/s00412-008-0181-5
- Casacuberta, E., and Pardue, M. L. (2003). Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the virilis group. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3363–3368. doi: 10.1073/pnas.0230353100
- Chew, J. S. K., Oliveira, C., Wright, J. M., and Dobson, M. J. (2002). Molecular and cytogenetic analysis of the telomeric (TTAGGG)_n repetitive sequences in the Nile tilapia, *Oreochromis niloticus* (Teleostei: Cichlidae). *Chromosoma* 111, 45–52. doi: 10.1007/s00412-002-0187-3
- Choo, K. H. (1997). *The Centromere*. New York, NY: Oxford University Press.
- Cioffi, M. B., Liehr, T., Trifonov, V., Molina, W. F., and Bertollo, L. A. (2013). Independent sex chromosome evolution in lower vertebrates: cytogenetic overview in the erythrinidae fish family. *Cytogenet. Genome Res.* 6, 2–9. doi: 10.1159/000354039
- Cioffi, M. B., Martins, C., and Bertollo, L. A. (2010). Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. *BMC Evol. Biol.* 10:271. doi: 10.1186/1471-2148-10-271
- Cioffi, M. B., Molina, W. F., Moreira-Filho, O., and Bertollo, L. A. (2011). Chromosomal distribution of repetitive DNA sequences highlights the independent differentiation of multiple sex chromosomes in two closely related fish species. *Cytogenet. Genome Res.* 134, 295–302. doi: 10.1159/000329481
- Devine, S. E., Chissoe, S. L., Eby, Y., Wilson, R. K., and Boeke, J. D. (1997). A transposon-based strategy for sequencing repetitive DNA in eukaryotic genomes. *Genome Res.* 7, 551–563.
- Enukashvily, N. I., Malashicheva, A. B., and Waisertrreiger, I. S. R. (2009). Satellite DNA spatial localization and transcriptional activity in mouse embryonic E-14 and IOUD2 stem cells. *Cytogenet. Genome Res.* 124, 277–287. doi: 10.1159/000218132
- Enukashvily, N. I., and Ponomartsev, N. V. (2013). Mammalian satellite DNA: a speaking dumb. *Adv. Protein Chem. Struct. Biol.* 90, 31–65. doi: 10.1016/B978-0-12-410523-2.00002-X
- Ferreira, D. C., Oliveira, C., and Foresti, F. (2011). Chromosome mapping of retrotransposable elements Rex1 and Rex3 in three fish species in the subfamily Hypoptopomatinae (Teleostei, Siluriformes, Loricariidae). *Cytogenet. Genome Res.* 132, 64–70. doi: 10.1159/000319620
- Ferreira, I. A., and Martins, C. (2008). Physical chromosome mapping of repetitive DNA sequences in Nile tilapia *Oreochromis niloticus*: evidences for a differential distribution of repetitive elements in the sex chromosomes. *Micron* 39, 411–418. doi: 10.1016/j.micron.2007.02.010
- Flicek, P., A. M., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., et al. (2012). Ensembl 2012. *Nucleic Acids Res.* 40, 84–90. doi: 10.1093/nar/gkr991
- Gjedrem, T. (2010). The first family-based breeding program in aquaculture. *Rev. Aqua* 2, 2–15. doi: 10.1111/j.1753-5131.2010.01011.x
- Gjedrem, T., and Baranski, M. (2009). *Selective Breeding in Aquaculture: An Introduction*. New York, NY: Springer.
- Guiguen, Y., Cauty, C., Fostier, A., Fuchs, J., and Jalabert, B. (1993). Reproductive cycle and sex inversion of the seabass, *Lates calcarifer*, reared in sea cages in

- French Polynesia: histological and morphometric description. *Environ. Biol. Fish* 39, 231–247. doi: 10.1007/BF00005126
- Han, Y. H., Zhang, Z. H., Liu, J. H., Lu, J. Y., Huang, S. W., and Jin, W. W. (2008). Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence *in situ* hybridization. *Cytogenet. Genome Res.* 122, 80–88. doi: 10.1159/000151320
- Jiang, F., Yang, M., Guo, W., Wang, X., and Kang, L. (2012). Large-scale transcriptome analysis of retroelements in the migratory locust, *Locusta migratoria*. *PLoS ONE* 7:e40532. doi: 10.1371/journal.pone.0040532
- Kantek, D. L. Z., Vicari, M. R., Peres, W. A. M., Cestari, M. M., Artoni, R. F., Bertollo, L. A. et al. (2009). Chromosomal location and distribution of As51 satellite DNA in five species of the genus *Astyanax* (Teleostei, Characidae, Incertae sedis). *J. Fish Biol.* 75, 408–421. doi: 10.1111/j.1095-8649.2009.02333.x
- Kawai, A., Nishida-Umehara, C., Ishijima, J., Tsuda, Y., Ota, H., and Matsuda, Y. (2007). Different origins of bird and reptile sex chromosomes inferred from comparative mapping of chicken Z-linked genes. *Cytogenet. Genome Res.* 117, 92–102. doi: 10.1159/000103169
- Kohany, O., Gentles, A. J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7:474. doi: 10.1186/1471-2105-7-474
- Komissarov, A. S., Gavrilova, E. V., Demin, S. J., Ishov, A. M., and Podgornaya, O. I. (2011). Tandemly repeated DNA families in the mouse genome. *BMC Genomics* 12:531. doi: 10.1186/1471-2164-12-531
- Kramerov, D. A., and Vassetzky, N. S. (2005). Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* 247, 165–221. doi: 10.1016/S0074-7696(05)47004-7
- Kuznetsova, T., Erukashvily, N., Trofimova, I., Gorbunova, A., Vashukova, E., and Baranov, V. (2012). Localisation and transcription of human chromosome 1 pericentromeric heterochromatin in embryonic and extraembryonic tissues. *Med. Genetics* 11, 19–24.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Lee, B.-Y., Lee, W.-J., Streelman, J. T., Carleton, K. L., Howe, A. E., Hulata, G., et al. (2005). A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* 170, 237–244. doi: 10.1534/genetics.104.035022
- Letunic, I., and Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39, W475–W478. doi: 10.1093/nar/gkr201
- Levan, A., Fredga, K., and Sandberg, A. A. (1964). Nomenclature for centromeric position on chromosomes. *Hereditas* 52, 201–220. doi: 10.1111/j.1601-5223.1964.tb01953.x
- Liu, Z., and Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37. doi: 10.1016/j.aquaculture.2004.05.027
- Luna, S. (2008). *Lates calcarifer, Barramundi: Fisheries, Aquaculture, Gamefish, Aquarium* [Online]. Fish Base (On-line) Available online at: <http://fishbase.sinica.edu.tw/summary/speciessummary.php?genusname=Lates&speciesname=calcarifer> [Accessed April 2, 2008].
- Mantovani, M., Abel, L. D. D. S., and Moreira-Filho, O. (2005). Conserved 5S and variable 45S rDNA chromosomal localisation revealed by FISH in *Astyanax scabripinnis* (Pisces, Characidae). *Genetica* 123, 211–216. doi: 10.1007/s10709-004-2281-3
- Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011
- Martins, C., Wasko, A. P., Chapter, X., Clyde, R., and Nova, W. (2004). “Organization and evolution of 5S ribosomal DNA in the fish genome,” in *Focus on Genome Research*, ed C. R. Williams (New York, NY: Nova Science Publishers), 335–363.
- Maxwell, P. (2004). *Molecular Genetic Analysis of TART (Telomere-Associated Retrotransposon) in Drosophila Melanogaster*. Ph.D. thesis, Syracuse University Biology.
- Merlo, M. A., Cross, I., Chairi, H., Manchado, M., and Rebordinos, L. (2010). Analysis of three multigene families as useful tools in species characterization of two closely-related species, *Dicentrarchus labrax*, *Dicentrarchus punctatus* and their hybrids. *Genes Genet. Syst.* 85, 341–349. doi: 10.1266/ggs.85.341
- Mestrović, N., Plohl, M., Mravinac, B., and Ugarković, D. (1998). Evolution of satellite DNAs from the genus *Palorus*—experimental evidence for the “library” hypothesis. *Mol. Biol. Evol.* 15, 1062–1068. doi: 10.1093/oxfordjournals.molbev.a026005
- Moore, R. (1979). Natural sex inversion in the giant perch (*Lates calcarifer*). *Aust. J. Mar. Freshwat. Res.* 30, 803–813. doi: 10.1071/MF9790803
- Mravinac, B., and Plohl, M. (2010). Parallelism in evolution of highly repetitive DNAs in sibling species. *Mol. Biol. Evol.* 27, 1857–1867. doi: 10.1093/molbev/msq068
- Nakajima, R. T., Cabral-De-Mello, D. C., Valente, G. T., Venere, P. C., and Martins, C. (2012). Evolutionary dynamics of rRNA gene clusters in cichlid fish. *BMC Evol. Biol.* 12:198. doi: 10.1186/1471-2148-12-198
- Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30:e36. doi: 10.1093/nar/30.9.e36
- Pinhal, D., Gadigll, O., Wasko, A. P., Oliveirai, C., Ron, E., Foresti, F., et al. (2008). Discrimination of *Shark species* by simple PCR of 5S rDNA repeats. *Genet. Mol. Biol.* 361–365. doi: 10.1590/S1415-4752008000200033
- Pons, J., Bruvo, B., Petitpierre, E., Plohl, M., Ugarkovic, D., and Juan, C. (2004). Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common ‘satellite DNA library’. *Heredity* 92, 418–427. doi: 10.1038/sj.hdy.6800436
- Poulter, R., Butler, M., and Ormandy, J. (1999). A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. *Gene* 227, 169–179. doi: 10.1016/S0378-1119(98)00600-3
- Pradeep, P. J., Srijaya, T. C., Zain, R. B., Papini, A., and Chatterji, A. K. (2011). A Simple technique for chromosome preparation from embryonic tissues of teleosts for ploidy verification cytology. *Caryologia* 64, 235–241. doi: 10.1080/00087114.2002.10589788
- Preuss, S., and Pikaard, C. S. (2007). rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. *Biochim. Biophys. Acta Gene Struct. Expr.* 1769, 383–392. doi: 10.1016/j.bbexp.2007.02.005
- Rizzi, N., Denegri, M., Chiodi, I., Corioni, M., Valgardsdottir, R., Cobiانchi, F., et al. (2004). Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. *Mol. Biol. Cell* 15, 543–551. doi: 10.1091/mbc.E03-07-0487
- Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: Manual, 3rd Edn.* Cold Spring, NY: Cold Spring Harbor Laboratory Press.
- Schmid, M. (1980). Chromosome banding in amphibia. IV. Differentiation of GC- and AT-rich chromosome regions in Anura. *Chromosoma* 77, 83–103. doi: 10.1007/BF00292043
- Schmid, M., and Guttenbach, M. (1988). Evolutionary diversity of reverse (R) fluorescent chromosome bands in vertebrates. *Chromosoma* 97, 101–114. doi: 10.1007/BF00327367
- Schweizer, D. (1980). Simultaneous fluorescent staining of R bands and specific heterochromatic regions (DA-DAPI bands) in human chromosomes. *Cytogenet. Cell Genet.* 27, 190–193. doi: 10.1159/000131482
- Shirak, A., Grabherr, M., Di Palma, F., Lindblad-Toh, K., Hulata, G., Ron, M., et al. (2009). Identification of repetitive elements in the genome of *Oreochromis niloticus*: tilapia repeat masker. *Mar. Biotechnol.* 12, 121–125. doi: 10.1007/s10126-009-9236-8
- Sutton, G. G., White, O., Adams, M. D., and Kerlavage, A. R. (1995). Tigr assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci Tech* 1, 9–19. doi: 10.1089/gst.1995.1.9
- Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Akhavanfar, S., Coffman, E. J., et al. (2011). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331, 593–596. doi: 10.1126/science.1200801
- Valente, G. T., Mazzuchelli, J., Ferreira, I. A., Poletto, A. B., Fantinatti, B. E., and Martins, C. (2011). Cytogenetic mapping of the retroelements Rex1, Rex3 and Rex6 among cichlid fish: new insights on the chromosomal distribution of transposable elements. *Cytogenet. Genome Res.* 133, 34–42. doi: 10.1159/000322888
- Valgardsdottir, R., Chiodi, I., Giordano, M., Cobiانchi, F., Riva, S., and Biamonti, G. (2005). Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. *Mol. Biol. Cell* 16, 2597–2604. doi: 10.1091/mbc.E04-12-1078

- Vicari, M. R., Nogaroto, V., Noletto, R. B., Cestari, M. M., Cioffi, M. B., Almeida, M. C., et al. (2010). Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. *J. Fish Biol.* 76, 1094–1116. doi: 10.1111/j.1095-8649.2010.02564
- Volf, J.-N., Bouneau, L., Ozouf-Costaz, C., and Fischer, C. (2003). Diversity of retrotransposable elements in compact pufferfish genomes. *Trends Genet.* 19, 674–678. doi: 10.1016/j.tig.2003.10.006
- Volf, J. N., Körting, C., and Scharl, M. (2000). Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. *Mol. Biol. Evol.* 17, 1673–1684. doi: 10.1093/oxfordjournals.molbev.a026266
- Willard, H. F. (1991). Evolution of alpha satellite. *Curr. Opin. Genetics Dev.* 1, 509–514. doi: 10.1016/S0959-437X(05)80200-X
- Xia, J. H., Feng, F., Lin, G., Wang, C. M., and Yue, G. H. (2010). A First Generation BAC-Based Physical Map of the Asian Seabass (*Lates calcarifer*). *PLoS ONE* 5:e11974. doi: 10.1371/journal.pone.0011974
- Zwick, M. G., Wiggs, M., and Paule, M. R. (1991). Sequence and organization of 5S RNA genes from the eukaryotic protist *Acanthamoeba castellanii*. *Gene* 101, 153–157. doi: 10.1016/0378-1119(91)90239-8
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 January 2014; accepted: 27 June 2014; published online: 25 July 2014.

Citation: Kuznetsova IS, Thevasagayam NM, Sridatta PSR, Komissarov AS, Saju JM, Ngho SY, Jiang J, Shen X and Orbán L (2014) Primary analysis of repeat elements of the Asian seabass (*Lates calcarifer*) transcriptome and genome. *Front. Genet.* 5:223. doi: 10.3389/fgene.2014.00223

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Kuznetsova, Thevasagayam, Sridatta, Komissarov, Saju, Ngho, Jiang, Shen and Orbán. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*)

Jørgen Ødegård^{1*}, Thomas Moen², Nina Santi², Sven A. Korsvoll¹, Sissel Kjølglum¹ and Theo H. E. Meuwissen³

¹ Breeding and Genetics, AquaGen AS, Trondheim, Norway

² Research and Development, AquaGen AS, Trondheim, Norway

³ Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Aas, Norway

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Jiuzhou Song, University of Maryland, USA

Roger Vallejo, United States Department of Agriculture, USA
Ignacy Misztal, University of Georgia, USA

*Correspondence:

Jørgen Ødegård, Department of Animal and Aquacultural Sciences, Norwegian University of Life Science, Arboretveien 6, PO Box 5003, Aas, NO-1432, Norway
e-mail: jorgen.odegard@aquagen.no

Reliability of genomic selection (GS) models was tested in an admixed population of Atlantic salmon, originating from crossing of several wild subpopulations. The models included ordinary genomic BLUP models (GBLUP), using genome-wide SNP markers of varying densities (1–220 k), a genomic identity-by-descent model (IBD-GS), using linkage analysis of sparse genome-wide markers, as well as a classical pedigree-based model. Reliabilities of the models were compared through 5-fold cross-validation. The traits studied were salmon lice (*Lepeophtheirus salmonis*) resistance (LR), measured as (log) density on the skin and fillet color (FC), with respective estimated heritabilities of 0.14 and 0.43. All genomic models outperformed the classical pedigree-based model, for both traits and at all marker densities. However, the relative improvement differed considerably between traits, models and marker densities. For the highly heritable FC, the IBD-GS had similar reliability as GBLUP at high marker densities (>22 k). In contrast, for the lowly heritable LR, IBD-GS was clearly inferior to GBLUP, irrespective of marker density. Hence, GBLUP was robust to marker density for the lowly heritable LR, but sensitive to marker density for the highly heritable FC. We hypothesize that this phenomenon may be explained by historical admixture of different founder populations, expected to reduce short-range linkage disequilibrium (LD) and induce long-range LD. The relative importance of LD/relationship information is expected to decrease/increase with increasing heritability of the trait. Still, using the ordinary GBLUP, the typical long-range LD of an admixed population may be effectively captured by sparse markers, while efficient utilization of relationship information may require denser markers (e.g., 22 k or more).

Keywords: Atlantic salmon, genomic selection, reliability, admixture, genetics

INTRODUCTION

Aquaculture populations are characterized by high male and female fecundity, typically resulting in large full-sib families. For invasive traits, traditional aquaculture selection programs involve sib-testing, which has limited reliability under classical selection schemes, as selection candidates are evaluated based on mid-parent means. Furthermore, this also leads to increased co-selection among close relatives, and enforcing restrictions on inbreeding will therefore hamper selection on such traits more than selection for individually evaluated traits. For individually evaluated traits, the sizeable family groups of aquaculture species give a substantial potential for within-family selection.

Marker-assisted selection can be used to select directly for favorable QTL alleles, a method that allows individual selection of genotyped animals even in absence of phenotyping. This requires that the QTL effects are known and that carriers of the favorable alleles can be identified through the markers. For Atlantic salmon, the method has been utilized with great success in selection for reduced incidence of infectious pancreatic necrosis (IPN), for which a single QTL explains (nearly) all genetic variation (Moen

et al., 2009). In situations where several QTL underlie the trait, MAS will be more complex, and power of QTL detection lower as each QTL explains a smaller fraction of the total genetic variance. For such traits genomic selection (GS) is a viable alternative, utilizing information from numerous genome-wide marker loci jointly in the genetic analysis (Meuwissen et al., 2001). The GS methods facilitates computation of individual breeding values for all genotyped animals and do not require any prior knowledge of the underlying QTL. In simulated aquaculture populations, superior performance of GS models compared with classical models has been documented in several publications (e.g., Nielsen et al., 2009; Ødegård et al., 2009; Ødegård and Meuwissen, 2014), while documentation from real aquaculture data has been largely absent so far.

The original idea behind GS was that the genome-wide markers would capture linkage disequilibrium between marker loci and QTL (Meuwissen et al., 2001). However, accuracy of GS has been shown to be non-zero even in absence of linkage disequilibrium (LD) (Habier et al., 2007), and the actual reliability of GS models can thus be explained by three types

of quantitative-genetic information sources contained in the genomic data (Habier et al., 2013):

- (1) Pedigree;
- (2) Co-segregation (linkage analysis information);
- (3) Population-wide linkage disequilibrium (LD).

The ancestry (pedigree) is indeed reflected through inheritance of marker loci and is thus implicitly included in the dense marker information, although pedigree is not used directly. Co-segregation is the deviation from independent segregation of alleles as a result of linkage (i.e., deviations between relationships estimated from pedigree and linkage analysis), while LD is the statistical dependency between alleles at different loci in the base generation (i.e., the generation with unknown parents). Information on (1) and (2) can thus explain the non-zero reliability of GS even in absence of LD. Furthermore, in populations of strong relationship structure (e.g., livestock and aquaculture populations) LD may not even be the most important of these factors under GS; Wientjes et al. (2013) showed that the level of family relationship between selection candidates and the reference population had a higher effect on reliability of GS than LD *per se*.

There are currently numerous available GS methodologies. The most widely used methods are GS models using identity-by-state (IBS) information on dense genome-wide SNP markers, including the so-called genomic BLUP (GBLUP) and Bayesian methods (e.g., BayesA, BayesB, BayesC, BayesD) (Meuwissen et al., 2001; Habier et al., 2011). Other methods involve use of SNP haplotypes (combining multiple SNPs), that also take identity-by-descent (IBD) information into account (Calus et al., 2008). Finally, GS may be performed based on linkage analysis of genome-wide markers, producing an IBD genomic relationship matrix (IBD-GS), completely ignoring LD information (Villanueva et al., 2005; Luan et al., 2012).

In the following, we will focus on two of these methodologies for use in aquaculture breeding: Ordinary GBLUP and IBD-GS. GBLUP can be implemented by ridge-regression on genome-wide marker genotypes (Meuwissen et al., 2001) or by an animal model using a realized genomic relationship matrix estimated from marker genotype similarities across the genome (Hayes et al., 2009). The latter method will be used here. The advantage of the IBD-GS model lies in its ability to utilize realized IBD relationships rather than expected relationships estimated through the pedigree, e.g., full-sibs (which are numerous in aquaculture) are no longer necessarily related by a coefficient of $\frac{1}{2}$, but their relationships depend on the actual length of shared IBD chromosome segments, which are traced by the markers through linkage analysis. Compared with other GS methods, IBD-GS has the advantage that it can be successfully implemented even at extremely low marker densities. This is due to the fact that number of recombinations from parent to offspring is usually low (i.e., averaging one per Morgan), and inheritance of long chromosomal blocks can thus be traced accurately even with a few genome-wide markers. A recent simulation study on an aquaculture-like population indicated that IBD-GS works effectively at densities where IBS-based methods are expected to

fail, e.g., with 10–20 SNPs/Morgan (Vela-Avitúa et al., in press). Thus, there is no need for dense marker panels, making IBD-GS attractive for cost-effective GS implementation. For dairy cattle, IBD-GS models have been shown to give similar reliability as ordinary GBLUP models with dense markers (Luan et al., 2012). Hence, for livestock populations with large family sizes, realized close relationships (pedigree and co-segregation) are essential for the reliability of any GS model, and GS methodology may thus have large potential even in absence of strong LD structures. Aquaculture populations typically have strong relationship structures, with selection candidates having numerous full-sibs and potentially both maternal and paternal half-sib groups.

The Norwegian AquaGen Atlantic salmon population originates from the first family-based selective breeding program on Atlantic salmon, going back to the 1970's, based on crossing of wild founders from numerous wild Norwegian river strains (Gjedrem et al., 1991). Originally, four parallel populations were created, one for each year class in a 4 year generation interval. Although as much as 41 river strains were originally included, contributions of the different rivers vary considerably, both between the original base populations of the 4 year classes and as result of subsequent selection. Hence, the original farmed populations were indeed heavily admixed. The year-class strains were selected for a common breeding goal, but kept largely separate for 7 generations until 2005, when they were merged into a single population. Hence, the AquaGen population can be regarded as an admixed/synthetic population comprised of genetic material from many wild subpopulations, which likely have been separated for a long time in nature.

Admixture between genetically distinct populations increases LD between all loci (linked and unlinked) that have different allele frequencies in the founding populations (Pfaff et al., 2001). However, LD between unlinked loci will quickly be removed through recombinations, while LD between linked loci will be more persistent, e.g., for loci separated by 1 or 10 cM, respectively 90 and 35% of the admixture-induced LD (ALD) is expected to remain even after 10 generations, while 82 and 12% remain after 20 generations. However, admixture will not only introduce long-range ALD, it will also reduce the short range LD, i.e., the LD existing in the original founder populations. The short-range LD will decrease as phase associations between marker and QTL alleles can differ depending of the origin of the chromosome segments (Thomasen et al., 2013), and haplotype segments with strong LD are thus shorter in admixed populations (Toosi et al., 2010). This can be illustrated by the following example, assuming two sub-populations for simplicity: The frequency of a M_1N_1 haplotype is $(p + \kappa)(q + \lambda) + D_I$ in population I, where $(p + \kappa)$ and $(q + \lambda)$ are the frequencies of the alleles M_1 and N_1 , expressed as deviations from the across population frequencies (p and q), with frequency deviations κ and λ , and D_I is the LD in population I. Similarly, $(p - \kappa)(q - \lambda) + D_{II}$ is the haplotype frequency in population II. The haplotype frequency in their crossbred-offspring (F_1) is thus: $(p + q + \bar{D})$, where \bar{D} is the average of D_I and D_{II} . The LD in the F_1 cross is $(\kappa\lambda + \bar{D})$, which comprises a ALD term $\kappa\lambda$ due to the crossbreeding (depends on frequency differences and is independent of distances between the loci), and the average of the original population-specific LD

coefficients between the loci. \bar{D} is on average smaller than either D_I or D_{II} since they may have opposite signs in the two populations, resulting in a reduced short-range LD in the admixed population.

The reduced short-range LD originating from founder populations may challenge accurate genomic prediction. Still, long-range ALD (the $\kappa\lambda$ term) can be effectively captured even by sparse markers, but may explain a limited fraction of the genetic variance, depending on the degree of differentiation between the founding populations.

Hence, effectiveness of GS in admixed populations depends on several layers of information: remaining LD from the founder populations, long-range ALD, and the relationship structure within the existing population. Furthermore, the relative importance of these factors likely depends on genetic architecture, marker density, heritability and the GS methodology used.

The aim of the study was to quantify the importance of marker density on the reliability of ordinary GBLUP models and to compare these estimates with IBD-based models completely ignoring LD, i.e., classical pedigree-based models and IBD-GS models. To this end, two traits measured on Atlantic salmon [fillet color (FC) and salmon lice resistance], with high and low heritability, using alternative GS models and marker densities were studied. So far, no QTL of large effect has yet been found for lice resistance, but major QTL have been found for FC, still these do not explain all genetic variance (Baranski et al., 2010).

MATERIALS AND METHODS

DATA

The fish used in the material were from the AquaGen population year-class first-fed in 2011. In total, 157 full-sib families (offspring of 99 dams and 97 sires) were sampled for salmon lice (*Lepeophtheirus salmonis*) challenge testing, and 30–40 fish from each of these families were transferred to Nofima at Averøy, Norway and put into sea net-cages in October 2011. Two separate lice tests were conducted the following year, with all families being represented in both tests. Test 1 was conducted in the period July 16–18, 2012 and Test 2 in the period October 17–19, 2012. The total number of challenge-tested fish was 5198, with 2850 and 2348 fish in Test 1 and 2, respectively. Lice challenge testing of the fish was approved by the Norwegian Animal Research Authority (S-2012/148773).

Challenge testing was conducted by closing the net cages with tarpaulins prior to adding *L. salmonis* copepodites to the water. The copepodites attach immediately to the fish and the test aimed at 10–20 copepodites per fish 10–15 days after infection, when number of lice per fish was recorded at the end of chalimus II stage (Hamre et al., 2013). Lice count (LC) on the surface of the skin was recorded by manual counting. Average LC per fish was ~21 lice in Test 1 and ~13 in Test 2. However, the distribution of LC was highly skewed (Figures 1, 2), with some animals having extremely high infestations (up to 238 parasites on a single fish). Skewness in distribution of parasite abundance traits are frequently observed, and such traits are thus often analyzed on the log-scale (e.g., Robert et al., 1990; Morand and Guegan, 2000; Rozsa et al., 2000; Davies et al., 2006). Hence, LC was normalized through log-transformation (LogLC), defined as:

$$\text{LogLC} = \log_e (\text{lice count} + 1)$$

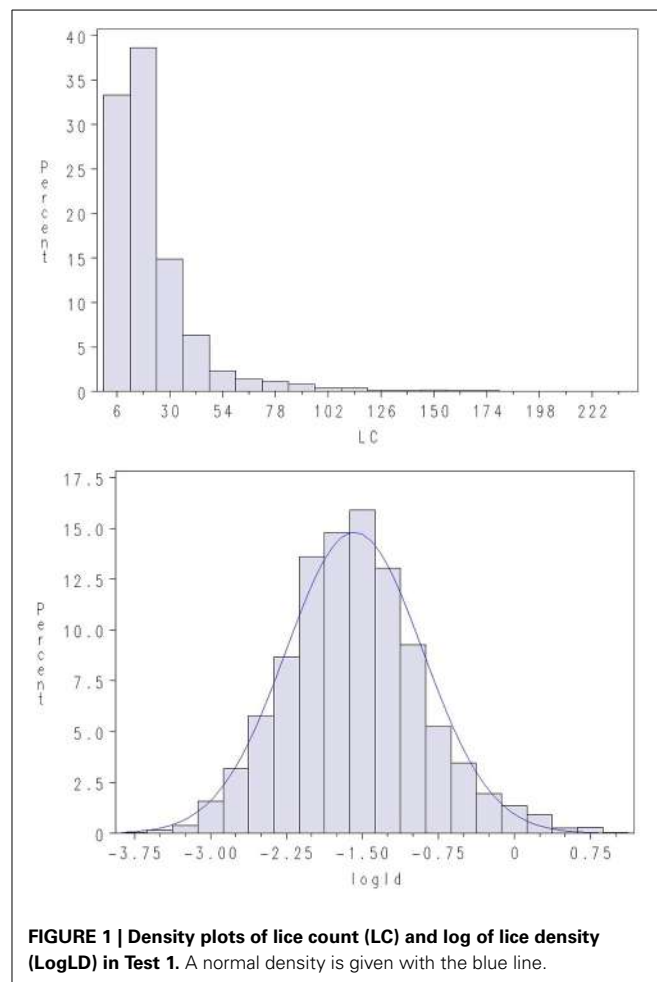


FIGURE 1 | Density plots of lice count (LC) and log of lice density (LogLD) in Test 1. A normal density is given with the blue line.

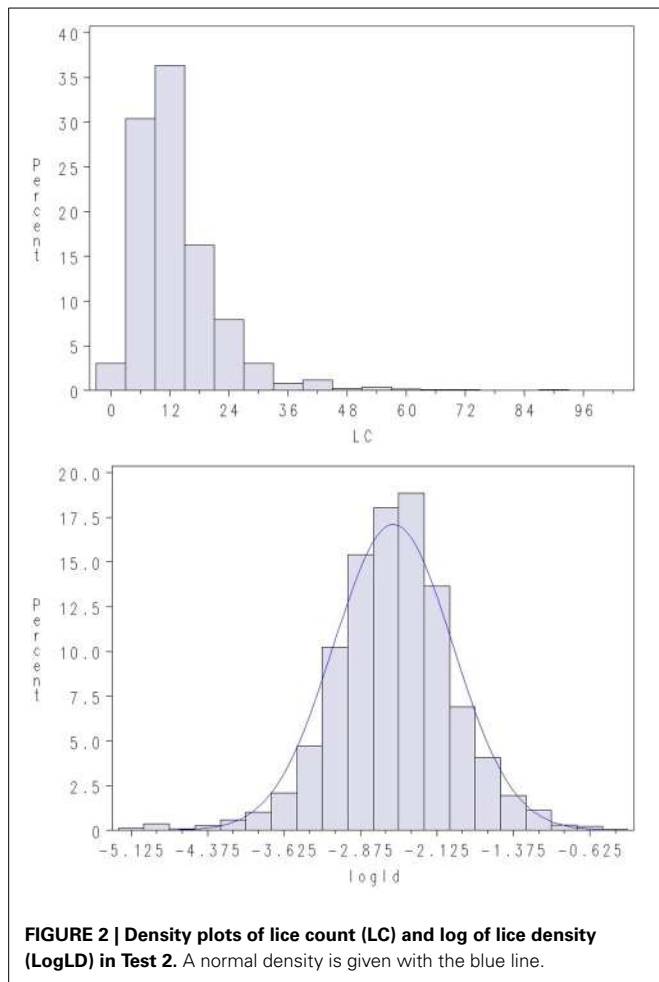
Lice counts were added a constant value of 1 to avoid computing errors due to fish with zero recorded lice. Furthermore, there is a tendency toward increasing LC with increasing body size (i.e., body surface) of the fish. Hence, Gjerde et al. (2011) developed an alternative measure of lice resistance, defined as estimated LiceD on the skin:

$$\text{LiceD} = \frac{\text{LC}}{\sqrt[3]{\text{BW}^2}}$$

where BW is body weight (g) at time of recording, and $\sqrt[3]{\text{BW}^2}$ is an approximate measure for the surface skin area of the fish. Still, considerable skewness was also observed for LiceD in the current dataset, while log-transformed LiceD (LogLD) was approximately normal (Figures 1, 2), indicating an approximate lognormal distribution of the trait:

$$\text{LogLD} = \log_e \left(\frac{\text{LC} + 1}{\sqrt[3]{\text{BW}^2}} \right)$$

Using the latter trait definition increased the estimated heritability (i.e., the fraction of variance explained by additive genetic



effects) increased substantially compared with a linear model applied to untransformed LiceD (results not shown).

FC was recorded in a subsequent slaughter test in April 2013, where fish originating from both lice challenge tests were jointly recorded for FC (majority of the recorded fish originated from Test 1). The trait FC was defined as the pigmentation (redness) of the fillet and was automatically measured using image analysis with PhotoFish equipment and software (Photofish AS, Ås, Norway). The recorded FC was found to be approximately normally distributed (Figure 3).

More descriptive statistics are shown in Tables 1–3.

GENOTYPING

A total of 1963 phenotyped individuals were genotyped with a 220 k Affymetrix genome-wide SNP-chip. About half the individuals in Test 1 were genotyped (1444 individuals), but a smaller fraction in Test 2 (519 individuals). The genotyping strategy was supposed to serve two purposes: (1) Application of GS; and (2) a genome-wide association study, potentially followed by marker-assisted selection (MAS) for the most significant SNP(s). The aim was to establish two experimental selection lines for, respectively, high and low sea lice resistance, using a combination of pedigree-based selection, GS and MAS. Hence, genotyping was not completely

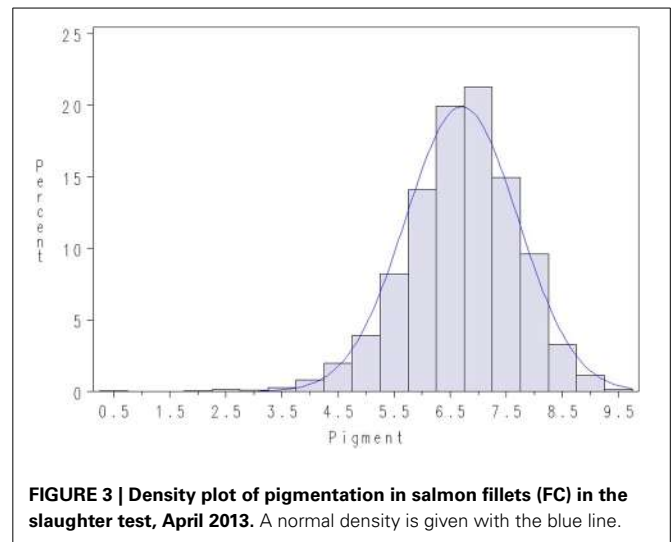


Table 1 | Descriptive statistics of data from fish participating in lice challenge test 1.

Variable	N	Mean	SD	Minimum	Maximum
BW (kg)	2850	0.80	0.29	0.1	1.98
LC	2850	20.96	19.68	1	238
LogLC	2850	2.83	0.70	0.69	5.48
LogLD	2850	-1.66	0.73	-4.40	1.00
FC	1426	6.80	0.98	0.32	9.56

Table 2 | Descriptive statistics of data from fish participating in lice challenge test 2.

Variable	N	Mean	SD	Minimum	Maximum
BW (kg)	2348	1.92	0.55	0.13	3.83
LC	2348	12.81	9.02	0	102
LogLC	2348	2.45	0.60	0.00	4.63
LogLD	2348	-2.55	0.58	-5.20	-0.35
FC	510	6.44	1.00	2.31	9.42

random, but particularly focused on the most extreme families (in both directions) with respect to lice resistance. Of the 1963 genotyped animals, 1869 had phenotype for FC.

STATISTICAL MODELS

A preliminary analysis showed that there was no significant genetic correlation between logLD and FC, and the two traits were therefore analyzed separately in subsequent analyses.

For logLD, an initial quantitative genetic analysis was run, treating phenotypes of the two lice tests as two correlated genetic traits. A bivariate linear animal model was used for analysis of the data:

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1\beta_1 + Z_1a_1 + e_1 \\ X_2\beta_2 + Z_2a_2 + e_2 \end{bmatrix},$$

Table 3 | Within-test Pearson correlation coefficient between the traits BW, LC, LogLC and LogLD, with coefficients for the tests 1 and 2 are given, respectively, above and below the diagonal.

Variable	BW	LC	LogLC	LogLD	FC
BW		0.20	0.27	-0.11	0.32
LC	0.23		0.86	0.81	0.07
LogLC	0.30	0.88		0.92	0.09
LogLD	-0.09	0.82	0.92		-0.03
FC	0.40	-0.01	0.03	-0.13	

where y_1 and y_2 are vectors of LogLD phenotypes from Test 1 and 2, respectively, β_1 and β_2 are vector of fixed effects (overall means of the two tests, and effect of observing person, nested within each effect), genetic effects are given in $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \sim N(0, A \otimes G_0)$,

residual effects are given in $\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} I\sigma_{e1}^2 & 0 \\ 0 & I\sigma_{e2}^2 \end{bmatrix}\right)$, Z_1 and

Z_2 are appropriate incidence matrices (assigning animal genetic effects to phenotypes), I is an identity matrix of appropriate size, A is the pedigree-based numerator relationship matrix, and G_0 is the additive genetic (co)variance matrix for the two genetic traits. As the two tests were performed on different individuals, the residual covariance between the two traits was assumed to be zero.

As the genetic correlation between logLD of the two tests was lower than unity (results shown below) and the majority of the genotyped animals came from lice challenge test 1, predictive ability of the genomic models for lice resistance was performed using phenotypes of the first test only. The FC was recorded on a later stage with fish originating both lice challenge tests, recorded at same age within the same slaughter test. FC of all fish was therefore analyzed jointly as a single genetic trait. Hence, predictive abilities of the different classical and genomic models for logLD (test 1) and FC were assessed using univariate animal models, with the following general characteristics:

$$y = Xb + Za + e$$

Where y is a vector of phenotypes (logLD or FC), $a \sim N(0, G\sigma_g^2)$ is a vector of random additive genetic effects, where G is a given relationship matrix (model dependent), and $e \sim N(0, I\sigma_e^2)$ is a vector of random residuals. The fixed effects (b) included person (responsible for counting) by day for logLD, and gender of fish for FC. Common environmental effects of family were also tested, but these effects were small and not significantly different from zero ($P > 0.20$) for both traits, and were thus dropped in the final model.

Univariate sub-models

The different models differed solely with respect to their specification of the relationship matrix G :

PED: Classical pedigree-based analysis, i.e., $G = A$ (numerator relationship matrix).

IBD-GS: Identity-by-descent GS, using a linkage-based IBD relationship matrix for the genotyped animals. The matrix was calculated from a sparse marker set containing 5590 mapped genome-wide SNP markers, using the LDMIP software (Meuwissen and Goddard, 2010). The number of mapped SNPs per chromosome varied from 52 to 396, and relationship matrices were thus computed for each chromosome separately and subsequently averaged over chromosomes to produce G .

GBLUP: Identity-by-state GS (ordinary GBLUP), calculating the G directly from genome-wide SNP markers using the second method by Vanraden (2007). Alternative G matrices were tested by extracting random sub-sets from the complete marker data set, including either (a) 1100 (1 K), (b) 2200 (2 k), (c) 4400 (4 k), (d) 22 000 (22 k), (e) 55 000 (55 k), or (f) all 220 000 (220 k) SNP markers, respectively. For (a) to (d) a total of 10 non-overlapping replicates (sub-sets of marker genotypes) were generated, while (e) was replicated 4 times. Results were averaged over replicates.

All models utilizing genomic information (IBD-GS and GBLUP) used one-step estimation of EBVs (Legarra et al., 2009; Christensen and Lund, 2010), combining relationships from genotyped and ungenotyped individuals into a unified relationship matrix H . Furthermore, the G matrices were adjusted to the same average rate of inbreeding and relationship as the numerator relationship matrix, using the ADJUST option in DMU (Christensen et al., 2012; Madsen and Jensen, 2013). Identical variance components were used in all models. These were estimated with the PED model using all phenotypic data.

Model comparison

Reliabilities of the different models were assessed through predictive ability, using five-fold cross-validation, i.e., individuals being both phenotyped and genotyped were randomly sampled into five validation sets, which were predicted one at a time, masking the phenotypes of the validation animals and using all the remaining phenotypes and genotypes as training data. Reliability was estimated as:

$$R_{EBV,BV}^2 = \frac{R_{EBV,y}^2}{h^2}$$

where $R_{EBV,y}^2$ is the squared correlation between EBVs of a given model (predicted from the training data, without the phenotype of the animal itself) and the recorded phenotype (y), while h^2 is the estimated heritability of the trait.

RESULTS

ESTIMATED HERITABILITIES AND GENETIC CORRELATIONS

Heritability of lice resistance (logLD) was estimated for the two tests using a bivariate PED model, as described in the above section. The estimated heritabilities for the two tests were low to moderate (0.14 ± 0.03 and 0.13 ± 0.03 for July and October, respectively), and the estimated genetic correlation between lice resistance in the two tests was high (0.72 ± 0.12). The estimated heritability of FC, based on an univariate PED model, was high

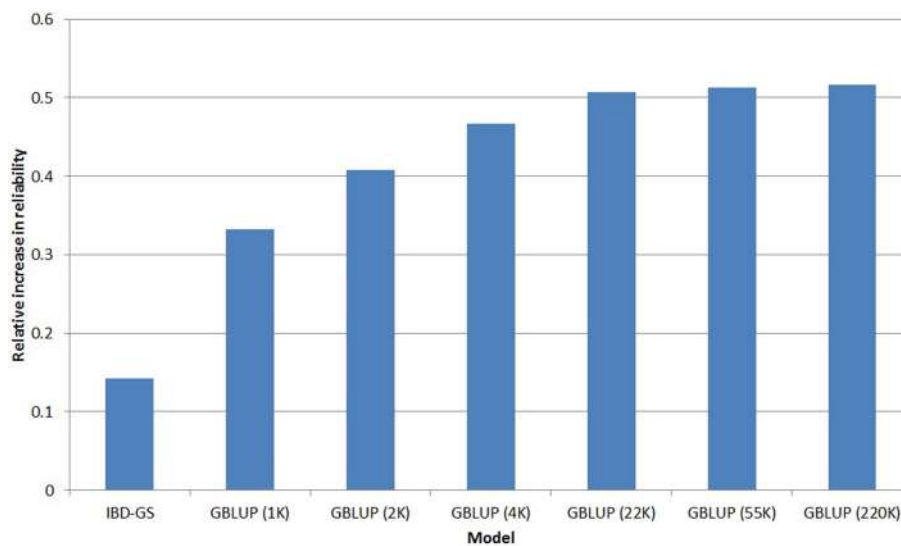


FIGURE 4 | Relative increase in reliability¹ of genomic selection models for LR compared with a classical pedigree-based model.

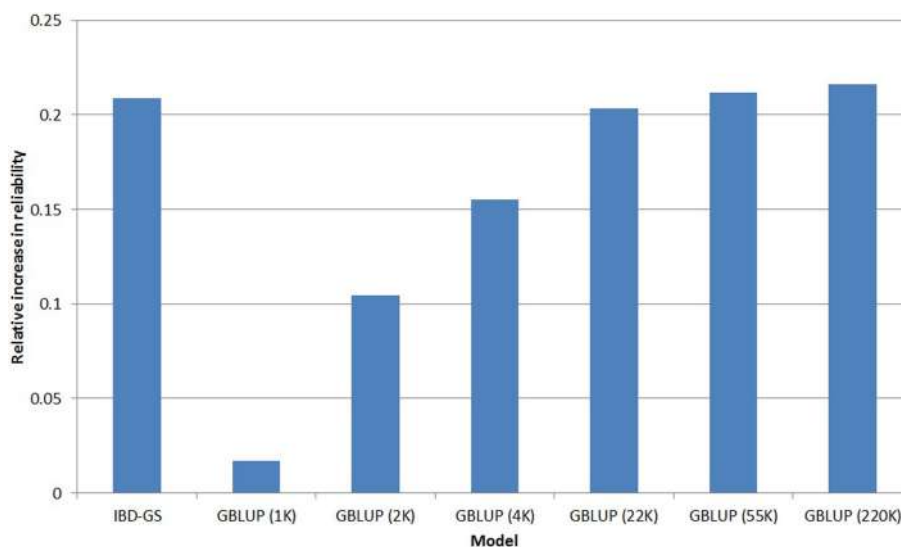


FIGURE 5 | Relative increase in reliability² of genomic selection models for FC compared with a classical pedigree-based model.

(0.43 ± 0.06). Based on likelihood ratio tests, genetic effects were highly significant for both traits ($P < 0.001$).

RELIABILITY OF DIFFERENT MODELS AND MARKER DENSITIES

Based on the five-fold cross validation, the reliability of the PED model was slightly higher for FC (0.36) than for lice resistance (0.34), the relative increases in reliabilities for the different GS models (compared with PED) are presented in **Figures 4, 5** for lice resistance and FC, respectively. In general, all GS models outperformed the classical PED model, but

the relative improvement varied considerably between models and traits. For lice resistance, the relative increase in reliability using GS was substantial (up to 52% for GBLUP with 220 k), but moderate for FC (21% for IBD-GS and 22% for GBLUP with 220 k).

Using GBLUP, higher marker densities were always favorable, but the relative advantage was considerably more expressed in FC than in lice resistance. For example, the relative increase in reliability of GBLUP for FC was 39% when going from 4 to 220 k, while the corresponding increase for lice resistance was only 11%. Nevertheless, GBLUP was superior to PED for both traits, even at the lowest marker densities (1 k). For both traits, going from 22 to 220 k SNPs increased reliability by only ~1%. Hence,

¹Reliability of LR using the PED model was 0.34.

²Reliability of FC using the PED model was 0.36.

increasing SNP density beyond 22 k would have little practical effect on selection.

Another striking result was the enormous difference between the traits with respect to the relative reliability of the IBD-GS model. For the lowly heritable lice resistance the relative improvement compared with a classical pedigree-based model was considerably lower for IBD-GS than (220 k) GBLUP (14 and 52% for IBD-GS and GBLUP, respectively). In contrast, for the more highly heritable FC, increases in reliability for the two models were similar (21 and 22% for IBD-GS and 220 k GBLUP, respectively).

DISCUSSION

The estimated heritabilities (0.14 ± 0.03 and 0.13 ± 0.03) for lice resistance (measured as log of LD) obtained in the current study was lower than recent estimates (0.26 ± 0.05) obtained for similar trait definitions (untransformed LD) and testing methods in a previous study on lice resistance in a different salmon population (Gjerde et al., 2011). It should be noted, however, that the previous test was conducted in tanks, while the current test was performed in sea-cages, with copepodids being added directly to the sea-cage.

As one of the aims of the project was to produce extreme high/low lines with respect to lice resistance, families with high/low lice resistance were over-represented among the genotyped animals, which may have some effect on the reliability of the PED and GS models for this trait. The analysis was in all cases validated based on genotyped animals, and extreme families for lice resistance are thus overrepresented among the validation animals, which is expected to inflate the between-family variation in the sample. In the PED model, predicted breeding values for animals with masked phenotypes is simply a function of the mid-parent means, and an inflation of the between-family variance in the training sample may thus increase the apparent reliability of the model. This may explain the relative small difference in reliabilities of the PED model for LR and FC (0.34 and 0.36, respectively), despite the considerable difference in heritability of the two traits (0.14 and 0.43, respectively). Despite this, the relative improvement of the reliability through GS was substantial for LR (up to 52%). For FC, selective genotyping with respect to LR had likely little impact, due to the low correlation between the traits.

The models used in this study utilize the sources of information contained in genomic data differently. The GBLUP model utilizes pedigree (implicitly contained in the genomic data), linkage analysis (animals sharing IBD chromosome segments will necessarily share marker alleles) as well as LD. However, its ability to utilize the different sources of information depends on marker density. IBD-GS utilizes pedigree and linkage analysis and is robust to marker density, while PED, by definition, utilizes the pedigree relationships only. For the GBLUP model, high marker density would be needed to capture both short-range LD and (tiny) variations in co-segregation among relatives. In contrast, the IBD-GS model will utilize linkage analysis information accurately, even at very low marker densities. Furthermore, the relative importance of the different types of information depends on several factors such as structure of the dataset (i.e., number of close

relatives in the population), historical N_e (i.e., amount of LD), as well as the heritability of the traits involved. In general, it is expected that for a lowly heritable trait, genetic effects estimated over larger groups of individuals, such as LD-associated effects (general association between marker genotypes and phenotypes) and mid-parent means would be more robust and thus relatively more important for the reliability, while linkage-analysis based deviations from pedigree relationships (i.e., largely minor individual deviations) would be relatively more important at higher heritabilities. Thus, the relative advantage of the GBLUP model may be largest at low heritability (e.g., lice resistance), while IBD-GS would be expected to perform relatively better at higher heritability (e.g., FC), which is consistent with results of this study. Another contributing factor may be the genetic architecture of the two traits; A major QTL has been published for FC (Baranski et al., 2010), and two more have recently been identified in the AquaGen population. All three QTL on FC were also detected in a genome-wide association study of the current data set, while, in contrast, no major QTL for lice resistance has been found (unpublished results). The GBLUP model assumes that genetic variance is uniformly distributed over the entire genome, which may fit better for lice resistance than FC. For this reason, more advanced Bayesian variable selection models (BayesB, BayesC, etc.) may have a larger potential in FC than LR.

Still, the factors discussed above do not explain the favorable performance of GBLUP for lice resistance at extremely low marker densities (e.g., 4 k), for which limited LD is usually expected (in homogenous populations), and linkage-based deviations from the expected relationships are unlikely to be accurately captured by IBS information. The explanation may thus lie in the selection history of farmed Atlantic salmon. As described in the introduction, admixture from several distinct wild strains is expected to introduce long-range LD, and simultaneously reduce the short-range LD in the population. This will likely reduce the relative advantage of dense SNP data, as a relatively larger fraction of the available LD may be captured even by sparse marker panels, potentially explaining the good performance of GBLUP for lice resistance even at extremely low marker densities. Current terrestrial livestock populations may also have been formed by admixtures of old populations, but these admixture events occurred longer ago and may have been less extreme than in Atlantic salmon. Still, some admixture effects on the LD structure may also be seen in terrestrial livestock species, and thus contribute to the rather small increases observed in accuracy of GS as marker density increases (Vanraden et al., 2011). A high marker density in GBLUP will be favorable for utilization of linkage analysis information, which is mainly an advantage at higher heritabilities and strong relationship structures (Ødegård and Meuwissen, 2014), i.e., as seen with FC.

The number of genotyped animals was rather limited in the current study. Genotyping larger fractions of the population would be expected to increase the reliability of GS models even further.

ACKNOWLEDGMENTS

The study was supported by the Norwegian Research Council through projects no. 200511/S40, 226266/E40 and 225181, and

by Regionale Forskningsfond Midt-Norge through project no. ES486711. We would also thank Bjarne Gjerde and Sissel Nergaard, Nofima for taking responsibility of the sea lice challenge tests.

REFERENCES

- Baranski, M., Moen, T., and Våge, D. (2010). Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genet. Sel. Evol.* 42:17. doi: 10.1186/1297-9686-42-17
- Calus, M. P. L., Meuwissen, T. H. E., De Roos, A. P. W., and Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178, 553–561. doi: 10.1534/genetics.107.080838
- Christensen, O., and Lund, M. (2010). Genomic prediction when some animals are not genotyped. *Gen. Sel. Evol.* 42:2. doi: 10.1186/1297-9686-42-2
- Christensen, O. F., Madsen, P., Nielsen, B., Ostensen, T., and Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal* 6, 1565–1571. doi: 10.1017/S1751731112000742
- Davies, G., Stear, M. J., Benothman, M., Abuagob, O., Kerr, A., Mitchell, S., et al. (2006). Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. *Heredity* 96, 252–258. doi: 10.1038/sj.hdy.6800788
- Gjedrem, T., Gjøen, H. M., and Gjerde, B. (1991). Genetic origin of Norwegian farmed Atlantic salmon. *Aquaculture* 98, 41–50. doi: 10.1016/0044-8486(91)90369-1
- Gjerde, B., Ødegård, J., and Thorland, I. (2011). Estimates of genetic variation in the susceptibility of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*. *Aquaculture* 314, 66–72. doi: 10.1016/j.aquaculture.2011.01.026
- Habier, D., Fernando, R., Kizilkaya, K., and Garrick, D. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190
- Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194, 597–607. doi: 10.1534/genetics.113.152207
- Hamre, L. A., Eichner, C., Caipang, C. M. A., Dalvin, S. T., Bron, J. E., Nilsen, F., et al. (2013). The salmon louse *Lepeophtheirus salmonis* (Copepoda: Caligidae) life cycle has only two chalimus stages. *PLoS ONE* 8:e73539. doi: 10.1371/journal.pone.0073539
- Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/S0016672308009981
- Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi: 10.3168/jds.2009-2061
- Luan, T., Wooliams, J. A., Ødegård, J., Dolezal, M., Roman-Ponce, S. I., Bagnato, A., et al. (2012). The importance of identity-by-state information for the accuracy of genomic selection. *Genet. Sel. Evol.* 44:28. doi: 10.1186/1297-9686-44-28
- Madsen, P., and Jensen, J. (2013). *DMU: A User's Guide. A Package for Analysing Multivariate Mixed Models*. Available online at: http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf
- Meuwissen, T. H. E., and Goddard, M. E. (2010). The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole-genome sequence density genotypic data. *Genetics* 185, 1441–1449. doi: 10.1534/genetics.110.113936
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368
- Morand, S., and Guegan, J. F. (2000). Distribution and abundance of parasite nematodes: ecological specialisation, phylogenetic constraint or simply epidemiology? *Oikos* 88, 563–573. doi: 10.1034/j.1600-0706.2000.880313.x
- Nielsen, H. M., Sonesson, A. K., Yazdi, H., and Meuwissen, T. H. E. (2009). Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture* 289, 259–264. doi: 10.1016/j.aquaculture.2009.01.027
- Ødegård, J., and Meuwissen, T. (2014). Identity-by-descent genomic selection using selective and sparse genotyping. *Genet. Sel. Evol.* 46:3. doi: 10.1186/1297-9686-46-3
- Ødegård, J., Yazdi, M. H., Sonesson, A. K., and Meuwissen, T. H. E. (2009). Incorporating desirable genetic characteristics from an inferior into a superior population using genomic selection. *Genetics* 181, 737–745. doi: 10.1534/genetics.108.098160
- Pfaff, C. L., Parra, E. J., Bonilla, C., Hiester, K., McKeigue, P. M., Kamboh, M. I., et al. (2001). Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am. J. Hum. Genet.* 68, 198–207. doi: 10.1086/316935
- Robert, F., Boy, V., and Gabrion, C. (1990). Biology of parasite populations: population dynamics of bothriocephalids (Cestoda-Pseudophyllidea) in teleostean fish. *J. Fish Biol.* 37, 327–342. doi: 10.1111/j.1095-8649.1990.tb05863.x
- Rozsa, L., Reiczgel, J., and Majoros, G. (2000). Quantifying parasites in samples of hosts. *J. Parasitol.* 86, 228–232. doi: 10.1645/0022-3395(2000)086[0228:QPISOH]2.0.CO;2
- Thomassen, J. R., Sørensen, A. C., Su, G., Madsen, P., Lund, M. S., and Guldbandsen, B. (2013). The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions. *J. Anim. Sci.* 91, 3105–3112. doi: 10.2527/jas.2012-5490
- Toosi, A., Fernando, R. L., and Dekkers, J. C. M. (2010). Genomic selection in admixed and crossbred populations. *J. Anim. Sci.* 88, 32–46. doi: 10.2527/jas.2009-1975
- Vanraden, P., O'Connell, J., Wiggans, G., and Weigel, K. (2011). Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10. doi: 10.1186/1297-9686-43-10
- Vanraden, P. M. (2007). Efficient estimation of breeding values from dense genomic data. *J. Dairy Sci.* 90, 374–375. doi: 10.3168/jds.2007-0980
- Vela-Avitúa, S., Meuwissen, T. H. E., Luan, T., and Ødegård, J. (in press). Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. *Genet. Sel. Evol.*
- Villanueva, B., Pong-Wong, R., Fernández, J., and Toro, M. A. (2005). Benefits from marker-assisted selection under an additive polygenic genetic model. *J. Anim. Sci.* 83, 1747–1752.
- Wientjes, Y. C. J., Veerkamp, R. F., and Calus, M. P. L. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193, 621–631. doi: 10.1534/genetics.112.146290

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 September 2014; accepted: 31 October 2014; published online: 21 November 2014.

Citation: Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S and Meuwissen THE (2014) Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Ødegård, Moen, Santi, Korsvoll, Kjøglum and Meuwissen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum*

David Marancik¹, Guangtu Gao¹, Bam Paneru², Hao Ma³, Alvaro G. Hernandez⁴, Mohamed Salem², Jianbo Yao³, Yniv Palti¹ and Gregory D. Wiens^{1*}

¹ National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV, USA

² Department of Biology, Middle Tennessee State University, Murfreesboro, TN, USA

³ Animal and Nutritional Sciences, West Virginia University, Morgantown, WV, USA

⁴ High-Throughput Sequencing and Genotyping Unit, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Edited by:

Scott Newman, Genus, plc, USA

Reviewed by:

Hugo Murua Escobar, Medical Faculty University of Rostock, Germany

Dan Nonneman, Agricultural Research Service, United States Department of Agriculture, USA

*Correspondence:

Gregory D. Wiens, National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, 11861 Leetown Rd., Kearneysville, WV 25430, USA
e-mail: greg.wiens@ars.usda.gov

Genetic improvement for enhanced disease resistance in fish is an increasingly utilized approach to mitigate endemic infectious disease in aquaculture. In domesticated salmonid populations, large phenotypic variation in disease resistance has been identified but the genetic basis for altered responsiveness remains unclear. We previously reported three generations of selection and phenotypic validation of a bacterial cold water disease (BCWD) resistant line of rainbow trout, designated ARS-Fp-R. This line has higher survival after infection by either standardized laboratory challenge or natural challenge as compared to two reference lines, designated ARS-Fp-C (control) and ARS-Fp-S (susceptible). In this study, we utilized 1.1 g fry from the three genetic lines and performed RNA-seq to measure transcript abundance from the whole body of naive and *Flavobacterium psychrophilum* infected fish at day 1 (early time-point) and at day 5 post-challenge (onset of mortality). Sequences from 24 libraries were mapped onto the rainbow trout genome reference transcriptome of 46,585 predicted protein coding mRNAs that included 2633 putative immune-relevant gene transcripts. A total of 1884 genes (4.0% genome) exhibited differential transcript abundance between infected and mock-challenged fish (FDR < 0.05) that included chemokines, complement components, tnfrsf receptor superfamily members, interleukins, nod-like receptor family members, and genes involved in metabolism and wound healing. The largest number of differentially expressed genes occurred on day 5 post-infection between naive and challenged ARS-Fp-S line fish correlating with high bacterial load. After excluding the effect of infection, we identified 21 differentially expressed genes between the three genetic lines. In summary, these data indicate global transcriptome differences between genetic lines of naive animals as well as differentially regulated transcriptional responses to infection.

Keywords: *Flavobacterium psychrophilum*, bacterial cold water disease, selective breeding, disease resistance, aquaculture, immune gene, tnfrsf, rainbow trout genome

INTRODUCTION

Selective breeding programs contribute to increased aquaculture production through the generation of animals with improved resistance/tolerance toward infectious disease causing microorganisms (Gjedrem, 1983, 2005; Van Muiswinkel et al., 1999; Cock et al., 2009; Gjedrem et al., 2012). In 2005, a family-based selective breeding program was initiated at the National Center for Cool and Cold Water Aquaculture (NCCCWA) to improve rainbow trout (*Oncorhynchus mykiss*) survival following exposure to *Flavobacterium psychrophilum* (Silverstein et al., 2009). This pathogen is the etiologic agent of bacterial cold water disease (BCWD) and rainbow trout fry syndrome (RTFS), and causes considerable losses to the rainbow trout aquaculture industry

within the U.S. and to trout and salmon populations worldwide (Nematollahi et al., 2003; Barnes and Brown, 2011). Infection of rainbow trout with *F. psychrophilum* typically results in mortality, ranging from 2 to 30% of the population, with higher losses caused by co-infection with the infectious hematopoietic virus. A further impact of the disease is that survival following infection has been associated with skeletal deformities (Kent et al., 1989; Madsen et al., 2001). Disease prevention is difficult as the pathogen is geographically wide spread, limited chemotherapeutics are available for treatment, and there is currently no commercial vaccine available in the U.S, although killed, sub-unit, and live-attenuated vaccines are all active areas of research (Gómez et al., 2014; Sundell et al., 2014).

A pedigreed line of rainbow trout, designated ARS-Fp-R, has been subjected to over three generations of selection and demonstrates increased survival following experimental injection challenge (Hadidi et al., 2008; Silverstein et al., 2009; Leeds et al., 2010) and natural exposure (Wiens et al., 2013a), relative to a disease susceptible line, ARS-Fp-S, and a randomly bred control line, ARS-Fp-C. Current research goals at the NCCCWA includes elucidating intrinsic factors associated with survival of the ARS-Fp-R line to better understand mechanisms of how selection has altered the genetic control of disease resistance. Phenotypic studies have thus far shown that ARS-Fp-R line fish have decreased organ damage as determined by histopathology (Marancik et al., 2014b) and fewer pathophysiologic changes in plasma biochemistry (Marancik et al., 2014a) during the acute-phase of disease following experimental challenge. Experiments that quantified splenic *F. psychrophilum* numbers on days 5 (Hadidi et al., 2008) and 9 post-infection (Marancik et al., 2014a) demonstrate significantly lower splenic bacterial loads in ARS-Fp-R line fish. It is likely these observed differences are a result of a differential immune response to infection.

Changes associated with host immunologic response can be elucidated by profiling alterations in host mRNA abundance between pathogen naive and infected animals (Martin et al., 2006; Beck et al., 2012; Langevin et al., 2012; Peatman et al., 2013; Pereiro et al., 2014; Shi et al., 2014). Previous studies of rainbow trout infected with *F. psychrophilum* demonstrate significant upregulation and downregulation of rainbow trout immune-relevant genes in the limited number of tissues examined (Overturf and LaPatra, 2006; Villarroel et al., 2008, 2009; Evenhuis and Cleveland, 2012; Langevin et al., 2012; Orioux et al., 2013; Henriksen et al., 2014). Microarray analysis of head kidney tissue from susceptible compared to resistant double-haploid rainbow trout lines, identified differences in basal gene expression as well as induction of antimicrobial peptides, complement, matrix metalloproteases, and chemokines 5 days post-infection (Langevin et al., 2012). Taken together, these studies suggest that the rainbow trout immune response to *F. psychrophilum* is likely multifactorial involving both innate and adaptive components.

In this manuscript, we quantify changes in gene transcript abundance between genetic lines with the following goals: (1) identify differentially regulated genes common to the host response to *F. psychrophilum* infection, (2) identify genes differentially regulated between lines in response to *F. psychrophilum* infection, and finally, (3) examine baseline differences in gene expression between naive animals that might contribute to the post-challenge phenotype. For this study, we utilized a comprehensive RNA-seq, transcriptome approach, starting with whole-body lysates from 1.1 g fry from pooled fish of naive or experimentally infected genetic lines. Experimental infection during the fry/juvenile life-stage allowed gene transcript profiling at a time when genetic lines express robust survival differences, and when epizootics and mortality are described as most severe in production environments (Branson, 1995; Decostere et al., 2000). Sequence reads were aligned to the recently released rainbow trout genome transcriptome (Berthelot et al., 2014) to which we added automated annotation and manual immune gene curation. In summary, we described common, whole-body gene transcriptional responses

to early *F. psychrophilum* infection and potential differences associated with the innate response between genetic lines, and finally, compare our results with published gene expression studies utilizing *F. psychrophilum* challenged rainbow trout.

MATERIALS AND METHODS

ETHIC STATEMENT

Fish were maintained at the NCCCWA and animal procedures were performed under the guidelines of NCCCWA Institutional Animal Care and Use Committee Protocols #053 and #076.

EXPERIMENTAL ANIMALS

The ARS-Fp-R, ARS-Fp-C, and ARS-Fp-S genetic lines were derived from the same base population, and thus differed only as a result of artificial selection for BCWD post-challenge survival (Wiens et al., 2013a). Single-sire × single-dam matings were made within genetic lines between 3-year-old females and 1-year-old neo-males as previously described (Marancik et al., 2014b). Water temperatures in the egg incubation jars were manipulated so that all families hatched within a 1-week period (Leeds et al., 2010). Eggs were pooled within-line at the eyed stage and reared in ~12.5°C flow-through spring water. The ARS-Fp-R egg pool consisted of contributions from 43 full-sib families, the ARS-Fp-C egg pool consisted of 10 full-sib families, and the ARS-Fp-S egg pool consisted of 11 full-sib families. The resistant line eggs were progeny of dams that had undergone three generations of BCWD selection while the sires had undergone four generations of selection. The control-line eggs were progeny of parents that had undergone one generation of selection for increased resistance (2007 year class) and since that time, randomly bred. The susceptible-line eggs were progeny of parents that had undergone one generation of selection for increased susceptibility (2007 year class) and since that time, randomly bred (Wiens et al., 2013a). Eggs were pooled from a larger number of resistant line families as more resistant families are generated within the breeding program to apply selection differential, and thus this sampling design more accurately captured the genetic diversity within the resistant line. In addition, the resistant-line egg pool was part of a germplasm release and was utilized in additional challenge studies that will be reported elsewhere (Wiens, unpublished data).

All brood-stock and fish used in this study were certified to be free of common salmonid bacterial and viral pathogens by two independent diagnostic laboratories as described previously, and were negative for *F. psychrophilum* infection (Leeds et al., 2010; Wiens et al., 2013a). Prior to challenge, fry were allowed 1 week to acclimatize to challenge tanks. Mean body weight of the ARS-Fp-R line fish was 1.11 ± 0.05 g, the ARS-Fp-C line was 1.12 ± 0.03 g, and the ARS-Fp-S line was 0.98 ± 0.04 g (± 1 SD, pooled weights of $n = 4$ tanks) at the initiation of the challenge. In the challenge facility, photoperiod was adjusted weekly to maintain a natural lighting cycle, and at the time of RNA-seq sample collection, was 14.5 h light: 9.5 h dark. Water quality parameters have been described previously (Wiens et al., 2013a).

RNA-seq EXPERIMENTAL DESIGN

Bacterial challenge was carried out in the NCCCWA challenge facility with *F. psychrophilum* strain CSF-259-93 (initially

provided by Dr. S. LaPatra, Clear Springs Foods, Inc.). This strain was previously isolated from a BCWD field-case and maintained at -80°C in TYES media supplemented with 10% (v/v) glycerol and has been consistently utilized as the challenge strain within the selective breeding program (Hadidi et al., 2008; Silverstein et al., 2009; Leeds et al., 2010; Wiens et al., 2013a) and the complete genome sequence determined (Wiens et al., 2014). Frozen stock was cultivated on TYES media for 5 days at 15°C , suspended in PBS and O.D.₅₂₅ adjusted to 0.4. Colony plate counts were performed in triplicate and recorded after 5 days incubation to estimate the challenge dose enumerated as viable CFU fish⁻¹.

Each tank held fifty, randomly assigned fish and were supplied with 2.4 L min^{-1} of $12.5 \pm 0.1^{\circ}\text{C}$ flow-through spring water. For each genetic line, two tanks of fish were challenged by *F. psychrophilum* injection and two tanks of fish were challenged by PBS injection and served as non-infected control animals. In total, 100 fish per line were anesthetized with 100 mg/L tricaine methanesulfonate (Tricaine-S, Western Chemical, Inc., Ferndale, WA) and intraperitoneally (IP) injected with 4.2×10^6 CFU fish⁻¹ *F. psychrophilum* suspended in 10 μL of chilled PBS or 10 μL of chilled PBS alone. Injections were performed using a repeater pipette (Eppendorf, Hauppauge, NY) fitted with a 27G \times 1/2 inch needle. Fish age at the time of challenge was 49 days post-hatch (617 temperature degree days).

Five fish were sampled per tank on day 1 and on day 5 post-injection for RNA extraction. Survival prior to and following sampling was monitored daily for 21 days with the exception of one PBS injected tank that was excluded on day 16 due to water failure. All fish were fed daily a standard commercial fishmeal-based diet by hand (Ziegler Bros, Inc., USA). The day 1 sampled fish were removed prior to being fed and the day 5 sampled fish were removed after being fed.

RNA EXTRACTION, LIBRARY PREPARATION, AND SEQUENCING

The sampled fish were euthanized with 150 mg mL⁻¹ tricaine methanesulfonate and individually flash frozen in liquid nitrogen and stored at -80°C . Total RNA was extracted from individual whole, ground fish using the standard TRIzol protocol (Invitrogen, Carlsbad, CA). Total RNA was extracted and integrity confirmed by running a 1% agarose gel. Equal amounts of RNA from five fish were pooled from each of the 12 tanks at each of the two time points (total of 24 pools, $n = 120$ fish total). The cDNA libraries were prepared using Illumina's TruSeq Stranded mRNA Sample Prep kit following the manufacturer's instructions. Briefly, mRNA was selected with oligo(dT) beads and chemically fragmented to a size of $\sim 100\text{--}400$ nt before annealing of random hexamers and first strand cDNA synthesis. The 24 indexed and barcoded libraries were randomly divided into three groups (eight libraries per group) and sequenced in three lanes of an Illumina HiSeq 2000 (single-end, 100 bp read length) at the University of Illinois at Urbana-Champaign. All raw RNA-seq reads were submitted to the NCBI Short Read Archive under accession number BioProject ID PRJNA259860 (accession number SRP047070). RNA sequence reads that matched the *F. psychrophilum* CSF259-93 genome sequence (GenBank accession CP007627.1) were separated from host RNA-seq data and counted in each library.

Frozen fish homogenate lysates (500 μL), stored at -80°C , were individually processed to isolate genomic DNA (TRIzol DNA Isolation Procedure) and qPCR detection of *F. psychrophilum* genomic DNA was performed as described previously (Marancik and Wiens, 2013). Bacterial genome equivalents were normalized to per 100 ng⁻¹ extracted DNA.

GENES DESCRIBED AS REGULATED IN RESPONSE TO

F. PSYCHROPHILUM CHALLENGE

Based on a meta-analysis of studies in which rainbow trout were challenged with *F. psychrophilum*, 23 genes encoding immune relevant factors with putative roles in inflammation, innate disease response, and adaptive immunity were utilized as a curated gene reference set that included *cd3*, *cd8*, *mx-1* (Overturf and LaPatra, 2006), *saa* (Villarreal et al., 2008), *igm*, *igt*, *inf- γ* , *il-8*, *tr- β* , *tlr5*, *tnf- α* (Evenhuis and Cleveland, 2012), *mt-a*, *sod-1*, *tgf- β* (Orieux et al., 2013), and *cd4*, *il-1 β* , *il-6*, *il-17c1*, *il-17c2*, *IL-4/13A*, *foxp3b*, *mhc-I*, and *mhc-II* (Henriksen et al., 2014).

RAINBOW TROUT REFERENCE GENOME, GO ANNOTATION, AND IDENTIFICATION OF IMMUNE RELEVANT GENES

In addition to the curated reference gene dataset described above, the recently released rainbow trout genome sequence accession CCAF000000000 (Berthelot et al., 2014) encoding 46,585 predicted coding mRNA sequences were utilized as a more complete reference set of gene models. Briefly, the *Onchorhynchus mykiss pep.fa* file (April 24, 2014 release), accessed at (<http://www.genoscope.cns.fr/trout-ggb/data>), was subjected to Blast2GO sequence annotation pipeline with the default parameters applied in the blast, mapping, and annotation steps using a local Blast2GO database (GO database file *go_201401-assocdb-data*). NCBI non-redundant protein sequence (nr) database (build 12/13/2013) was used as the reference in the blast step. Of the 46,585 protein-coding gene models with supporting protein evidence from other vertebrates, a total of 46,103 were assigned a "Sequence Description" from the blast step and most of these predicted protein sequences assigned GO terms (Supplementary Data Sheet 1) from the Blast2GO annotation step. A subset of 2633 genes were identified as "putative immune relevant" either by sorting genes identified by GO annotation as "Immune Response" or by manual annotation based on sequence description (Supplementary Data Sheet 2). We excluded from analyses gene models encoding long non-coding RNAs and microRNAs. Splice variants were not analyzed and will be the subject of further study.

IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES

To identify the differentially expressed genes, reads from each library were mapped against the annotated gene database and the coding sequences from the rainbow trout genome assembly. Based on the high quality score distribution of the RNA-seq reads, the whole 100 bp of the sequences were used in this step. Bowtie short read aligner (Langmead et al., 2009) was used in mapping the reads to the references with a maximum of two mismatches allowed and no gaps. The output of Bowtie was filtered with an in-house Perl script to generate a count table, where a

number at the corner of row i and column j represents the number of total mapped reads to the transcript i from the library j . For principle component analyses (PCA), reads were normalized to reads per kilobase of exon model per million mapped reads (RPKM) using an in-house script (Gao, available upon request) using the formula described previously (Mortazavi et al., 2008).

$$\begin{aligned} RPKM &= N_{read} / (L_{exon} / 10^3) / (N_{total} / 10^6) \\ &= 10^9 N_{read} / (L_{exon} N_{total}) \end{aligned}$$

Where N_{read} is defined as the number of reads mapped to the gene; L_{exon} is defined as the total bases of the sequence of the gene, and N_{total} is defined as the total number of reads mapped to the whole reference (sum of the N_{read} for all genes). Raw data and normalized RNA-seq data (RPKM) for all samples including immune relevant genes are included as Supplementary Data (Supplementary Data Sheet 3).

Principal component and nearest neighbor network analyses were performed using Qlucore Omics Explorer (v3.0). Read count data from the 24 samples (RPKM) were \log_2 transformed and subjected to normalization (mean = 0 and variance = 1) and variables (genes) subjected to multi-group comparison with a false discovery rate (FDR) $q < 0.05$ ($R^2 \geq 0.36$) and $p < 0.002$ [$F_{(1, 22)} \geq 12.42$].

The raw count table was also input into the R package DESeq2 (Love et al., 2014), datasets selected for pair-wise comparisons, and the standard differential analysis steps of DESeq2 applied to the selected datasets. The output table of DESeq2 contains a column of adjusted p -value (p_{adj}) obtained using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), and we utilized a cut-off of $p_{adj} < 0.05$ as a criteria for differential expression with no filtering based on fold-change. For more stringent data filtering and visualization, data were first sorted by variance 2.5%, then filtered by $q < 0.01$ and \log_2 fold change of >3 (Qlucore Omics Explorer v3.0).

GO ANNOTATION ENRICHMENT ANALYSES

The GOSSIP program embedded in the Blast2GO package was used for GO enrichment analysis (Bluthgen et al., 2005a). This program examines each GO term for gene annotation enrichment by comparing a test set with a reference set using Fisher's exact-test method. In our analysis, the differentially expressed genes were selected as the test set and all the peptide sequences from the rainbow trout genome assembly for which GO terms were assigned, excluding the test set, were used as the reference set. FDR is computed by GOSSIP using an analytical equation (Bluthgen et al., 2005b) and the final list of the enriched GO terms selected at the FDR $< 5\%$.

VALIDATION OF RNA-seq DATA BY qPCR

Six genes and primer sets were chosen for qPCR analysis to determine fold-change differences between control and infected fish on day 5 post-infection (Table S1). Extracted RNAs were treated with Optimize™ DNAase I (Fisher Bio Reagents, Hudson, NH) to remove contaminating genomic DNA. One microgram

of DNAase treated RNA was used to make cDNA in a total reaction volume of 20 μ L. cDNA was synthesized using the Verso cDNA Synthesis Kit (Thermo Scientific, Hudson, NH) following the manufacturer described protocol. Reverse transcription reaction was carried out using My Cycler™ Thermal Cycler (Bio Rad, Hercules, CA) at 42°C for 30 min (one cycle amplification) followed by 95°C for 2 min (inactivation). Anchored oligo(dT) primer, at a final concentration of 25 ng/ μ L, was used to prime the reverse transcription reaction.

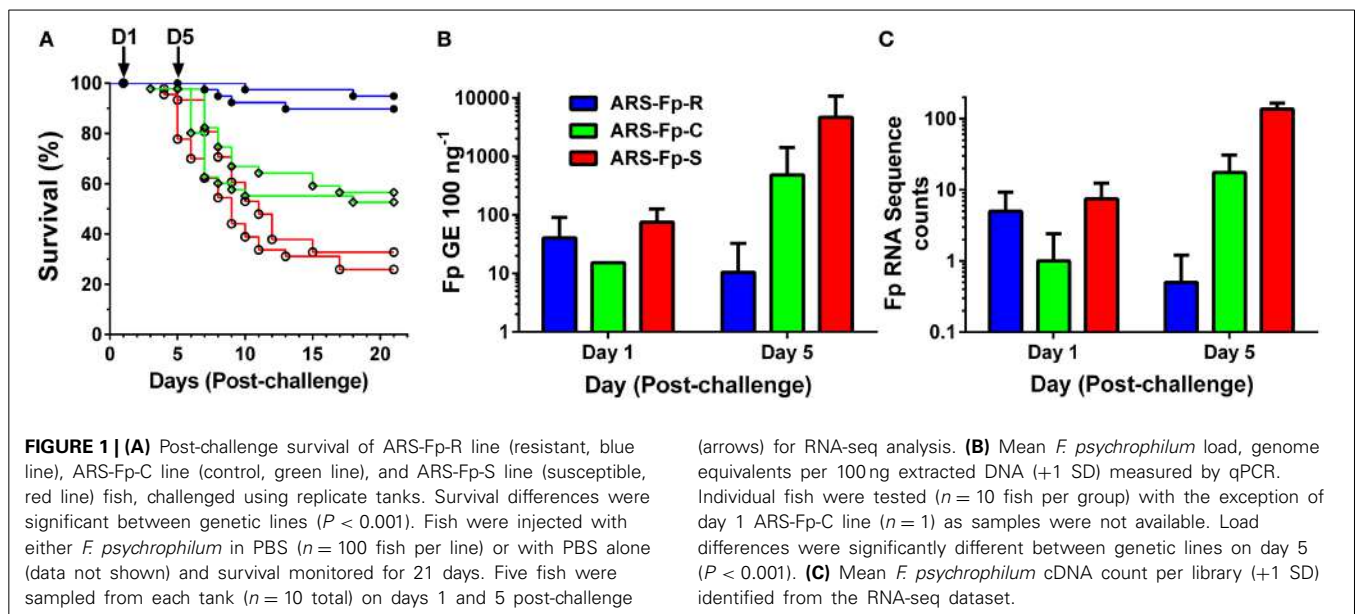
Relative gene expression was determined by CFX96™ Real Time System (Bio Rad, Hercules, CA). Forward and reverse primer sequences were manually aligned to their respective genes for validation. cDNA amplification was performed using DyNAmo Flash SYBR Green Master Mix (Thermo Scientific, Hudson, NH) containing 0.1 nm/ μ L forward and reverse primers and 0.0025 μ g/ μ L of cDNA template in a total reaction volume of 20 μ L. Initial denaturation was done at 95°C for 7 min. Forty-three cycles of amplification were carried out at the condition: 95°C for 0.1 min (denaturation), 57–64°C (annealing and extension) and 60°C for 5 min (final extension). Different annealing temperatures were used for different primers depending on their melting temperature (Table S1).

Real time data were analyzed using the software Bio-Rad CFX Manager (Bio Rad, Hercules, CA). Differential gene expression was calculated using the standard curve method in which β -actin (Accession: AJ438158) was used as endogenous reference to normalize the target gene. β -actin expression levels demonstrated in RNA-seq data were similar in PBS and *F. psychrophilum*-injected fish (data not shown). qPCR data were quantified using delta delta Ct ($\Delta\Delta$ Ct) methods (Schmittgen and Livak, 2008). Ct-values of β -actin were subtracted from Ct-values of the target gene to calculate the normalized value (Δ Ct) of the target gene in both the calibrator samples (PBS-injected) and test samples (*F. psychrophilum*-injected). The Δ Ct value of the calibrator sample was subtracted from the Δ Ct value of the test sample to get the $\Delta\Delta$ Ct value. Fold change in gene expression in the test sample relative to the calibrator sample was calculated by the formula $2^{-\Delta\Delta$ Ct} and the normalized target Ct values in each infected and non-infected group was averaged. Correlation between gene expression fold-change measured by qPCR and RNA-seq was performed by Pearson correlation. All statistics were performed with a significance of $P < 0.05$.

RESULTS

DISEASE RESISTANCE PHENOTYPE AND *F. PSYCHROPHILUM* LOAD IN THREE GENETIC LINES FOLLOWING EXPERIMENTAL CHALLENGE

Following injection challenge, survival significantly differed between the three genetic lines ($P < 0.001$) with the ARS-Fp-R line exhibiting highest survival (92.3%), the ARS-Fp-C line intermediate survival (54.6%), and the ARS-Fp-S line exhibiting the lowest survival (29.4%) (Figure 1A). The 62.9 percentage point survival difference between the ARS-Fp-R and ARS-Fp-S lines is consistent with the expected percentage point survival difference calculated from estimated parental breeding values (data not shown), analyzed as previously described (Leeds et al., 2010). Survival of PBS injected fish was $>98\%$ per line over the 21 days challenge period. Whole body *Fp* loads, measured by qPCR,



demonstrated no difference between genetic lines on day 1 but significant differences on day 5 ($P < 0.001$), with ARS-Fp-R line mean loads of 10 ± 22 *F. psychrophilum* genome equivalents (GE) ($n = 10$ fish, ± 1 SD), increasing to 484 ± 937 GE in the ARS-Fp-C line ($n = 10$, ± 1 SD), and highest mean load was present in the ARS-Fp-S line 4683 ± 6011 GE ($n = 10$, ± 1 SD) (Figure 1B). In the RNA-seq libraries, a small number of sequences were present that matched the *F. psychrophilum* CSF259-93 genome. These presumably represent either cDNA from bacterial expressed genes or contaminating genomic DNA. The abundance of the *F. psychrophilum* read counts were similar between genetic lines on day 1, and by day 5, lowest in ARS-Fp-R line fish and highest in ARS-Fp-S line fish (Figure 1C). In summary, by day 5 post-challenge the phenotypes of the ARS-Fp-R, ARS-Fp-C and ARS-Fp-S lines displayed expected differences in pathogen load and subsequently post-challenge mortality.

RAINBOW TROUT RNA-seq REFERENCE DATASET, GO ANNOTATION, AND MANUAL IMMUNE GENE CURATION

A total of 520 million sequence reads were generated from the 24 libraries with an average of 21.7 million (M) RNA-seq reads per library ranging from 17.4 M to 24.4 M reads (Table 1). Approximately half of the reads aligned to the reference transcriptome coding sequence, averaging 11.2 M (51.8%) per library with a range of 9.3 M–12.5 M. Of the 46,585 genes identified in the rainbow trout genome having protein coding evidence, an average of 43,068 (92.4%), exhibited detectable levels of expression, defined as >1 sequence read per gene (Table 1). The number of genes with an average of ≥ 10 read counts across the 24 libraries fell to an average of 32,830 (70.5%). In order to normalize gene expression across samples, data were converted to reads per kilobase of exon model per million mapped reads (RPKM) format (Supplementary Data Sheet 3). Of the putative immune relevant genes, 1797 (68.2%) had an average of ≥ 1 RPKM across the 24 libraries (Supplementary Data Sheet 3).

GLOBAL TRANSCRIPT ABUNDANCE DIFFERENCES BETWEEN CONTROL AND INFECTED AND BY SAMPLE DAY

Across the complete dataset, comparison of *F. psychrophilum* infected vs. PBS injected groups by principal component and nearest neighbor network analysis identified samples grouped by day and infection status (Figure 2). Most tank replicates were connected by nearest-neighbor analysis although there was variation observed between tanks. The replicate tanks for the day 5 PBS injected ARS-Fp-S line (red colored balls), day 1 *F. psychrophilum* infected ARS-Fp-R line (blue colored balls), and day 5 *F. psychrophilum* ARS-Fp-C line (green colored balls) grouped together but were not directly connected by the network analysis.

DIFFERENTIALLY EXPRESSED GENES COMMON TO THE THREE GENETIC LINES IN RESPONSE TO *F. PSYCHROPHILUM* INFECTION

Two-group comparison by infection (Qlucore) identified a total of 1884 genes as differentially regulated ($q < 0.05$) accounting for approximately 4.04% of the coding genes identified in the genome (see Table S2, Qlucore $q < 0.05$ for the complete list). Of the differentially regulated genes, 279 (14.8%) were categorized as immune relevant by GO or manual annotation. In order to more precisely identify differences between samples, 24 pairwise comparisons were tabulated by DESeq2 using a $q < 0.05$ (Table 2). There was a decrease in the number of differentially regulated genes in the infected ARS-Fp-R line fish compared to PBS injected fish between day 1 ($n = 515$ genes) and day 5 ($n = 428$ genes) time points. In contrast, the number of differentially regulated genes in the infected ARS-Fp-C line fish compared to PBS injected fish, increased from day 1 ($n = 20$ genes) to day 5 ($n = 2201$ genes). The number of differentially regulated genes in the infected ARS-Fp-S line fish compared to PBS injected fish increased from day 1 ($n = 1663$) to day 5 ($n = 2225$). In general, there were relatively few differentially regulated genes between the three PBS injected genetic lines, ranging from 3 to 246 genes (Table 2). Bacterial challenge increased the number

Table 1 | Summary statistics of 24 RNA-seq libraries listed by genetic line, time, infection status, and tank replicate.

Genetic line	Day	Infection, Tank	Biosample Accession No	Total reads	Mapped reads	Percentage reads mapped	Uniquely mapped reads	Number of expressed genes ^a
ARS-Fp-R	1	Fp, Tk25	SAMN03014722	20,061,852	10,210,344	50.89%	9,552,696	42,919
		Fp, Tk26	SAMN03014723	20,226,280	10,591,198	52.36%	9,963,273	42,044
	5	Fp, Tk25	SAMN03014726	21,409,329	10,973,051	51.25%	10,445,614	43,000
		Fp, Tk26	SAMN03014727	23,958,681	11,795,961	49.23%	11,135,354	43,654
	1	PBS, Tk27	SAMN03014724	22,129,914	11,745,879	53.08%	10,998,947	43,303
		PBS, Tk28	SAMN03014725	23,909,110	12,294,973	51.42%	11,410,381	43,567
	5	PBS, Tk27	SAMN03014728	24,361,298	12,124,063	49.77%	11,507,535	43,446
		PBS, Tk28	SAMN03014729	23,318,224	11,741,589	50.35%	11,171,392	43,504
ARS-Fp-C	1	Fp, Tk33	SAMN03014738	20,940,097	11,050,478	52.77%	10,225,748	43,229
		Fp, Tk34	SAMN03014739	19,151,755	10,393,600	54.27%	9,629,194	42,162
	5	Fp, Tk33	SAMN03014742	21,117,398	10,817,377	51.22%	10,299,653	42,904
		Fp, Tk34	SAMN03014743	21,763,498	10,846,416	49.84%	10,311,809	43,308
	1	PBS, Tk35	SAMN03014740	20,314,994	11,045,087	54.37%	10,314,237	41,735
		PBS, Tk36	SAMN03014741	20,372,060	10,581,232	51.94%	9,871,584	43,173
	5	PBS, Tk35	SAMN03014744	24,480,995	12,593,510	51.44%	11,896,143	43,345
		PBS, Tk36	SAMN03014745	20,535,582	10,422,413	50.75%	9,824,416	43,013
ARS-Fp-S	1	Fp, Tk29	SAMN03014730	17,389,164	9,377,592	53.93%	8,617,417	42,438
		Fp, Tk30	SAMN03014731	21,095,859	11,103,652	52.63%	10,198,695	42,765
	5	Fp, Tk29	SAMN03014734	21,467,982	11,073,977	51.58%	10,409,769	43,263
		Fp, Tk30	SAMN03014735	23,427,268	11,922,324	50.89%	11,172,000	43,479
	1	PBS, Tk31	SAMN03014732	20,622,290	11,310,300	54.85%	10,482,544	42,903
		PBS, Tk32	SAMN03014733	24,139,642	12,522,402	51.87%	11,594,067	43,561
	5	PBS, Tk31	SAMN03014736	22,371,928	11,319,845	50.60%	10,753,477	43,607
		PBS, Tk32	SAMN03014737	22,338,190	11,425,396	51.15%	10,750,682	43,319
Average				21,704,308	11,220,111	51.77%	10,522,359	43,068

Total number of 100 bp reads and the number that match reference genes are listed.

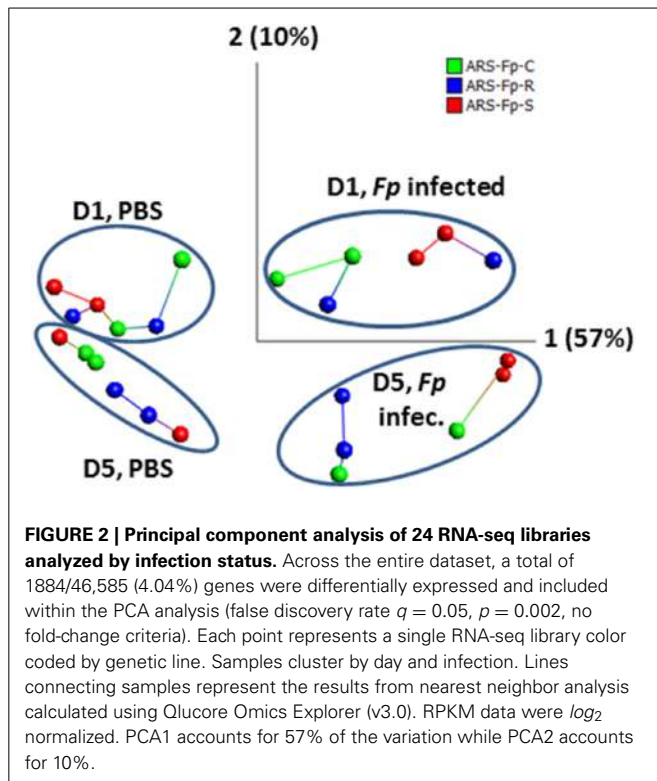
^aThe number of expressed genes is defined as genes having at least one read count with 2 bp or less mismatch and no gaps ($n = 46,585$ reference genes).

of differentially expressed genes between genetic lines from day 1 to day 5 in all between-line comparisons. There was a strong sampling time effect within each line including PBS injected groups possibly due to differences in feeding or post-injection recovery.

Analysis of GO term enrichment within the pair-wise comparisons revealed a larger number of over-represented terms within the dataset as compared to under-represented terms (Table 3). Overrepresented processes in the ARS-Fp-S line included defense response to bacterium, inflammatory response, leukotriene and arachidonic acid production, complement activation, humoral immune response, antigen processing, chemotaxis, B cell homeostasis, interleukin-2 mediated pathway signaling and cellular iron homeostasis (Supplementary Data Sheet 5 for complete list). The ARS-Fp-R line Day 1 GO term enrichment included several categories that include response to wounding and wound healing and was enriched for cytokine and chemokine activity. ARS-Fp-R line day 5 GO term enrichment included complement activation, inflammatory response, B cell mediated immunity and defense response to bacterium.

Pair-wise analysis of differentially regulated genes shared between infected and PBS injected fish within each sampling day identified no common genes on day 1, although 274 were

shared between ARS-Fp-R and ARS-Fp-S line fish (Figure 3). This may in part be due to the low numbers of genes differentially expressed between the ARS-Fp-C line replicates. Among all day five samples, 175 common genes were differentially regulated, by pair-wise comparisons, in all three lines (see Table S2, Pair-wise for the complete list). Of these, the majority (89%, $n = 156$) were upregulated, while only 19 were consistently downregulated. Application of stringent data filtering to the entire dataset (Qlucore, $q = 0.01$ and >3 -fold \log_2 expression difference) identified 110 genes that were the most robust predictors of infection status that collapsed to 51 common sequence descriptions (Figure 4 and Table S2: tab Qlucore $q < 0.01$, $\log_2 > 3$ fold for complete gene list). Many shared coordinated patterns of gene expression across samples were identified by unsupervised hierarchical clustering (Figure 4). The most highly regulated immune relevant genes included serum amyloid A (GSONMT00016296001, GSONMT00005013001), complement c1q-like protein 2-like (GSONMT00002696001), differentially regulated trout protein 1 precursor (GSONMT00048193001, GSONMT00025517001), leukocyte cell-derived chemotaxin 2 precursor (GSONMT00024746001, GSONMT00065856001), toll-like receptor 5 membrane form (GSONMT00013855001), c-type lectin domain family 4 member e (GSONMT00005166001), c



type lectin receptor b (GSONMT00023806001), cd59b glycoprotein (GSONMT00025518001), and interleukine-1 receptor type 2-like (GSONMT00066304001). Interestingly, a number of putative metabolic genes were also highly expressed in infected ARS-Fp-S line fish compared to the ARS-Fp-R line including cis-aconitate decarboxylase-like (GSONMT00057407001), tbt-binding partial (GSONMT00003889001), l-serine dehydratase l-threonine deaminase-like (GSONMT00025010001), leptin (GSONMT00002603001), and growth differentiation factor 15 (GSONMT0000024001), catechol o-methyltransferase domain-containing protein 1-like (3 genes) and hepcidin (GSONMT00082379001). Multiple paralogues of several less abundantly expressed immune relevant genes included complement component 3 (4 genes), interferon-induced guanylate-binding protein 1 (3 genes), interferon-induced guanylate-binding protein 1-like (5 genes), interferon-induced protein 44-like (3 genes) and microfibril-associated glycoprotein 4-like (3 genes). Differentially regulated cytokine genes included interleukin 11 (GSONMT00009406001) and interleukin 1-beta (GSONMT00005489001) as well as nine chemokine genes that included: cc chemokine (GSONMT00017873001, GSONMT00007278001), c-c motif chemokine 19 precursor (GSONMT00014841001, GSONMT00057082001), c-c motif chemokine 4-like (GSONMT00080018001), chemokine ck-1 (GSONMT00024124001), cxc chemokine (GSONMT00080684001), and interleukin 8 (GSONMT00038968001, GSONMT00059090001). Immune genes of interest also included, programmed cell death 1 ligand-1 (GSONMT00040812001), tnfr receptor superfamily member 5 (GSONMT00012579001), tnfr receptor superfamily member 6b (GSONMT00034343001) and tnfr receptor

Table 2 | Summary of pairwise comparison between treatment groups including infection status, genetic line, and day ($q \leq 0.05$).

Comparisons	Day, Genetic line, and Infection status ^a	Diff. expressed genes
Infected vs. PBS	Day1 R-line (Fp) vs. R-line (PBS)	515
	Day5 R-line (Fp) vs. R-line (PBS)	428
	Day1 C-line (Fp) vs. C-line (PBS)	20
	Day5 C-line (Fp) vs. C-line (PBS)	2201
	Day1 S-line (Fp) vs. S-line (PBS)	1663
	Day5 S-line (Fp) vs. S-line (PBS)	2225
Genetic lines-PBS	Day1 R-line (PBS) vs. S-line (PBS)	76
	Day1 R-line (PBS) vs. C-line (PBS)	3
	Day1 S-line (PBS) vs. C-line (PBS)	28
	Day5 R-line (PBS) vs. S-line (PBS)	45
	Day5 R-line (PBS) vs. C-line (PBS)	246
	Day5 S-line (PBS) vs. C-line (PBS)	61
Genetic lines-Fp Inf.	Day1 R-line (Fp) vs. S-line (Fp)	150
	Day5 R-line (Fp) vs. S-line (Fp)	1016
	Day1 R-line (Fp) vs. C-line (Fp)	28
	Day5 R-line (Fp) vs. C-line (Fp)	159
	Day1 S-line (Fp) vs. C-line (Fp)	37
	Day5 S-line (Fp) vs. C-line (Fp)	1758
Between time points	Day5 vs. Day1 R-line (PBS)	1286
	Day5 vs. Day1 C-line (PBS)	294
	Day5 vs. Day1 S-line (PBS)	376
	Day5 vs. Day1 R-line (Fp)	334
	Day5 vs. Day1 C-line (Fp)	2469
	Day5 vs. Day1 S-line (Fp)	2434

^aAbbreviations: R, ARS-Fp-R; C, ARS-Fp-C; S, ARS-Fp-S; Fp, *F. psychrophilum* challenged; PBS, phosphate buffered saline injected.

superfamily member 9 (GSONMT00057532001), and b-cell receptor cd22-like isoform x2 (GSONMT00072668001).

GENE EXPRESSION IN NAIVE FISH

Gene expression profiles were compared in PBS-injected ARS-Fp-R, ARS-Fp-C and ARS-Fp-S line fish to examine the effect of selective breeding on the transcriptome of naive animals. In the first analysis, pair-wise comparisons between lines were examined. On day 1 post challenge, three genes demonstrated significantly different transcript abundance between PBS-injected ARS-Fp-R and ARS-Fp-C fish, 76 genes were different between the ARS-Fp-R and ARS-Fp-S line fish, and 28 genes were different between the ARS-Fp-C and ARS-Fp-S line fish (Table 2 and Supplementary Data Sheet 4: **tabs R_C PBS Day 1, R_S PBS Day 1, and S_C PBS Day 1, respectively**). None of these genes were common to all three comparisons and none were found to be similarly differentially regulated between genetic lines after challenge with *F. psychrophilum*. On day 5 post injection, the number of differentially regulated genes in naive fish increased with 246 genes demonstrating significant differences in transcript counts between the ARS-Fp-R and ARS-Fp-C lines, 45 genes between the

Table 3 | GO enrichment categories determined from pair-wise comparison.

Pair-wise comparison ^a	Total	Over-represented	Under-represented
GORich_Fp_PBS_D1_R	53	50	3
GORich_Fp_PBS_D1_C	11	11	0
GORich_Fp_PBS_D1_S	361	218	143
GORich_Fp_PBS_D5_R	326	156	170
GORich_Fp_PBS_D5_C	909	611	298
GORich_Fp_PBS_D5_S	607	333	274

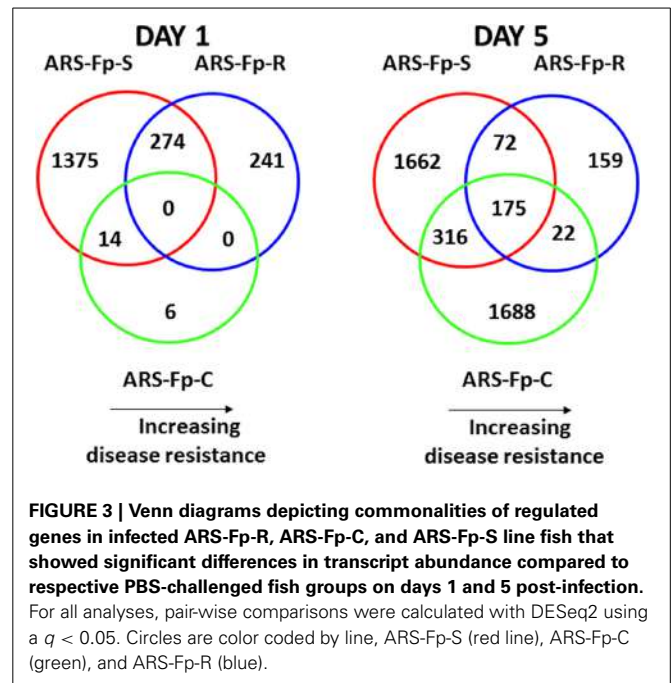
^aAbbreviations: Fp, *F. psychrophilum* challenged vs. PBS, phosphate buffered saline injected; D1, day1; D5, day5; R, ARS-Fp-R line; C, ARS-Fp-C line; S, ARS-Fp-S line.

ARS-Fp-R and ARS-Fp-S lines, and 61 genes between the ARS-Fp-C and ARS-Fp-S lines (Table 2). No genes were common to all three comparisons and eight genes exhibited a significant difference within two comparisons with similar trends in infected fish (Supplementary Data Sheet 4). Of these genes, only complement factor h-like (GSONMT00015052001) has a purported immunologic role. Fold gene expression differences were 2.8 and 2.1 between naive resistant and susceptible and infected resistant and susceptible fish, respectively. Complement factor h-like was 1.8 fold-higher in naive control fish compared to naive susceptible fish and 5.0 fold different between the infected lines (Supplementary Data Sheet 4).

A global analysis of expression differences between lines was performed on the entire dataset. In these analyses, data were first filtered for variance $\geq 2.5\%$ and then subjected to $q < 0.05$ and infection status was removed as a factor. Within the complete dataset, 21 differentially expressed genes were identified (Figure 5). Several genes exhibited the pattern of low expression in ARS-Fp-S line, intermediate expression in ARS-Fp-C line and highest expression in ARS-Fp-R line. These included immune relevant genes *tnf* receptor superfamily member 14b-like isoform x1 and interleukine-1 receptor-like 1, genes. We also analyzed only day 1 samples, and identified additional immune relevant candidates that included complement c1q-like protein 4 precursor, protein *nlr3*-like and *gata*-binding factor 2-like isoform x2 (Figure 6). Interestingly, there were also genes that exhibited the opposite expression pattern: higher in ARS-Fp-S line and lower in the ARS-Fp-C and ARS-Fp-R lines. This included putative immune relevant gene *nlr5*-partial (Figure 5) and *ig*-like v-type domain-containing protein family 187a-like (Figure 6).

MANUAL EXAMINATION OF KNOWN IMMUNE GENES EXPRESSED IN RESPONSE TO *F. PSYCHROPHILUM* INFECTION

In order to compare this dataset to previous analyses of gene expression following bacterial challenge, 23 known immune relevant genes were selected and *de novo* sequence alignment of the complete RNA-seq dataset was performed. There was a significant difference in transcript abundance between *F. psychrophilum* and PBS-challenged fish on day 1 and/or day 5 post-infection for 10/23 published gene sequence that were searched against the RNA-seq data set (Table 4) ($P_{adj} < 0.05$). Seven genes demonstrated a significant difference in transcript counts between



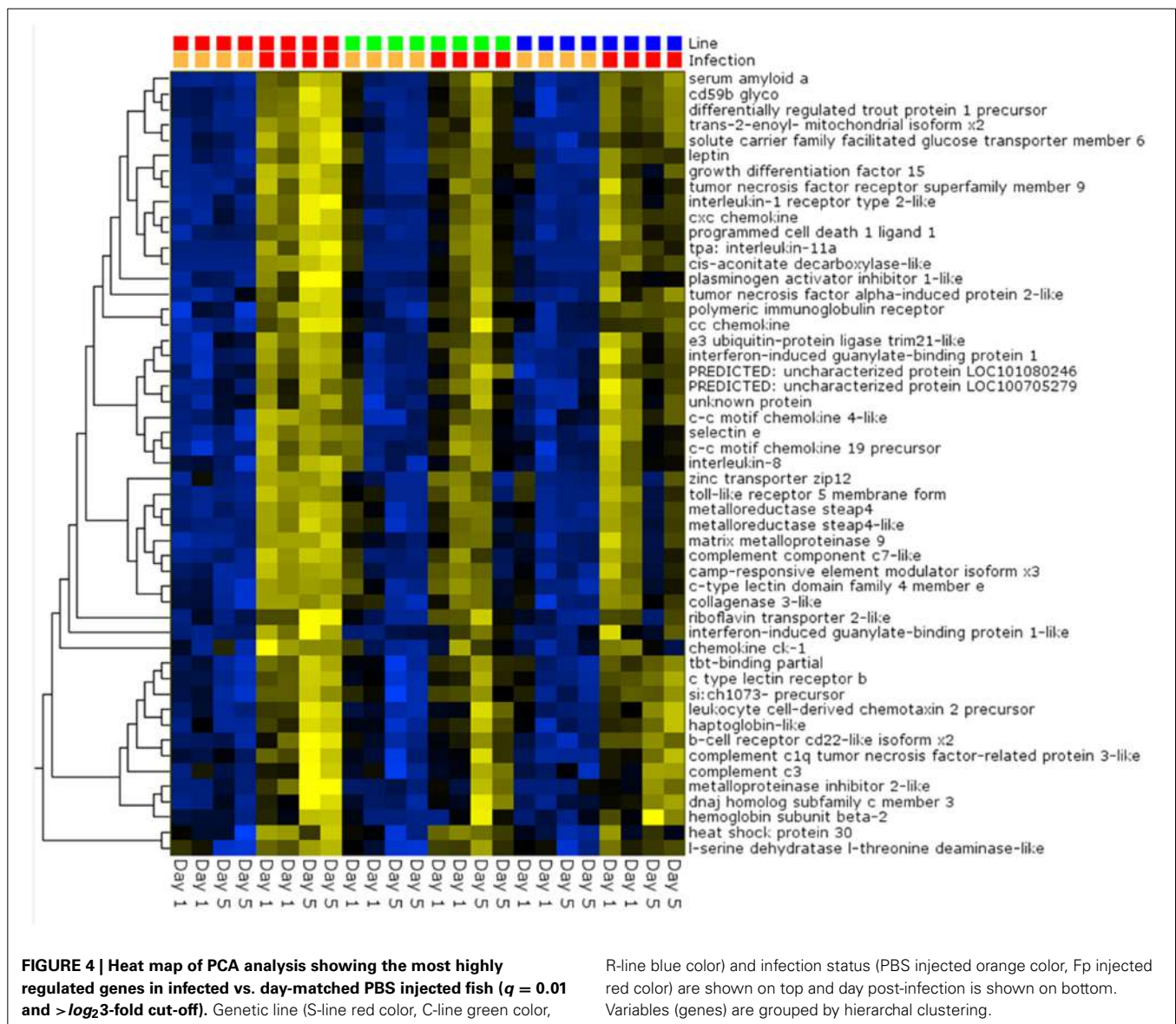
genetic lines on day 5 post-infection, with no significant differences occurring between genetic lines on day 1 (Table 4) ($P_{adj} < 0.05$). There was no significant difference in transcript abundance between infected and control fish or between genetic lines for *tgf- β* , *tcr- β* , *sod1*, *mx-1*, *mt-a*, *mhc-I*, *il-6*, *il-4/13A*, and *foxp3b*. Transcript counts were too low to in at least one comparison to provide a significant P_{adj} value for *tnf- α* , *il1- β* , *il-17c2*, and *inf- γ* .

VALIDATION OF RNA-seq DATA BY qPCR

In order to begin validating RNA-seq transcript abundance estimates, selected differentially expressed genes were identified for qPCR analysis (Table S1). There was significant correlation between transcript fold-change values determined by RNA-seq and qPCR in control and infected fish on day 5 post-infection ($R = 0.75$, $P < 0.001$) (Figure 7).

DISCUSSION

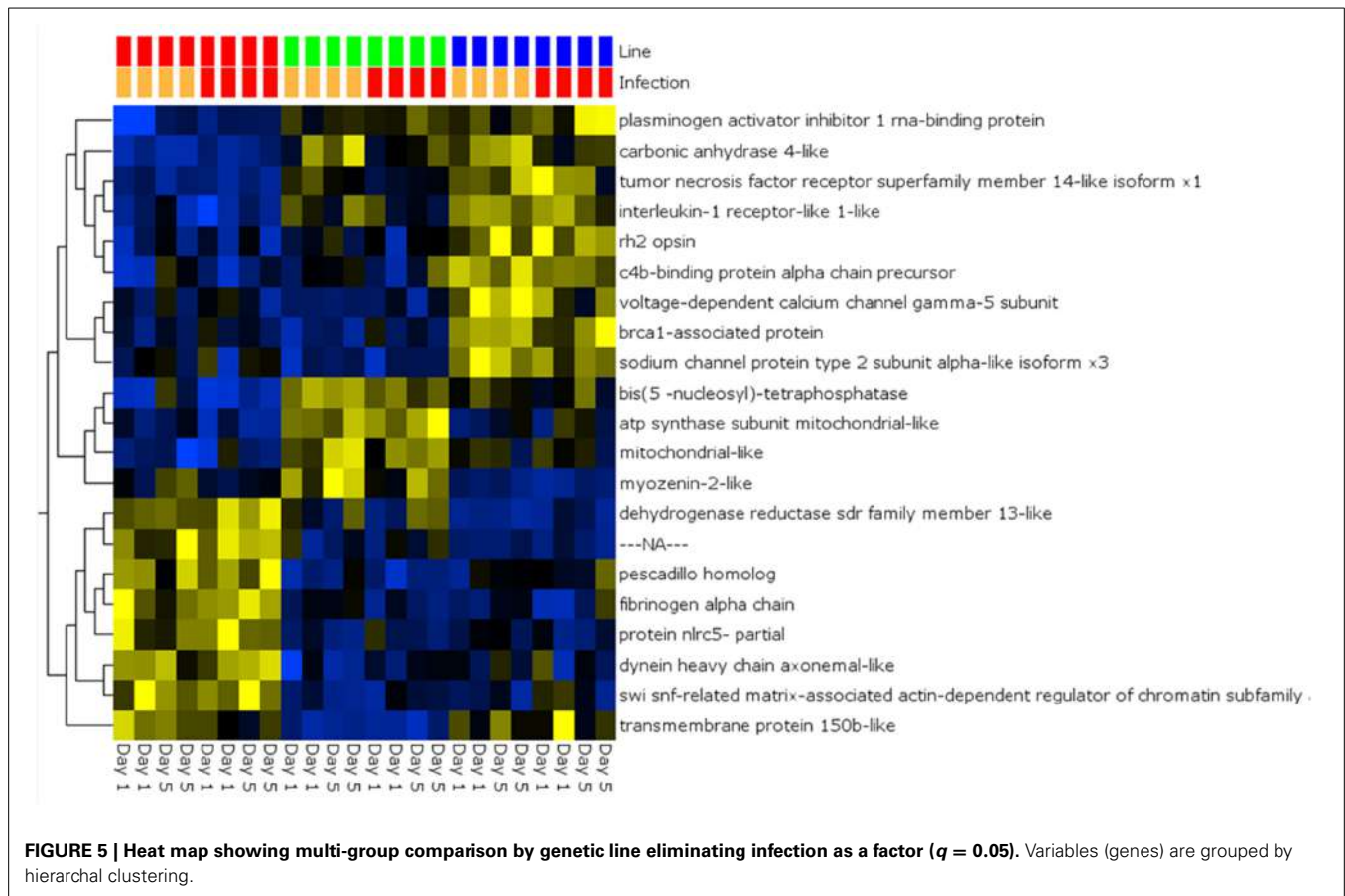
To our knowledge this is the first study describing whole-body transcriptome analysis and comparison of three divergently selected lines of rainbow trout exhibiting graded survival differences in response to standardized *F. psychrophilum* challenge. We identified large numbers of differentially expressed genes between genetic lines that increased in number with time and bacterial load. Since we utilized whole fish in these experiments, we interpreted differences in transcript abundance as upregulation/downregulation of genes or alteration of transcript stability. Importantly, the use of entire fish for RNA extraction rules out the potential for cell migration to confound expression differences inherent in studies utilizing defined immune tissue source (i.e., spleen, blood, or anterior kidney). A limitation of our approach is the likelihood of missing differentially expressed genes expressed only in rare subsets of cells. Nevertheless, our



depth of sequence coverage allowed quantification of transcript abundance from about 70.5% of all identified protein coding genes identified within the rainbow trout genome and approximately 68.2% of putative immune relevant genes that we identified by automated and manual annotation. Genes identified as differentially expressed were primarily associated with the acute phase response to bacterial infection but we also identified genes associated with innate and adaptive immune responses, physiologic and metabolic processes, and wound healing. This experimental design allowed for inclusion of both mucosal and systemic immune system tissue sampling and demonstrates previously unrecognized changes in gene transcription that occur with BCWD. These findings illustrate the high degree of transcriptional complexity involved in the rainbow trout BCWD response and provide a reference data set to begin to understand the impact of selective breeding on the genetic basis of disease resistance.

INDUCTION OF THE ACUTE PHASE RESPONSE AND INNATE IMMUNITY

The transcriptional response included induction of complement factors, acute-phase proteins, cytokines, chemokines and other genes associated with innate immunity. A relatively high number of complement factors were identified as upregulated after infection. This included multiple transcript alignments to genes encoding complement c3, complement c9, complement c4-b-like, and complement c1q-like proteins. Generation of complement during the acute-phase response has been well-described in rainbow trout and likely imparts direct bactericidal activity (Sunyer and Lambris, 1998; Whyte, 2007). The magnitude and range of complement factor expression suggests the complement system constitutes an important aspect of the host response to *F. psychrophilum* although factors were not found to be significantly different between resistant and susceptible fish. A wide range of acute phase protein encoding genes, including serum



amyloid A and *tlr5* were identified as highly expressed after infection. Similar trends have been described from tissues of rainbow trout affected by BCWD (Overturf and LaPatra, 2006; Langevin et al., 2012). Differentially expressed trout protein is a relatively recently recognized immune factor shown to be expressed during the salmonid acute phase response after bacterial infection (Bayne et al., 2001; Tsoi et al., 2004). Along with IL-1, IL-11, and IL-17-c1, these factors likely exert diverse pro-inflammatory and defensive actions including recruitment of inflammatory cells and further amplification of the acute phase response (Stadnyk, 1994; Jorgensen et al., 2001; Carrington et al., 2004; Wang et al., 2009). Significantly higher transcript counts for acute phase proteins and cytokines in the ARS-Fp-S line on day 5 may represent higher induction of pro-inflammatory conditions compared to the ARS-Fp-R line that correlates with higher bacterial loads.

DIFFERENTIAL EXPRESSION OF GENES THAT CONTRIBUTE TO ADAPTIVE IMMUNITY

A limited number of differentially regulated genes were identified that may be associated with adaptive immune processes. There was modest up-regulation of *igm* and *igt* genes and cellular factors associated with cell signaling and activation including *mhc-II*, *cd3*, *cd4*, and *cd8* genes. Higher *igm* gene transcript levels in resistant and control-line fish compared to the susceptible-line suggest an earlier development of antibody

mediated processes, although the converse pattern was observed for *mhc-II* gene transcript counts on the same days post-infection.

GENES INVOLVED IN WOUND HEALING

A number of genes associated with wound healing and wound response showed significant expression differences between naive and infected fish but no trends between genetic lines or sample day. These included syndecan-4-like isoform x2, plasminogen activator inhibitor 1-like, and ras-related c3 botulinum toxin substrate 2 precursor. All have roles in localized tissue repair in mammals through augmenting extracellular matrix reorganization, cellular growth and proliferation, and regulation of cell signaling (Romer et al., 1996; Woods and Couchman, 1998; Lin et al., 2005; Ojha et al., 2008). A limited number of genes with described roles in wound healing in other species (i.e., abhydrolase domain-containing protein 2-a-like and a collagen alpha-1 chain-like factor) demonstrated reduced expression in infected fish. Wound repair has not been well-characterized in fish and the general effect of these genes on the host response to infection is unknown. Necrosis of internal organs, peripheral skeletal muscle and skin is likely directly associated with host morbidity and mortality (Nilsen et al., 2011; Marancik et al., 2014b) and further studies are needed to determine how wound healing impacts recovery and survival.

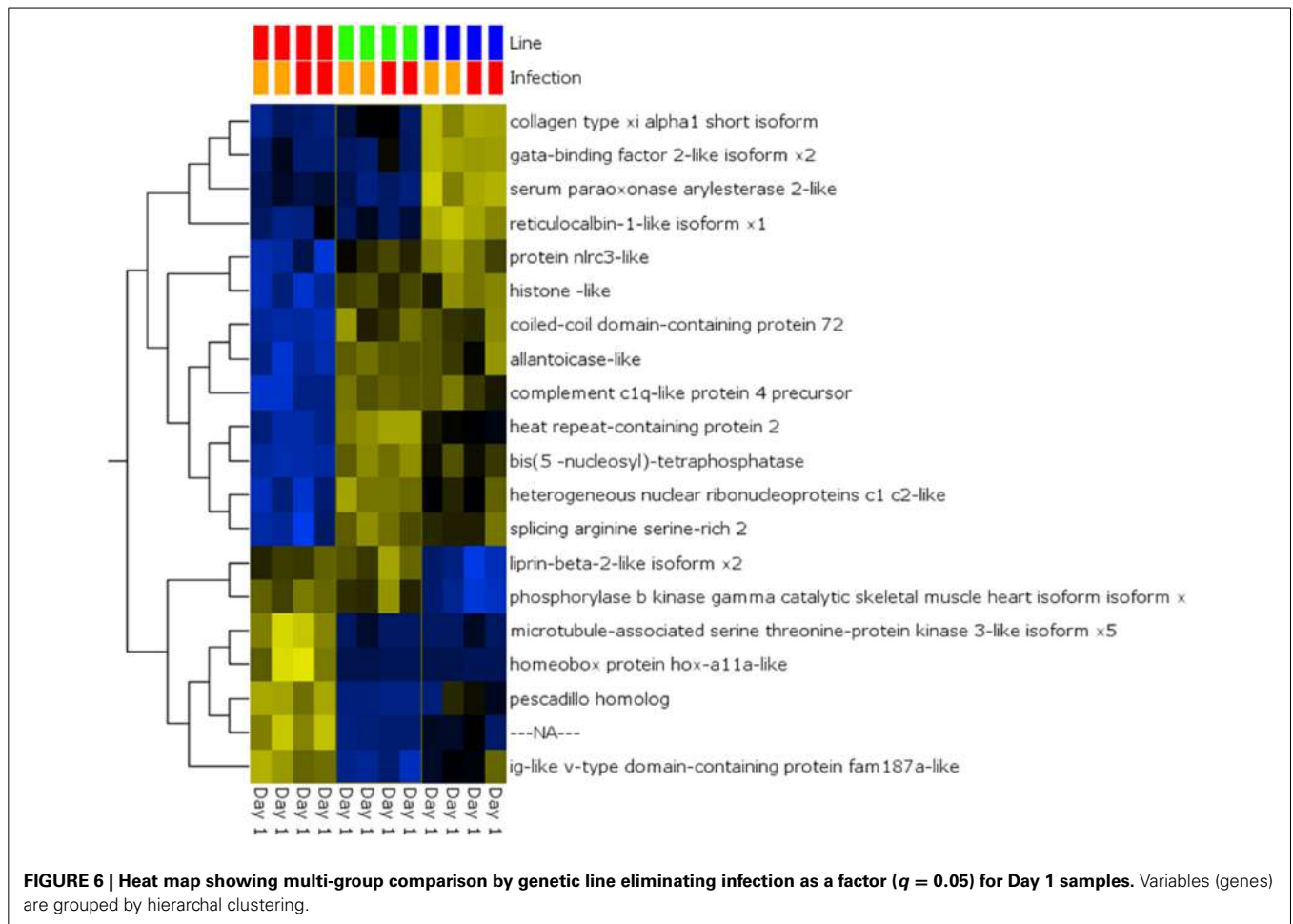


Table 4 | Fold-change in transcript abundance of genes described as modulated by *Fp* challenge.

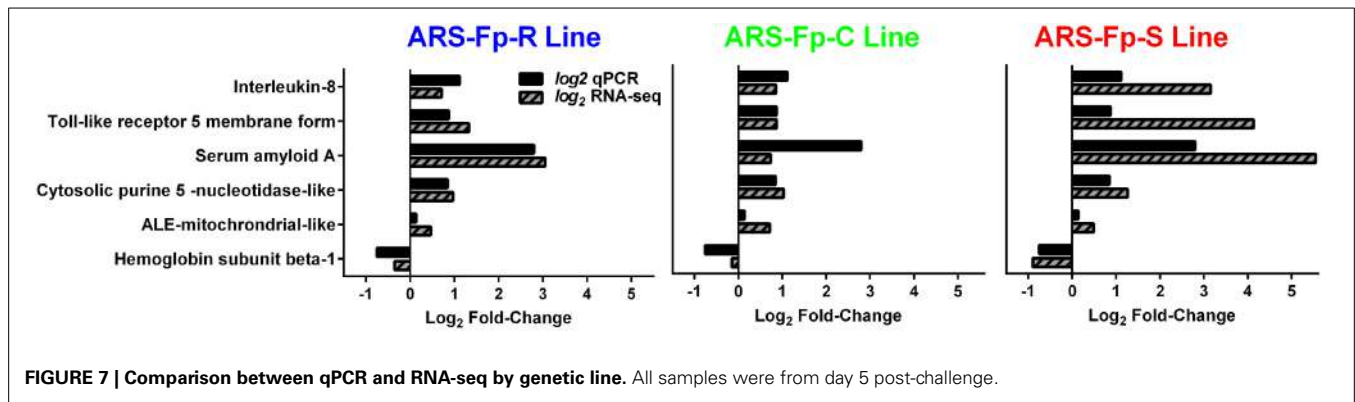
Annotation	Accession No	<i>Fp</i> vs. PBS, Day1			<i>Fp</i> vs. PBS, Day5			<i>Fp</i> vs. <i>Fp</i> , Day5		
		<i>R</i>	<i>C</i>	<i>S</i>	<i>R</i>	<i>C</i>	<i>S</i>	<i>R</i> vs. <i>S</i>	<i>R</i> vs. <i>C</i>	<i>C</i> vs. <i>S</i>
PROINFLAMMATORY/ACUTE PHASE RESP.										
<i>il-17-c1</i>	FM955453	3.9	–	172	8.8	3.6	41.8	–	–	–
<i>tlr5</i>	AB062504	–	–	8.0	2.6	2.1	18.4	0.4	–	–
<i>Saa</i>	AM422446	2.7	–	3.6	–	1.9	10.6	0.3	–	–
<i>il-8</i>	AJ279069	2.9	–	5.3	–	2.0	12.6	0.2	–	–
ADAPTIVE PROCESSES										
<i>cd3</i>	L24433	–	–	1.6	2.6	2.5	3.0	–	–	–
<i>cd4</i>	AY973028	–	–	–	1.9	2.4	1.7	–	–	–
<i>cd8</i>	AF178054	–	–	–	–	–	1.9	–	–	–
<i>lgm</i>	S63348	–	–	–	2.1	2.8	–	1.9	–	2
<i>lgt</i>	AY870265	–	–	–	2.1	2.8	–	1.9	–	–
<i>mhc-II</i>	AF115533	–	–	–	–	–	1.6	0.6	–	0.7

Quantification was performed between *Fp* and PBS challenged ARS-*Fp*-R (*R*), ARS-*Fp*-C (*C*), and ARS-*Fp*-S (*S*) line fish and between genetic lines on days 1 and 5 post-infection. There were no significant differences between genetic lines on day 1 post-infection.

IDENTIFICATION OF LARGE PARALOGOUS FAMILIES OF PUTATIVE IMMUNE-RELEVANT GENES

The recent rainbow trout genome contains expansion of putative gene families including protein *nlr3* and *nlr3*-like ($n = 111$),

protein *nlr5* ($n = 9$), polymeric immunoglobulin receptor-like ($n = 14$), perforin-1-like ($n = 8$), B-cell receptor *cd22*-like proteins ($n = 38$), *cd209*-like ($n = 34$), macrophage mannose receptor 1-like ($n = 28$), *Irr* and *pyd* domain-containing proteins



($n = 37$), and fish virus induced trim proteins ($n = 44$) (Supplementary Data Sheet 3). The availability of the genome allows a more comprehensive analysis of partially characterized immune gene families including tumor necrosis factor superfamily of ligands ($n = 26$) and toll-like receptors ($n = 27$) (Palti, 2011; Wiens and Glenney, 2011). Of note, a large number of tumor necrosis factor receptor superfamily members (tnfrsf, $n = 59$) were identified (Supplementary Data Sheet 3). While the precise phylogenetic nomenclature and grouping of tnfrsf remains, our analysis indicates many members are differentially regulated in response to *F. psychrophilum* infection including tnfrsf 1a precursor (GSONMT00019008001), tnfrsf 1a-like (GSONMT00061996001), tnfrsf 5 precursor (GSONMT00012579001, GSONMT00003531001), tnfrsf 5-like (GSONMT00013182001), tnfrsf 6 (GSONMT00082555001), tnfrsf 6b (GSONMT00034343001), tnfrsf 6b-like (GSONMT00020936001), tnfrsf 9 (GSONMT00057532001), tnfrsf 9-like (GSONMT00050654001), tnfrsf 19-like isoform x1 (GSONMT00055755001) and tnfrsf 19-like isoform x2 (GSONMT00069336001). In addition, basal expression of one paralogue of tnfrsf 14-like isoform x1 (GSONMT00000915001, from a total $n = 11$ paralogues present in the genome) is modestly higher in the ARS-Fp-R line as compared to the ARS-Fp-S line (Figure 5). While Blast2GO v.2 provides a description for novel sequences based on natural language text mining functionality (Gotz et al., 2008), we emphasize that detailed annotation and phylogenetic analysis of these genes remains to be undertaken. Some of these genes may be pseudogenes and it is likely that the sequence description of many of the automated annotations we present here will require revision following expert curation, and may also change as improvements are made to the reference rainbow trout genome sequence and analysis of transcript variants.

DIFFERENTIALLY EXPRESSED GENES BETWEEN NAIVE FISH FROM THE THREE GENETIC LINES

There was tight clustering of data and a low number of gene expression differences between PBS-injected fish. This suggests phenotypic differences between the ARS-Fp-R, -C, and -S lines are largely induced by infection and that selective breeding appears to have had a relatively low impact on basal gene expression during the normal physiologic state. Expression of

complement factor-h like was observed to be significantly different between both naive and infected resistant and susceptible fish. Complement factor-h is a regulatory protein of the alternative complement pathway and although isolated from rainbow trout (Anastasiou et al., 2011), its contribution to the rainbow trout immune response has not yet been characterized.

COMPARISON OF DIFFERENTIALLY REGULATED GENES WITH Langevin et al. STUDY

This study expands on the work of Langevin et al. (2012) who described differential regulation of select genes by microarray and qPCR in the anterior kidney of BCWD resistant and susceptible rainbow trout clones after experimental infection. Both studies showed an increase in transcription of genes encoding pro-inflammatory cytokines, anti-bacterial effectors and matrix metalloproteases. Notably, complement factors, serum amyloid A, mannose-binding proteins, chemokines, and interleukins 1 and 8 all showed significant transcriptional increase on day 5 post-infection. RNA-seq data provides further evidence for upregulation of immunorelevant factors not previously identified, including tlr5, leptin, haptoglobin, and C-type lectin and genes associated with adaptive, physiologic, structural, and intracellular process. There was no evidence to support differential regulation of interferon-gamma during infection in either study although potential interferon induced genes are differentially regulated. Tumor necrosis factor-alpha was also notably absent from both studies despite previously published (Evenhuis and Cleveland, 2012) and unpublished data (Wiens, unpublished data) suggesting upregulation during infection. Assay sensitivity may have been confounded by whole-body RNA isolation which could reduce the ability to detect low abundant cytokines expressed in specific tissues.

There was variability between studies when transcriptomic responses were compared between resistant and susceptible line fish. Both studies observed higher bacterial loads in susceptible fish and increased transcriptional response of a number of genes associated with pro-inflammatory conditions including interleukin-1, cc chemokine and matrix metalloproteinases 1 and 19. Our study further identified serum amyloid A, differentially regulated trout protein, and cytochrome p450 1a as significantly upregulated in susceptible fish. A number of metalloproteinase inhibitors were identified by Langevin et al. (2012) as significantly

upregulated in resistant but not susceptible fish but showed no significant differences between genetic lines when analyzed in our study. This variability likely extends from apparent differences in experimental design including challenge route, bacterial dose and strain, tissue sampled, and microarray vs transcript count quantification and analysis. However, even with these differences in method, both studies demonstrated a robust immune response in *F. psychrophilum* challenged fish with significant differences in the transcriptome of resistant and susceptible fish, associated with immune relevant genes.

CONCLUDING REMARKS

Complex transcriptional differences were identified between lines following infection with *F. psychrophilum* strain CSF259-93. Most of the differentially regulated genes exhibited increased transcript abundance and correlated with higher bacterial loads. It is likely that most of the changes we identified are consequences of the differential expansion of *F. psychrophilum* following infection, especially in the ARS-Fp-S line. However, differences in bacterial load cannot account for transcriptional differences observed on day 1, as bacterial loads were similar between genetic lines. Future efforts will be directed at dissecting early time points following exposure and will focus on identifying inter-individual differences as this study examined pools of fish due to sequencing cost limitations. This data will be integrated with proteomic studies to examine the relationship between transcript and protein levels and to assist in exploring biomarkers of infection, immunocompetency and disease resistance. Also, in this study we did not analyze long non-coding RNA, microRNA and splice variants. Despite these limitations, we suggest that this data set represents an important future resource for exploration of candidate genes identified from QTL analyses being conducted in parallel with this study. Quantitative trait loci (QTL) mapping has identified nine QTL on seven chromosomes that have moderate to large effects on resistance (Vallejo et al., 2014a). We have focused on an Omy19 QTL (Wiens et al., 2013b) and confirmed inheritance in a subsequent generation (Vallejo et al., 2014b). By combining chromosomal position known for many of the genes within this dataset, we have begun to identify differentially regulated genes present on Omy19 and other chromosomes as proof of principle for this approach. Further studies are needed to validate and fine-map the BCWD QTL and these studies are currently underway. These complex patterns support a polygenetic architecture of BCWD resistance and will serve as a reference dataset for identifying mechanisms associated with the genetic basis of disease resistance.

ACKNOWLEDGMENTS

We acknowledge technical contributions from Travis Moreland and Joel Caren. Author contributions: conceived and designed the experiments: Yniv Palti, Mohamed Salem, Jianbo Yao, Gregory D. Wiens; performed the experiments Gregory D. Wiens, Bam Paneru, Hao Ma, Mohamed Salem, Alvaro G. Hernandez; analyzed the data David Marancik, Guangtu Gao, Gregory D. Wiens; contributed reagents/materials/analysis tools GG; wrote the paper David Marancik, Gregory D. Wiens. All authors reviewed and approved the manuscript. This work was supported

by Agricultural Research Service CRIS Project 1930-32000-005 “Integrated Approaches for Improving Aquatic Animal Health in Cool and Cold Water Aquaculture,” CRIS Project 1930-32000-009 “Utilization of Genomics for Improving Production Traits in Cool and Cold Water Aquaculture” and Agriculture and Food Research Initiative competitive grant no. 2012-67015-30217 from the USDA National Institute of Food and Agriculture to Gregory D. Wiens. The authors have no conflict of interest to declare. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity employer.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00453/abstract>

REFERENCES

- Anastasiou, V., Mikrou, A., Papanastasiou, A. D., and Zarkadis, I. K. (2011). The molecular identification of factor H and factor I molecules in rainbow trout provides insights into complement C3 regulation. *Fish Shellfish Immunol.* 31, 491–499. doi: 10.1016/j.fsi.2011.06.008
- Barnes, M. E., and Brown, M. L. (2011). A review of *Flavobacterium psychrophilum* biology, clinical signs, and bacterial cold water disease prevention and treatment. *Open Fish Sci. J.* 4, 40–48. doi: 10.2174/1874401X01104010040
- Bayne, C. J., Gerwick, L., Fujiki, K., Nakao, M., and Yano, T. (2001). Immune-relevant (including acute phase) genes identified in the livers of rainbow trout, *Oncorhynchus mykiss*, by means of suppression subtractive hybridization. *Dev. Comp. Immunol.* 25, 205–217. doi: 10.1016/S0145-305X(00)00057-4
- Beck, B. H., Farmer, B. D., Straus, D. L., Li, C., and Peatman, E. (2012). Putative roles for a rhamnose binding lectin in *Flavobacterium columnare* pathogenesis in channel catfish *Ictalurus punctatus*. *Fish Shellfish Immunol.* 33, 1008–1015. doi: 10.1016/j.fsi.2012.08.018
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noel, B., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5, 3657. doi: 10.1038/ncomms4657
- Bluthgen, N., Brand, K., Cajavec, B., Swat, M., Herzel, H., and Beule, D. (2005a). Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform.* 16, 106–115. doi: 10.11234/gi1990.16.106
- Bluthgen, N., Kielbasa, S. M., and Herzel, H. (2005b). Inferring combinatorial regulation of transcription *in silico*. *Nucleic Acids Res.* 33, 272–279. doi: 10.1093/nar/gki167
- Branson, E. (1995). Rainbow trout fry syndrome. *Fish Vet. J.* 1, 1–7.
- Carrington, A. C., Collet, B., Holland, J. W., and Secombes, C. J. (2004). CpG oligodeoxynucleotides stimulate immune cell proliferation but not specific antibody production in rainbow trout (*Oncorhynchus mykiss*). *Vet. Immunol. Immunopathol.* 101, 211–222. doi: 10.1016/j.vetimm.2004.04.022
- Cock, J., Gitterle, T., Salazar, M., and Rye, M. (2009). Breeding for disease resistance of Penaeid shrimps. *Aquaculture* 286, 1–11. doi: 10.1016/j.aquaculture.2008.09.011
- Decostere, A., Lammens, M., and Haesebrouck, F. (2000). Difficulties in experimental infection studies with *Flavobacterium psychrophilum* in rainbow trout (*Oncorhynchus mykiss*) using immersion, oral and anal challenges. *Res. Vet. Sci.* 69, 165–169. doi: 10.1053/rvsc.2000.0408
- Evenhuis, J. P., and Cleveland, B. M. (2012). Modulation of rainbow trout (*Oncorhynchus mykiss*) intestinal immune gene expression following bacterial challenge. *Vet. Immunol. Immunopathol.* 146, 8–17. doi: 10.1016/j.vetimm.2012.01.008
- Gjedrem, T. (1983). Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture* 33, 51–71. doi: 10.1016/0044-8486(83)90386-1

- Gjedrem, T. (2005). *Selection and Breeding Programs in Aquaculture*. Dordrecht: Springer. doi: 10.1007/1-4020-3342-7
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350–353, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Gómez, E., Méndez, J., Cascales, D., and Guijarro, J. A. (2014). *Flavobacterium psychrophilum* vaccine development: a difficult task. *Microb. Biotech.* 7, 414–423. doi: 10.1111/1751-7915.12099
- Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176
- Hadidi, S., Glenney, G. W., Welch, T. J., Silverstein, J. T., and Wiens, G. D. (2008). Spleen size predicts resistance of rainbow trout to *Flavobacterium psychrophilum* challenge. *J. Immunol.* 180, 4156–4165. doi: 10.4049/jimmunol.180.6.4156
- Henriksen, M. M., Kania, P. W., Buchmann, K., and Dalsgaard, I. (2014). Effect of hydrogen peroxide and/or *Flavobacterium psychrophilum* on the gills of rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J. Fish Dis.* doi: 10.1111/jfd.12232. [Epub ahead of print].
- Jorgensen, J. B., Zou, J., Johansen, A., and Secombes, C. J. (2001). Immunostimulatory CpG oligodeoxynucleotides stimulate expression of IL-1beta and interferon-like cytokines in rainbow trout macrophages via a chloroquine-sensitive mechanism. *Fish Shellfish Immunol.* 11, 673–682. doi: 10.1006/fsim.2001.0344
- Kent, M. L., Groff, J. M., Morrison, J. K., Yasutake, W. T., and Holt, R. A. (1989). Spiral swimming behavior due to cranial and vertebral lesions associated with *Cytophaga psychrophila* infections in salmonid fishes. *Dis. Aquat. Org.* 6, 11–16. doi: 10.3354/dao006011
- Langevin, C., Blanco, M., Martin, S. A., Jouneau, L., Bernardet, J. F., Houel, A., et al. (2012). Transcriptional responses of resistant and susceptible fish clones to the bacterial pathogen *Flavobacterium psychrophilum*. *PLoS ONE* 7:e39126. doi: 10.1371/journal.pone.0039126
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25
- Leeds, T. D., Silverstein, J. T., Weber, G. M., Vallejo, R. L., Palti, Y., Rexroad, C. E. III, et al. (2010). Response to selection for bacterial cold water disease resistance in rainbow trout. *J. Anim. Sci.* 88, 1936–1946. doi: 10.2527/jas.2009-2538
- Lin, H. J., Shaffer, K. M., Sun, Z., Jay, G., He, W. W., and Ma, W. (2005). Glial-derived nexin, a differentially expressed gene during neuronal differentiation, transforms HEK cells into neuron-like cells. *Int. J. Dev. Neurosci.* 23, 9–14. doi: 10.1016/j.ijdevneu.2004.09.007
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *bioRxiv* 15:550. doi: 10.1101/002832
- Madsen, L., Arnbjerg, J., and Dalsgaard, I. (2001). Radiological examination of the spinal column is farmed rainbow trout *Oncorhynchus mykiss* (Walbaum): experiments with *Flavobacterium psychrophilum* and oxytetracycline. *Aquacult. Res.* 32, 235–241. doi: 10.1046/j.1365-2109.2001.00552.x
- Marancik, D. P., Camus, M. S., Camus, A. C., Leeds, T. D., Weber, G. M., and Wiens, G. D. (2014a). Biochemical reference intervals and pathophysiological changes in *Flavobacterium psychrophilum*-resistant and -susceptible rainbow trout lines. *Dis. Aquat. Org.* 111, 239–248. doi: 10.3354/dao02777
- Marancik, D. P., Leeds, T. D., and Wiens, G. D. (2014b). Histopathologic changes in disease-resistant-line and disease-susceptible-line juvenile rainbow trout experimentally infected with *Flavobacterium psychrophilum*. *J. Aquat. Anim. Health* 26, 181–189. doi: 10.1080/08997659.2014.920735
- Marancik, D. P., and Wiens, G. D. (2013). A real-time polymerase chain reaction assay for identification and quantification of *Flavobacterium psychrophilum* and application to disease resistance studies in selectively bred rainbow trout *Oncorhynchus mykiss*. *FEMS Microbiol. Lett.* 339, 122–129. doi: 10.1111/1574-6968.12061
- Martin, S. A., Blaney, S. C., Houlihan, D. F., and Secombes, C. J. (2006). Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*). *Mol. Immunol.* 43, 1900–1911. doi: 10.1016/j.molimm.2005.10.007
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nematollahi, A., Decostere, A., Pasmans, F., and Haesebrouck, F. (2003). *Flavobacterium psychrophilum* infections in salmonid fish. *J. Fish Dis.* 26, 563–574. doi: 10.1046/j.1365-2761.2003.00488.x
- Nilsen, H., Olsen, A. B., Vaagnes, O., Hellberg, H., Bottolfsen, K., Skjelstad, H., et al. (2011). Systemic *Flavobacterium psychrophilum* infection in rainbow trout, *Oncorhynchus mykiss* (Walbaum), farmed in fresh and brackish water in Norway. *J. Fish Dis.* 34, 403–408. doi: 10.1111/j.1365-2761.2011.01249.x
- Ojha, N., Roy, S., He, G., Biswas, S., Velayutham, M., Khanna, S., et al. (2008). Assessment of wound-site redox environment and the significance of Rac2 in cutaneous healing. *Free Radic. Biol. Med.* 44, 682–691. doi: 10.1016/j.freeradbiomed.2007.10.056
- Orieux, N., Douet, D. G., Le Henaff, M., and Bourdineaud, J. P. (2013). Prevalence of *Flavobacterium psychrophilum* bacterial cells in farmed rainbow trout: characterization of metallothionein A and interleukin1-beta genes as markers over-expressed in spleen and kidney of diseased fish. *Vet. Microbiol.* 162, 127–135. doi: 10.1016/j.vetmic.2012.08.015
- Overturf, K., and LaPatra, S. (2006). Quantitative expression (Walbaum) of immunological factors in rainbow trout, *Oncorhynchus mykiss* (Walbaum), after infection with either *Flavobacterium psychrophilum*, *Aeromonas salmonicida*, or infectious haematopoietic necrosis virus. *J. Fish Dis.* 29, 215–224. doi: 10.1111/j.1365-2761.2006.00707.x
- Palti, Y. (2011). Toll-like receptors in bony fish: from genomics to function. *Dev. Comp. Immunol.* 35, 1263–1272. doi: 10.1016/j.dci.2011.03.006
- Peatman, E., Li, C., Peterson, B. C., Straus, D. L., Farmer, B. D., and Beck, B. H. (2013). Basal polarization of the mucosal compartment in *Flavobacterium columnare* susceptible and resistant channel catfish (*Ictalurus punctatus*). *Mol. Immunol.* 56, 317–327. doi: 10.1016/j.molimm.2013.04.014
- Pereiro, P., Dios, S., Boltaña, S., Coll, J., Estepa, A., Mackenzie, S., et al. (2014). Transcriptome profiles associated to VHSV infection or DNA vaccination in turbot (*Scophthalmus maximus*). *PLoS ONE* 9:e104509. doi: 10.1371/journal.pone.0104509
- Romer, J., Bugge, T. H., Pyke, C., Lund, L. R., Flick, M. J., Degen, J. L., et al. (1996). Impaired wound healing in mice with a disrupted plasminogen gene. *Nat. Med.* 2, 287–292. doi: 10.1038/nm0396-287
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Shi, M., Huang, R., Du, F., Pei, Y., Liao, L., Zhu, Z., et al. (2014). RNA-seq profiles from grass carp tissues after reovirus (GCRV) infection based on singular and modular enrichment analyses. *Mol. Immunol.* 61, 44–53. doi: 10.1016/j.molimm.2014.05.004
- Silverstein, J. T., Vallejo, R. L., Palti, Y., Leeds, T. D., Rexroad, C. E. III., Welch, T. J., et al. (2009). Rainbow trout resistance to bacterial cold-water disease is moderately heritable and is not adversely correlated with growth. *J. Anim. Sci.* 87, 860–867. doi: 10.2527/jas.2008-1157
- Stadnyk, A. W. (1994). Cytokine production by epithelial cells. *FASEB J.* 8, 1041–1047.
- Sundell, K., Högfors-Rönholm, E., and Wiklund, T. (2014). “23 vaccination against diseases caused by flavobacteriaceae species,” in *Fish Vaccination*, eds A. L. Roar Gudding and O. Evensen (Chichester: Wiley Blackwell), 273–282. doi: 10.1002/9781118806913.ch23
- Sunyer, J. O., and Lambris, J. D. (1998). Evolution and diversity of the complement system of poikilothermic vertebrates. *Immunol. Rev.* 166, 39–57. doi: 10.1111/j.1600-065X.1998.tb01251.x
- Tsoi, S. C., Ewart, K. V., Penny, S., Melville, K., Liebscher, R. S., Brown, L. L., et al. (2004). Identification of immune-relevant genes from atlantic salmon using suppression subtractive hybridization. *Mar. Biotechnol.* 6, 199–214. doi: 10.1007/s10126-002-0101-2
- Vallejo, R. L., Palti, Y., Liu, S., Evenhuis, J. P., Gao, G., Rexroad, C. E. III., et al. (2014a). Detection of QTL in rainbow trout affecting survival when challenged with *Flavobacterium psychrophilum*. *Mar. Biotechnol.* 16, 349–360. doi: 10.1007/s10126-013-9553-9
- Vallejo, R. L., Palti, Y., Liu, S., Marancik, D. P., and Wiens, G. D. (2014b). Validation of linked QTL for bacterial cold water disease resistance and spleen size on rainbow trout chromosome Omy19. *Aquaculture* 432, 139–143. doi: 10.1016/j.aquaculture.2014.05.003
- Van Muiswinkel, W. B., Wiegertjes, G. F., and Stet, R. J. M. (1999). The influence of environmental and genetic factors on the disease resistance of fish. *Aquaculture* 172, 103–110. doi: 10.1016/S0044-8486(98)00444-X

- Villarroel, F., Casado, A., Vasquez, J., Matamala, E., Araneda, B., Amthauer, R., et al. (2008). Serum amyloid A: a typical acute-phase reactant in rainbow trout? *Dev. Comp. Immunol.* 32, 1160–1169. doi: 10.1016/j.dci.2008.03.004
- Villarroel, F., Zambrano, A., Amthauer, R., and Concha, M. I. (2009). Detection of up-regulated serum amyloid A transcript and of amyloid AA aggregates in skeletal muscle lesions of rainbow trout infected with *Flavobacterium psychrophilum*. *Vet. Immunol. Immunopathol.* 130, 120–124. doi: 10.1016/j.vetimm.2009.02.002
- Wang, T., Martin, S. A., and Secombes, C. J. (2009). Two interleukin-17C-like genes exist in rainbow trout *Oncorhynchus mykiss* that are differentially expressed and modulated. *Dev. Comp. Immunol.* 34, 491–500. doi: 10.1016/j.dci.2009.11.011
- Whyte, S. K. (2007). The innate immune response of finfish—a review of current knowledge. *Fish Shellfish Immunol.* 23, 1127–1151. doi: 10.1016/j.fsi.2007.06.005
- Wiens, G. D., and Glenney, G. W. (2011). Origin and evolution of TNF and TNF receptor superfamilies. *Dev. Comp. Immunol.* 35, 1324–1335. doi: 10.1016/j.dci.2011.03.031
- Wiens, G. D., LaPatra, S. E., Welch, T. J., Evenhuis, J. P., Rexroad, C. E. III., and Leeds, T. D. (2013a). On-farm performance of rainbow trout (*Oncorhynchus mykiss*) selectively bred for resistance to bacterial cold water disease: effect of rearing environment on survival phenotype. *Aquaculture* 388–391, 128–136. doi: 10.1016/j.aquaculture.2013.01.018
- Wiens, G. D., LaPatra, S. E., Welch, T. J., Rexroad, C. E. III., Call, D. R., Cain, K. D., et al. (2014). Complete genome sequence of *Flavobacterium psychrophilum* strain CSF259-93 used select rainbow trout for increased genetic resistance against bacterial cold water disease. *Genome Announcements* 2, e00889–14. doi: 10.1128/genomeA.00889-14
- Wiens, G. D., Vallejo, R. L., Leeds, T. D., Palti, Y., Hadidi, S., Liu, S., et al. (2013b). Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout: identification of QTL on chromosome Omy19. *PLoS ONE* 8:e75749. doi: 10.1371/journal.pone.0075749
- Woods, A., and Couchman, J. R. (1998). Syndecans: synergistic activators of cell adhesion. *Trends Cell Biol.* 8, 189–192. doi: 10.1016/S0962-8924(98)01244-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 October 2014; paper pending published: 24 November 2014; accepted: 11 December 2014; published online: 08 January 2015.

Citation: Marancik D, Gao G, Paneru B, Ma H, Hernandez AG, Salem M, Yao J, Palti Y and Wiens GD (2015) Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum*. *Front. Genet.* 5:453. doi: 10.3389/fgene.2014.00453
This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Marancik, Gao, Paneru, Ma, Hernandez, Salem, Yao, Palti and Wiens. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress

Jesús Fernández^{1*}, Miguel Á. Toro², Anna K. Sonesson³ and Beatriz Villanueva¹

¹ Departamento de Mejora Genética Animal, INIA, Madrid, Spain

² Departamento de Producción Animal, ETSI Agrónomos, Universidad Politécnica de Madrid, Madrid, Spain

³ Nofima, Ås, Norway

Edited by:

Ross Houston, University of Edinburgh, UK

Reviewed by:

Jeff Silverstein, United States Department of Agriculture, Agricultural Research Service, USA
Ross Houston, University of Edinburgh, UK

*Correspondence:

Jesús Fernández, Departamento de Mejora Genética Animal, INIA, Ctra. Coruña Km 7,5, 28040 Madrid, Spain
e-mail: jmj@inia.es

The success of an aquaculture breeding program critically depends on the way in which the base population of breeders is constructed since all the genetic variability for the traits included originally in the breeding goal as well as those to be included in the future is contained in the initial founders. Traditionally, base populations were created from a number of wild strains by sampling equal numbers from each strain. However, for some aquaculture species improved strains are already available and, therefore, mean phenotypic values for economically important traits can be used as a criterion to optimize the sampling when creating base populations. Also, the increasing availability of genome-wide genotype information in aquaculture species could help to refine the estimation of relationships within and between candidate strains and, thus, to optimize the percentage of individuals to be sampled from each strain. This study explores the advantages of using phenotypic and genome-wide information when constructing base populations for aquaculture breeding programs in terms of initial and subsequent trait performance and genetic diversity level. Results show that a compromise solution between diversity and performance can be found when creating base populations. Up to 6% higher levels of phenotypic performance can be achieved at the same level of global diversity in the base population by optimizing the selection of breeders instead of sampling equal numbers from each strain. The higher performance observed in the base population persisted during 10 generations of phenotypic selection applied in the subsequent breeding program.

Keywords: base populations, aquaculture breeding programs, SNP markers, optimal contributions, fish genomics

INTRODUCTION

The success of an aquaculture breeding program critically depends on the way in which the base population of breeders is constructed (Hayes et al., 2006; Holtsmark et al., 2006, 2008a,b). In particular, all the genetic variability that can be used for selection on the traits initially included in the breeding objective is that found in the original breeders. Also, the decisions taken when creating the base population will have consequences on the genetic progress for any other additional trait that may be part of future breeding goals, whatever production or fitness related traits.

Traditionally, base populations in aquaculture breeding programs were created from wild or from domesticated strains not subjected to any formal selection program (see Gjedrem et al., 1991, for details on the formation of the base population in the Norwegian breeding program for Atlantic salmon and Eknath et al., 1993, 1998, for details on the formation of the base population in the GIFT breeding program for tilapia). In this context, Holtsmark et al. (2006, 2008a) investigated through computer simulation the effects of the number of strains contributing to the base population and the mating strategy (within and across strains) on the genetic gain achieved in the subsequent selection

program. These studies assumed wild populations and no knowledge of the genetic structure (i.e., within and between strain diversity) or phenotypic levels for the trait of interest when setting up the base population. Consequently, the simulated strategy was randomly sampling the same number of individuals from each strain to form the base population. However, the solution leading to the highest level of diversity and acceptable levels of phenotypic performance when starting the breeding program surely will depart from equal proportions.

In the absence of genealogies (which could be the case even in improved commercial strains), molecular markers can be used to estimate relationships between and within populations. Hayes et al. (2006) compared random with molecular-based optimized selection of breeders in terms of the genetic variance captured for growth and for two disease traits in Atlantic salmon. However, no considerations about the phenotypic level of candidates were made when sampling the breeders and the number of markers they used was scarce (237 AFLP).

The present availability of large panels of SNPs makes marker diversity more informative on the global genetic variability in the genome than diversity computed from genealogical data

(Gómez-Romano et al., 2013). Nowadays, dense SNP chips have been already developed for Atlantic salmon (130K, Houston et al., 2014), rainbow trout (57K, Palti et al., 2014), catfish (250K, Liu et al., 2014) and carp (250K, Xu et al., 2014). This development of an increasing number of markers for aquaculture species will make it possible in a near future to have accurate measures of genetic relationships between and within strains. In this new scenario, the optimal number of individuals to be sampled from each strain could be calculated in a similar way as when optimizing the construction of mixed populations from different origins in conservation programs aimed at capturing the highest levels of genetic diversity (Eding and Meuwissen, 2001; Caballero and Toro, 2002; Eding et al., 2002).

Nowadays, for some aquaculture species already genetically improved strains are available. The use of improved rather than wild strains when creating base populations would allow the new breeding program to begin from higher phenotypic levels for the trait of interest, making it competitive from the start. The necessity of taking into account the phenotypic value of each candidate strain is even clearer when searching for new breeders to be included in an already established breeding program whose genetic variability has been greatly reduced. The increase in genetic variability should be achieved without compromising the gain in performance previously obtained through artificial selection.

The objective of this paper was to study, using computer simulations, the consequences of using genome-wide molecular information to compute genetic relationships and phenotypic records to optimize the sampling of individuals from different strains when creating base populations in aquaculture breeding programs. Different scenarios varying in the type of strains (wild or commercial), the degree of relationships within and between strains and the level of information (individual or strain information) were considered for generic aquaculture population designs. Results were compared in terms of phenotypic level for the trait of interest and the diversity achieved in the base population and in subsequent generations of selection.

MATERIALS AND METHODS

GENOME STRUCTURE

Diploid individuals were simulated with a genome comprising 20 chromosomes of 1 Morgan each. This genome length could be representative of the genomic architecture of the most cultivated fish species in aquaculture (which have genomes ranging from 10 to 35 Morgans). Each chromosome carried 25,000 neutral biallelic loci that will be referred to as “non-marker loci” thereafter. Additionally, 5–5000 biallelic markers per chromosome (equivalent to SNPs) were simulated interspersed with the neutral loci. Therefore, the total number of available markers ranged from 100 to 100,000 and the density of markers ranged from 5 to 5000 markers/M. All loci were evenly spaced within each chromosome. Markers were used to optimize the contributions of the different strains to the base population while non-marker loci (i.e., the 25,000 neutral loci per chromosome) were used to monitor the effect on genetic diversity of the different strategies evaluated (see below). With dense marker panels we expect that most non-marker loci in the genome will be in linkage disequilibrium (LD)

with at least some of the markers and, thus, managing diversity using marker genotypes will lead to the maintenance of diversity in the rest of the genome.

GENERATION OF CANDIDATES

The creation of the candidates to contribute to the base population followed a two-step process. First, a large population in mutation-drift equilibrium was generated. Then, individuals were sampled from this population to form different strains which were allowed to diverge for several generations before being available for the selection of the individuals to be included in the base population.

Equilibrium population

First, to obtain a realistic pattern of linkage disequilibrium, a large population ($N = 1000$ with equal number of males and females) under random selection was simulated for 1000 discrete generations. Each generation, sires and dams were sampled with replacement and population size was kept constant across generations. Initially, genotypes for the individuals were assigned at random independently for each locus (markers and non-markers). Consequently, the initial allelic frequencies were 0.5 for all loci and Hardy-Weinberg and linkage equilibria existed within and between loci, respectively. During this period, mutation was allowed to occur throughout the genome. The mutation rate per locus and generation was $\mu = 2.5 \times 10^{-3}$ for both types of loci (marker and non-marker). The number of new mutations simulated at every generation was sampled from a Poisson distribution with mean $2Nn_c\mu n_l$ where n_c is the number of chromosomes and n_l is the total number of loci (markers and non-markers) per chromosome. Mutations were then randomly distributed across individuals, chromosomes and loci, switching allele 0 to allele 1 and vice versa. When generating the gametes, the number of crossovers per chromosome was drawn from a Poisson distribution with mean equal to 1. Crossovers were randomly distributed without interference. At the end of the process the expected heterozygosity of the population had already reached an equilibrium value.

Creation of strains

The second step of the process consisted in randomly sampling 10 different groups of individuals (mimicking 10 different strains) from the population at equilibrium. At this step, a quantitative trait with phenotypic mean (μ), initial phenotypic variance [$V_{P(0)}$] and heritability [$h_{(0)}^2$] of 100, 30, and 0.4, respectively, was defined. The trait, measured in both sexes, was controlled by 1000 additive loci (thereafter called selective loci). These selective loci were chosen at random from the previously simulated loci (markers and non-markers). The additive effect of locus i (a_i) was sampled from a normal distribution with mean 0 and variance $V_A(0)/[2p(1-p)n_{sel}]$, where $V_A(0)$ is the initial additive variance ($h_{(0)}^2 V_P(0) = 12$), p is the average frequency across selective loci and n_{sel} is the number of selective loci (i.e., 1000). Note that, in this way, the expected additive variance summed over all loci equals $V_A(0)$, assuming that covariances between loci generated by LD are negligible. The phenotypic value for a particular individual j was obtained as $P_j = \mu + \sum_{i=1}^{n_{sel}} x_i a_i + e_j$,

where x_i is an indicator variable that takes values 1, 0 or -1 for homozygous 11, heterozygous or homozygous 00, respectively, and e_j is the individual environmental deviation that was sampled from a normal distribution with mean 0 and variance $V_E = V_A(0)(1-h_{(0)}^2)/h_{(0)}^2$. The environmental variance (V_E) was initially calculated for each replicate and strain in order to assure that all started with the same $h_{(0)}^2$ value and was kept constant across generations.

Once the strains were created, they were allowed to diverge during 20 discrete generations (with constant population size) under three different regimes: (i) random selection; (ii) artificial selection with different selection pressures in order to mimic already improved strains; and (iii) stabilizing selection with different optima in order to mimic wild strains with local adaptations in nature. Four different scenarios varying in the type of strains they included were then defined:

- (1) Drift Scenario. All strains were created under random selection but they differed in size (from 10 to 40 individuals, half of each sex) in order to simulate different degrees of genetic drift. Smaller sizes will lead to lower within strain diversity but higher differentiation between strains.
- (2) Selection Scenario. All strains were created under artificial directional selection for the simulated trait. They had all the same size (20 males and 20 females) but differed in the strength of the selection imposed. The highest selection pressure corresponded to selecting the two males and the two females with the highest phenotypic values for the trait, and the lowest pressure corresponded to selecting five males and five females.
- (3) Stabilizing Scenario. All strains were created under stabilizing selection for the simulated trait. The probability of survival of an individual with phenotypic value P was modeled using a Gaussian distribution (Turelli, 1984; Bürger et al., 1989):

$$W(P) = \exp\left(-\frac{(P - P_{opt})^2}{2\omega^2}\right)$$

where P_{opt} is the optimum phenotype in a particular environment and ω^2 is an inverse measure of the strength of stabilizing selection. For some strains P_{opt} was set to the original mean (100), for others the optimum was set to a lower value (90) and for others the optimum was set to a higher value (110). Selection pressure was relatively strong ($\omega^2 = 5$) and population size was equal in all strains (20 males and 20 females).

- (4) Mixed Scenario. The set comprised strains of the three types described above (i.e., randomly selected, under directional selection and under stabilizing selection).

Table 1 summarizes the specific parameters used when creating the 10 strains under each scenario. The sizes of the strains simulated during the 20 generations were rather small in order to force a rapid divergence between them and, thus, real differences in relatedness and phenotypic levels at the end of the differentiation stage. After this period, subpopulations were expanded in order to have a large number of candidates to form the base population. Specifically, the population size of each strain increased to 50 males and 50 females in a single generation of random selection and mating.

FOUNDATION OF THE BASE POPULATION

From the 1000 available candidates (50 males and 50 females from each strain) for each particular scenario, base populations were constructed by selecting 100 males and 100 females following different strategies: (i) Taking at random equal numbers of individuals from each strain (strategy E); (ii) Determining optimal strain proportions for maximizing the expected heterozygosity (H_e) calculated from the mean coancestry values within and between strains (strategy MC); (iii) Determining optimal strain proportions for maximizing the mean (strain) phenotypic value with a restriction on coancestry (strategy MP); (iv) as in (ii) but using individual relationships instead of strain means (strategy IC); and (v) as in (iii) but maximizing individual values of the selected individuals instead of strain means (strategy IP). Note that, in these abbreviations, M and I stand for mean and individual information, respectively, C indicates that the objective is

Table 1 | Parameters used to generate each strain for the four different scenarios simulated.

Strains	Drift	Selection		Stabilizing		Mixed		
	Size	Size	Numbers selected	Size	Optimum phenotype	Size	Optimum phenotype	Numbers selected
1	5	20	2	10	90	5	–	–
2	5	20	2	10	90	10	–	–
3	5	20	3	10	90	10	–	–
4	10	20	3	10	100	20	–	–
5	10	20	4	10	100	10	90	–
6	10	20	4	10	100	10	100	–
7	20	20	4	10	100	10	110	–
8	20	20	5	10	110	10	–	5
9	20	20	5	10	110	10	–	3
10	20	20	5	10	110	10	–	1

Size and Numbers selected refer to the number of individuals per sex.

just to minimize coancestry (i.e., maintain high levels of H_e) and P indicates that the objective is to create base populations with a high phenotypic level.

Strategy E is equivalent to that used by Holtmark et al. (2006, 2008a,b) and provides a reference point for comparisons. Strategy MC followed the methodology presented by Eding and Meuwissen (2001) and Caballero and Toro (2002). The particular objective function to minimize was $\sum_{i=1}^{N_s} c_i \sum_{j=1}^{N_s} c_j f_{ij}$, where N_s is the number of strains, c_i is the proportion of individuals to be sampled from strain i and f_{ij} is the mean coancestry coefficient between strains i and j calculated from marker genotypes. Contributions were forced to be in the interval $[0,1]$ and to sum up to 1.

Strategy MP searched for the solution that maximized the objective function $\sum_{i=1}^{N_s} c_i P_i$, where P_i is the mean phenotype of strain i but imposing the restriction $\sum_{i=1}^{N_s} c_i \sum_{j=1}^{N_s} c_j f_{ij} \leq C_E$, where C_E is the mean coancestry of the base population obtained under strategy E. The restrictions imposed in MC were also applied to MP. In the two strategies relying on strain mean values (MC and MP) the actual number of individuals to be sampled from each strain was obtained by multiplying c_i by the total number of individuals to be selected and rounding to the nearest even integer. This procedure was implemented to ensure that half of the individuals from each strain were males and half females.

The objective under strategies IC and IP was to minimize $\sum_{i=1}^N \sum_{j=1}^N x_i x_j f_{ij}$ and to maximize $\sum_{i=1}^N x_i P_i$, respectively. Here N is the total number of candidates (i.e., all individuals from every strain), f_{ij} is the coancestry between individuals i and j , P_i is the phenotype of individual i and x_i is an indicator variable that takes a value of 1 if individual i is to be selected and 0 otherwise. The sum of x 's for males and females was forced to be equal to the number of individuals to be selected for creating the base population (i.e., 100 of each sex). Strategy IP also included a restriction to guarantee that solutions had a global coancestry lower or equal than that obtained under E strategy. All the optimizations were performed using "simulated annealing" algorithms (Kirkpatrick et al., 1983).

It must be emphasized that in strategies MC and MP the output of the optimizations is the proportion of individuals to be taken from each strain but the specific individuals are sampled at random. This could be the situation when prior knowledge of the mean genetic relationship or performance of the strains is available but individual information for the candidates is not. In strategies IC and IP the outputs are the particular individuals to be selected as we assume that their genotypes and phenotypic values are known.

As described above, restrictions imposed under strategies MP and IP were set to the level of coancestry (calculated on the markers) obtained under strategy E. Thus, these three strategies could be compared in terms of phenotypic level at the same diversity level.

ARTIFICIAL SELECTION

To explore the consequences of each strategy for creating the base population on genetic gain and diversity in the subsequent breeding programme, 10 generations of artificial selection were simulated for each combination of marker density, type of

available strains and strategy. At generation $t = 0$ (i.e., base population), founders were mated at random to form 100 families and 10 offspring were obtained from each couple. Consequently, 1000 individuals were available as candidates for selection. The 100 males and 100 females with the highest phenotypic value for the simulated trait were selected to produce generation $t+1$ (i.e., phenotypic truncation selection was conducted). Selected individuals were mated at random and, again, 10 offspring were generated from each couple. Therefore, the proportion of selected individuals was 20% that corresponds to a selection intensity of 1.4.

VARIABLES FOR COMPARISON

In the base population, comparisons between strategies were made in terms of the mean phenotypic value (\bar{P}) and H_e of the group of selected individuals. Note that H_e , calculated on the non-marker loci, is a measure of the genetic variation of the population and its ability to adapt to new environments. The contributions of strains to the base population, measured as the proportion of breeders selected from each of the strains, was also considered in the comparisons.

In the artificial selection step, the strategies were also compared in terms of mean breeding value (BV) and additive variance (V_A) for the target trait, genealogical inbreeding (F) and coancestry (f) coefficients and rates of gain, inbreeding (ΔF) and coancestry (Δf). For the computation of inbreeding and coancestry, founders in the base population were assumed to be unrelated and non-inbred. In all scenarios, values presented are averages of 100 replicates.

RESULTS

CONTRIBUTIONS OF STRAINS TO THE BASE POPULATION

The proportional contribution of each available strain to the base population is shown in **Table 2** for the most extreme marker densities, the four scenarios and the five strategies simulated. In general, the observed patterns were the same when using high or low density of markers. Particular differences in performance due to marker density are highlighted below.

In the Drift scenario no meaningful differences in contributions were observed between strategies MC, MP, and IC for a particular strain. This was due to the fact that no selection on the quantitative trait was exerted during the generation of the strains and, therefore, the phenotypic mean was equal to the initial value before the divergence period (i.e., 100) for all strains. However, differences arose between strains. The higher the population size of a particular strain the higher was its contribution. Strains with a small size (strains 1, 2, and 3) had the lowest contributions due to the large loss of genetic diversity during the divergence period. Therefore, these strains were less useful for increasing the amount of diversity stored in the synthetic base population. The opposite happened with large size strains (7, 8, 9, and 10) that contributed more than 14% each to the base population (see Drift scenario in **Table 2**). The lowest variance of contributions between strains (beyond strategy E) was found under strategy IP. This could be explained by the fact that strong drift in small populations could result in the existence of individuals with extreme high phenotype. Thus, as the main objective of strategy IP is

Table 2 | Contributions of each strain to the base population (in percentage) under different strategies to select individuals for the base population for the four scenarios considered and for different number of markers.

Markers		Drift					Selection				Stabilizing				Mixed			
		E	MC	MP	IC	IP	MC	MP	IC	IP	MC	MP	IC	IP	MC	MP	IC	IP
100	1	10.0	-3.3	-3.6	-3.6	-2.9	-1.5	-3.8	-1.6	-4.3	0.6	-4.1	-0.1	-7.6	-0.9	-2.9	-1.2	-4.1
	2	10.0	-5.0	-4.0	-4.4	-1.8	-1.8	-4.5	-1.7	-5.0	-0.7	-5.1	-0.6	-7.6	4.0	1.0	3.3	-1.5
	3	10.0	-5.1	-4.4	-4.5	-2.1	0.1	-0.2	-0.3	-0.5	-1.1	-5.4	-0.7	-7.8	2.1	-1.2	2.5	-2.5
	4	10.0	-0.5	0.6	-0.6	1.0	-0.9	-1.5	-0.7	-1.2	1.5	1.7	0.8	0.4	12.4	5.4	9.8	0.1
	5	10.0	-2.6	-1.4	-1.6	0.5	0.5	1.6	0.1	1.3	0.7	0.7	0.9	0.2	-2.3	-7.7	-1.7	-8.9
	6	10.0	-2.2	-3.1	-1.5	-1.9	-0.4	0.1	0.0	0.8	0.3	-0.1	0.6	-0.6	-1.9	-3.8	-1.1	-4.2
	7	10.0	4.6	4.3	3.9	2.0	-0.3	0.7	0.1	1.3	-0.5	-0.7	-0.1	-0.7	-2.7	-0.4	-2.0	1.0
	8	10.0	4.2	3.8	3.9	2.1	2.0	3.2	1.3	2.7	-0.3	4.3	-0.7	7.7	-1.7	7.0	-1.0	10.9
	9	10.0	5.8	5.3	5.1	1.9	0.0	1.0	0.5	1.6	-0.9	3.8	-0.5	7.6	-4.1	3.8	-3.5	8.3
	10	10.0	4.1	2.3	3.5	1.1	2.3	3.3	2.3	3.4	0.4	5.0	0.4	8.5	-4.9	-1.2	-5.1	0.8
		<i>0.0</i>	<i>16.2</i>	<i>12.7</i>	<i>12.6</i>	<i>3.4</i>	<i>1.6</i>	<i>6.2</i>	<i>1.3</i>	<i>7.1</i>	<i>0.6</i>	<i>13.4</i>	<i>0.4</i>	<i>36.7</i>	<i>23.6</i>	<i>18.0</i>	<i>16.2</i>	<i>31.1</i>
100,000	1	10.0	-4.0	-3.8	-4.8	-3.3	-2.0	-3.5	-2.0	-3.0	0.4	-1.3	-0.4	-2.0	-1.2	-2.8	-2.0	-3.2
	2	10.0	-4.9	-4.7	-4.8	-3.5	-2.1	-3.6	-1.9	-3.0	-0.6	-2.3	-0.5	-2.2	3.3	0.9	2.7	0.4
	3	10.0	-5.0	-4.8	-4.7	-3.5	-0.9	-0.7	-1.0	-0.9	-0.7	-2.4	-0.4	-2.1	2.5	0.3	2.8	0.6
	4	10.0	-0.9	-0.5	-1.5	-0.3	-1.2	-1.1	-1.0	-0.8	1.3	1.3	0.7	0.6	12.4	8.3	12.2	6.8
	5	10.0	-1.8	-1.3	-1.6	-0.2	1.0	1.8	0.4	0.9	0.4	0.4	0.7	0.6	-2.4	-7.4	-2.2	-6.5
	6	10.0	-1.8	-1.2	-1.6	-0.3	0.0	0.6	0.3	0.8	0.5	0.4	0.7	0.7	-1.6	-3.2	-1.3	-2.6
	7	10.0	4.3	2.5	4.6	1.7	-0.1	0.2	0.2	0.4	0.2	0.3	0.6	0.5	-2.6	-0.6	-2.4	-0.3
	8	10.0	5.2	6.0	4.9	4.0	2.4	2.8	1.7	1.9	-0.2	1.6	-0.4	1.3	-1.4	5.4	-1.2	5.3
	9	10.0	4.5	3.8	4.8	2.7	1.4	1.7	1.7	1.8	-0.6	1.0	-0.4	1.3	-3.9	1.6	-3.7	1.8
	10	10.0	4.5	3.9	4.8	2.8	1.4	1.8	1.6	2.0	-0.7	0.9	-0.5	1.2	-5.1	-2.4	-5.1	-2.2
		<i>0.0</i>	<i>15.9</i>	<i>13.6</i>	<i>16.7</i>	<i>7.0</i>	<i>2.1</i>	<i>4.4</i>	<i>1.8</i>	<i>3.2</i>	<i>0.4</i>	<i>2.0</i>	<i>0.3</i>	<i>1.9</i>	<i>23.0</i>	<i>18.0</i>	<i>22.1</i>	<i>14.1</i>

Contributions for strategies MC, MP, IC, and IP are given as deviations from those of strategy, E. The variance of contributions across strains is given in italics. E, equal numbers from each strain; MC, minimization of coancestry based on mean strain values; MP, maximization of phenotype with restriction on coancestry based on mean strain values; IC, minimization of coancestry based on individual values; IP, maximization of phenotype with restriction on coancestry based on individual values.

achieving a high phenotypic level in the base population, it would be worthy to keep individuals not only from large but also from small strains. In this situation a high phenotypic level can be obtained without reducing too much the diversity maintained.

In the Selection scenario, the variance of contributions across strains was much lower than in the Drift scenario (1.3–6.2 vs. 12.6–16.7; **Table 2**) except for strategy IP with low marker density. This is a consequence of the greater uniformity between strains, at least for the genetic variability of the trait. In the Selection scenario all strains had been under directional selection for the same trait and all had the same size. The only difference between strains was the strength of selection. Those under the weakest selection pressure (8, 9, and 10) had a higher effective population size (N_e), maintained higher levels of genetic diversity and, thus, in average contribute more to the base population. Strains subjected to a strong selection pressure were those with the lowest contributions, even under the strategies directed to keep high levels of trait performance (see strains 1 and 2, Selection scenario in **Table 2**). This is due to the relative long time period since the separation of the strains (20 generations). The small N_e induced by the selection pressure erodes rapidly not only neutral variability but also the genetic variance of the trait. Therefore, strains 1 and 2 reached a selection limit before the rest and presented lower phenotypic

mean values at the time the base population was created (data not shown). Consequently, they cannot contribute much to diversity nor to trait value either.

In the Stabilizing scenario, when the criteria for choosing individuals to constitute the base population included considerations on the phenotypic performance (i.e., strategies MP and IP) the contribution of a strain was proportional to its phenotypic mean (**Table 2**). Contrarily, contributions were almost equalized when the only concern was to keep the highest levels of diversity (strategies MC and IC), given that all strains had an identical population size and a similar selection pressure in the divergence period. An interesting observation was that the variance of contributions across strains greatly decreased for strategies MP and IP when a large panel of SNPs (100,000) was used (see lower section of **Table 2**). With dense genotyping, diversity at selective loci is tightly linked to neutral diversity and, thus, groups of individuals with high phenotype will also have low diversity at the markers. Therefore, optimal solutions include the selection of fewer individuals from the same high performance strain to cope with the restriction on genetic diversity. The lower diversity of high performance groups of individuals is not detected with sparse marker coverage as diversity at selective loci is loosely linked to neutral diversity.

The Mixed scenario included all kind of strains and, thus, the performance was somehow more complex. Notwithstanding, the general patterns highlighted before (i.e., dependence of contributions on the historical size, the intensity of selection and the optimal phenotypic value, respectively) are still observable (see right part of **Table 2**). A particular observation was that the contribution of strains adapted to a low phenotypic value (i.e., strains created under stabilizing selection with a low optimum) was no longer required even with massive genotyping. Also, it was observed that under strategies MP and IP the low diversity of high performance individuals may be compensated for the high diversity of the “drift” strains with reasonable phenotypic levels.

DIVERSITY AND PHENOTYPIC LEVEL IN THE BASE POPULATION

The mean phenotypic value for the trait of interest and the genetic diversity captured in the group of individuals conforming the base population are presented in **Table 3**. It must be pointed out that all strategies except E are explicitly concerned with the maintenance of diversity. However, only MP and IP strategies included considerations about the phenotypic level in the objective function to optimize.

Using strategy IP with a large number of markers in the Mixed scenario led to a base population with mean phenotypic value for the trait 7% higher than under strategy E at the same diversity. The advantage of strategy IP over E in other scenarios ranged from about 4% (Selection scenario) to 6% (Drift scenario), being proportional to the degree of differentiation between strains. This is a logical result if we realize that the margin for improvement with unequal contributions was lower when strains were more similar (for example in the Selection scenario). When relying on the information of few markers, IP led to higher trait means than with dense genotyping (left part of **Table 3**) because diversity at selective loci and at the markers was more loosely linked. Then, groups of individuals with high phenotypic mean for the trait

can be found also showing high levels of diversity at the markers and, thus, coping with the restrictions in the optimization. However, the global diversity at non-marker loci will be low (as stated before) yielding solutions that do not achieve the intended balance between phenotype and diversity (right part of **Table 3**).

In general, the higher the number of markers used to estimate relationships the higher the H_e retained in the base population. However, the differences in H_e with different marker densities were small. The largest difference occurred in the Mixed scenario under strategy IP (3% higher H_e when using 100,000 markers instead of 10 markers). Improvements in the level of H_e maintained under IC, which should be the most efficient strategy in terms of diversity captured were never larger than 1% (**Table 3**).

The genetic diversity maintained in the base population under different strategies was very similar across scenarios. Except for some cases with low number of markers, the strategy capturing the highest levels of neutral H_e was IC, because no other factor but diversity was included in the objective and decisions were taken on the genotypes of the individual candidates and not on the mean strain values. The advantage of this strategy compared to sampling equal number of individuals from each strain (strategy E) ranged from 0.4% (in the Selection scenario) to 1.1% (in the Mixed scenario). This result was obtained because the Selection scenario and Mixed scenario present the highest and the lowest degree of homogeneity between the available strains, respectively.

When using the average strain values (i.e., all individuals from the same strain assumed equivalent) levels of diversity obtained under MC were always lower than with IC although, as stated before, differences were small. With increasing number of markers differences diminished and, eventually, disappeared (see **Table 3**, Stabilizing scenario).

Strategies MP and IP were intended to select individuals with high phenotypic performance but keeping the same level of diversity than E. When using a low number of markers these strategies

Table 3 | Average phenotypic value and expected heterozygosity (in percentage) under different strategies to select individuals for the base population for the four scenarios considered and for different number of markers (n_m).

Scenario	n_m	Phenotypic value					Expected heterozygosity				
		E	MC	MP	IC	IP	E	MC	MP	IC	IP
Drift	100	99.93	99.88	101.44	99.97	106.14	46.21	46.23	45.63	46.32	45.77
	1000	99.81	99.73	100.98	99.76	105.78	46.22	46.49	46.11	46.51	46.14
	100,000	100.06	100.00	101.29	99.98	105.97	46.18	46.48	46.12	46.60	46.19
Selection	100	129.02	129.12	130.45	129.09	134.41	44.54	44.17	43.69	44.26	43.64
	1000	127.96	128.11	128.82	128.07	132.81	44.55	44.56	44.41	44.58	44.36
	100,000	128.26	128.37	129.01	128.04	132.87	44.51	44.57	44.47	44.67	44.51
Stabilizing	100	100.00	99.95	102.70	100.01	107.94	45.14	44.78	44.37	44.87	43.81
	1000	99.99	100.03	102.65	100.08	107.91	45.13	44.75	44.33	44.84	43.80
	100,000	99.02	98.95	99.97	98.99	104.15	44.66	45.00	44.62	45.00	44.65
Mixed	100	106.27	104.07	109.49	104.32	115.50	45.24	45.42	44.70	45.49	43.93
	1000	105.95	103.94	108.15	104.13	113.05	45.34	45.71	45.27	45.74	45.08
	100,000	106.08	103.97	108.12	103.97	112.58	45.33	45.72	45.31	45.84	45.33

E, equal numbers from each strain; *MC*, minimization of coancestry based on strains values; *MP*, maximization of phenotype with restriction on coancestry based on strains values; *IC*, minimization of coancestry based on individual values; *IP*, maximization of phenotype with restriction on coancestry based on individual values. Standard errors of phenotypic value ranged from 0.13 to 0.25 and those for expected heterozygosity were lower than 0.01%.

did not fulfill the restriction (i.e., H_e was lower under MP or IP than under E; see **Table 3**) as this was introduced in the formulation through the molecular coancestry calculated from the markers but diversity results were obtained from the non-markers genotypes. With the largest panel (i.e., 100,000 SNPs) IP did maintain the same diversity level than E but MP maintained slightly lower values because of the random sampling of individuals within strains.

GENETIC GAIN AND INBREEDING FROM THE BREEDING PROGRAM

The capability to respond to artificial selection depends on the amount of additive genetic variance (V_A) present for the trait. **Figure 1** shows V_A along the 10 generations of phenotypic truncation selection for all scenarios and strategies used to construct the base population. The highest initial V_A corresponded to the Mixed scenario as this was the most heterogeneous scenario in terms of types of available strains. The order for the rest of scenarios was Stabilizing, Drift and Selection, following thus the same pattern as that observed for general variability described in the previous section. Within each scenario, patterns of V_A for each strategy were very similar. The highest values were observed for strategy E and the lowest for strategy IP. A large decrease in V_A was observed in early generations in all scenarios due to the Bulmer effect (Falconer and Mackay, 1996).

Higher V_A values in the founders of the breeding program turned into higher initial responses to selection (right panels in **Figure 2**). After the initial generations of selection, gain increased at a lower rate and, after the 10 generations observed values were, in general, inversely related to the initial V_A (i.e., scenarios with higher initial V_A maintained lower BV gain for the whole period of selection).

The mean breeding value (BV) of the base population was higher for those scenarios including already selected strains (i.e., Selection and Mixed) and close to zero for the other two scenarios (left panels in **Figure 2**). Irrespective of the differences in the rate of gain for each scenario pointed out before, the rank of scenarios in terms of BV remained the same for the 10 generations of selection.

For a particular scenario, strategy IP always provided the highest BVs (**Figure 2**), even although generally started from the lowest V_A (**Figure 1**). When the number of markers used in the creation of the base population was small the advantage (in terms of mean BV) of strategy IP was greater than that presented in **Figure 2** (data not shown) due to the higher initial differences with the rest of strategies in trait performance already shown when constructing the base population. But, as discussed before, strategy IP also resulted in lower H_e values (**Table 3**). In all scenarios, strategies MC and IC performed almost identically for V_A , BV and gain, especially when using a large panel of markers (left panel in **Figure 1** and both in **Figure 2**). Strategy E yielded similar results to those from MC and IC except for the Mixed scenario at early generations where the average BV was higher for E (left panels in **Figure 2**). However, at the end of the 10 generations of selection average BV for E, MC, and IC equalized.

It must be highlighted that the trait under selection was simulated with an additive gene action within and across loci. This is the reason for a continuous decay of V_A for the trait and the

corresponding decrease of genetic gain between consecutive generations (**Figures 1, 2**, respectively). In traits with an important non-additive component the selection process may generate new additive variance which could lead to the maintenance of levels of response to selection larger than expected under a pure additive model.

The Drift scenario started from the highest H_e levels and also showed the lowest rate of loss of diversity along the breeding program (right panels in **Figure 1**). On the other hand, the fastest decrease in diversity was observed in the Mixed scenario and this was related to the large initial responses obtained under this scenario.

For all scenarios, populations arising from strategies accounting for the phenotypic level of the founders (i.e., MP and IP) lost more H_e during the 10 generation of selection than strategies aiming just at keeping diversity (right panels in **Figure 1**). This could be due to the fact that in groups of selected individuals the genetic variance for the trait would be more correlated to the global genetic diversity under MP and IP strategies than in the other strategies. Consequently, during the breeding programme the reduction in V_A inherent to the selection process also imply greater reductions in H_e across generations. When initial breeders were chosen based on the genotypes for few markers, populations obtained following IC strategy maintained higher levels of diversity along the generations of selection than when using MC (data not shown) but when a large panel of SNPs was used the performance of both strategies was similar (right panels in **Figure 1**).

As mating was at random throughout the selection process, average inbreeding (F) and coancestry (f) coefficients run in parallel, with the expected lag for F . Therefore, only results for f are shown. Especially for the Mixed scenario (and to a lesser extent for the Stabilizing scenario) f was higher for strategies MC and IC than for the rest (left panels in **Figure 3**). The reason is that strategies MC and IC keep higher numbers of “low performance” individuals whose descendants will not be selected, leading to higher Δf and, thus, to lower N_e than expected. This effect was not detectable in the Selection and Drift scenarios due to the higher homogeneity between strains and individuals for the phenotypic level of the trait.

Irrespective of the scenario, at the beginning of the breeding program there was an increase in Δf (right panels in **Figure 3**) that was due to the removal of individuals with low genetic BV for the trait. Afterwards, Δf stabilized in the Drift and Selection scenarios around 0.4%. This figure is higher than the expected rate ($\Delta f = 0.25\%$) for a random selection population of size 200 (i.e., the number of selected individuals each generation; Woolliams and Bijma, 2000) because between-family selection occurs. In the Stabilizing and Mixed scenarios, Δf monotonically decreased reaching levels closer to 0.25%.

DISCUSSION

The present study has shown that the use of phenotypic information of the candidate strains and the use of genome-wide marker information to infer relationships within and between strains can help to optimize the proportion of individuals to be sampled from each strain when creating base populations for breeding programs

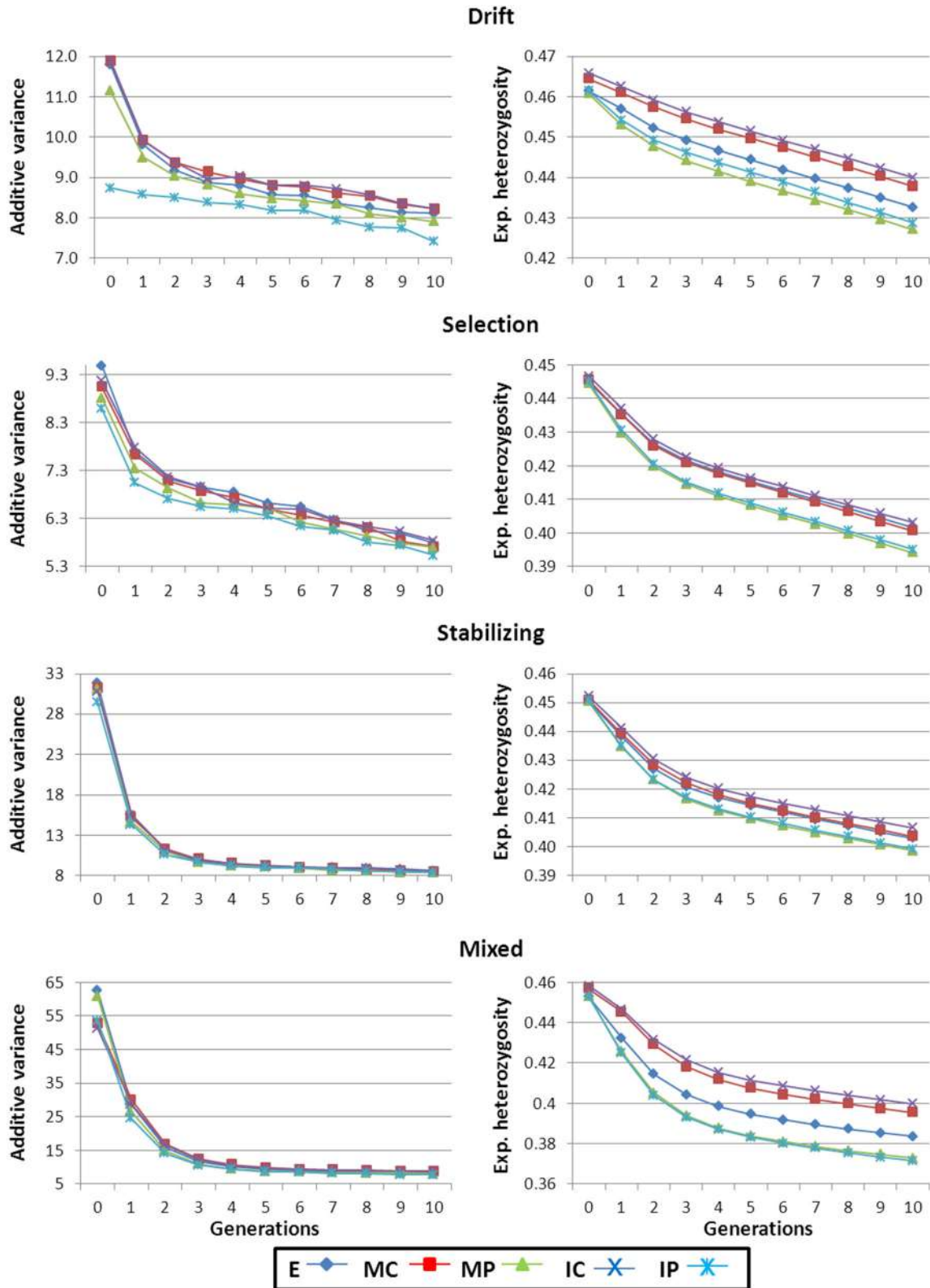


FIGURE 1 | Additive genetic variance for the selected trait and expected heterozygosity for the non-marker loci along the generations of selection. Results shown correspond to base populations obtained using 100,000 markers. E, equal numbers from each strain; MC,

minimize mean strain coancestry values; MP, maximize mean strain phenotypic value with a restriction on coancestry; IC, minimize individual coancestry; and IP, maximize individual phenotypic value with a restriction on coancestry.

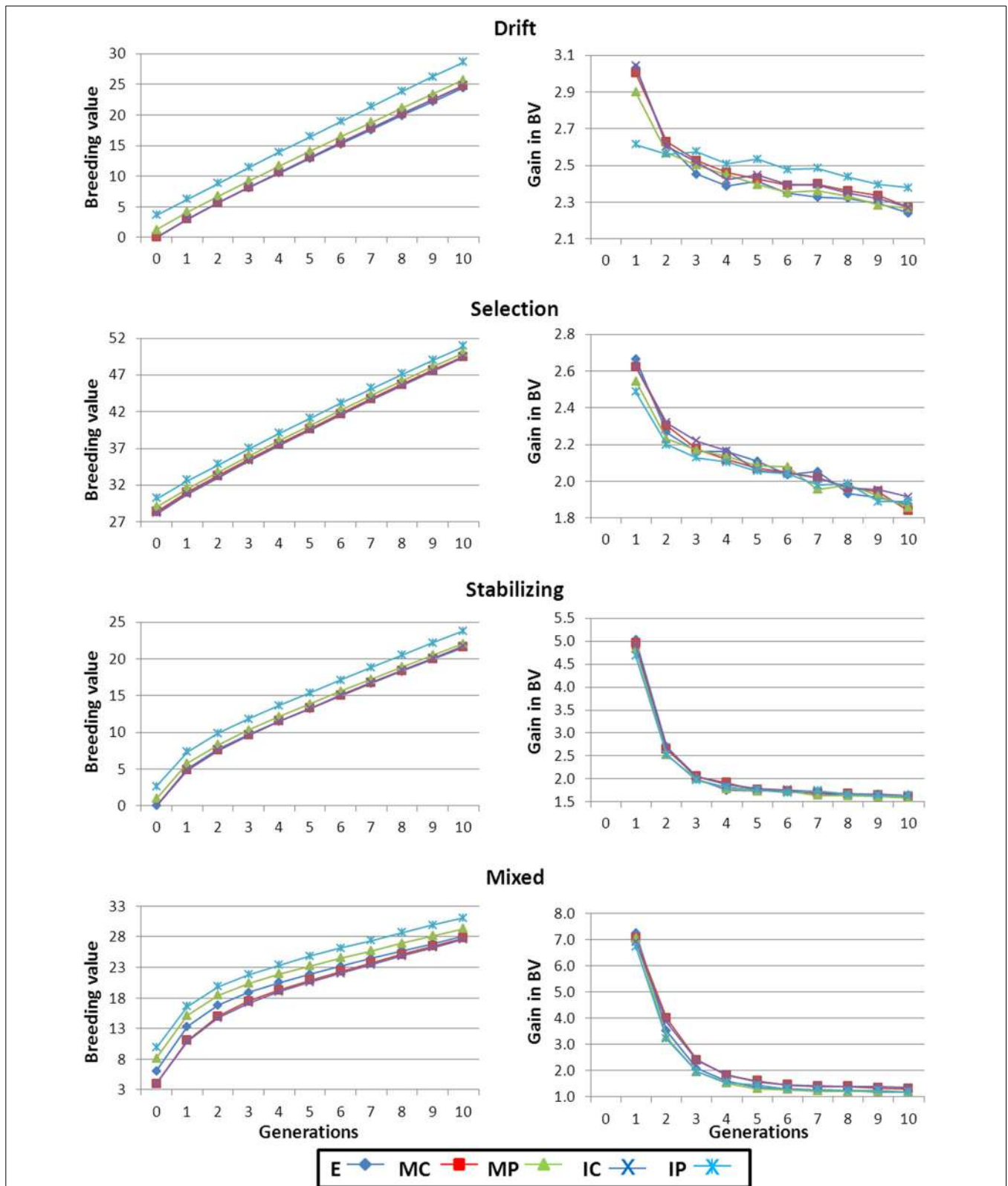


FIGURE 2 | Average breeding value and gain in breeding value along the generations of selection. Results shown correspond to base populations obtained using 100,000 markers. E, equal numbers from each strain; MC,

minimize mean strain coancestry values; MP, maximize mean strain phenotypic value with a restriction on coancestry; IC, minimize individual coancestry; and IP, maximize individual phenotypic value with a restriction on coancestry.

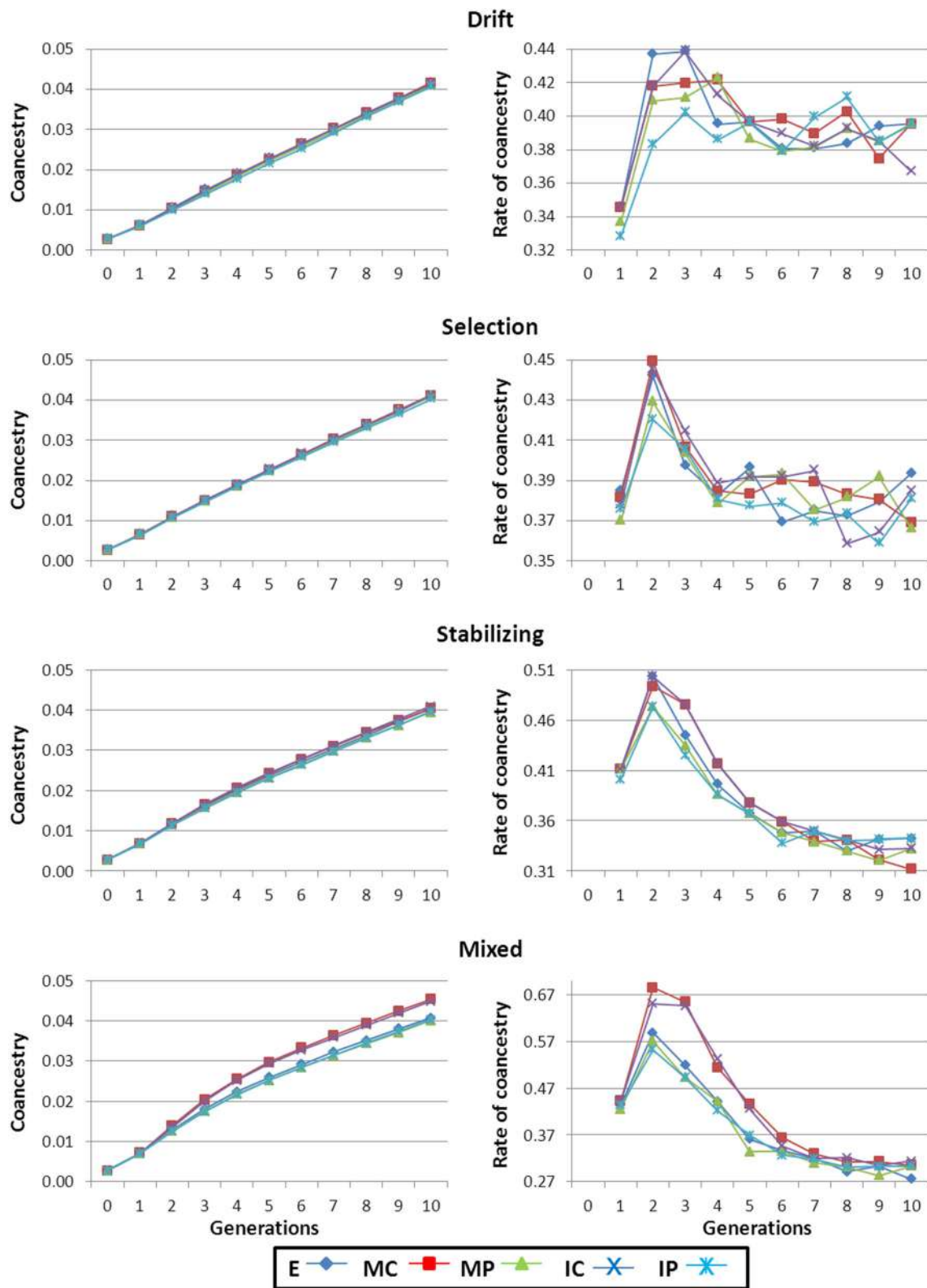


FIGURE 3 | Average genealogical coancestry coefficient and rate of coancestry along the generations of selection. Results shown correspond to base populations obtained using 100,000 markers. E, equal numbers from each strain; MC, minimize mean strain

coancestry values; MP, maximize mean strain phenotypic value with a restriction on coancestry; IC, minimize individual coancestry; and IP, maximize individual phenotypic value with a restriction on coancestry.

in aquaculture species. The advantage of using this information is reflected in the phenotypic and breeding values obtained at the beginning of the program and at the genetic diversity captured by the base population. The advantage remains during the subsequent generations of selection, making the breeding program more profitable. The study has not been designed for a particular species. Instead, we have considered a general genome architecture and a population structure that fit most aquaculture species.

Traditionally, when creating base populations for the establishment of a breeding program, no information was available about the genetic relationships between candidate strains or between individuals within strains. Consequently, the usual strategy was to collect equal number of individuals from as many strains as possible (Holtmark et al., 2006, 2008a,b). However, the increasing amount of molecular markers developed for aquaculture species provide us with the opportunity of estimating genetic relationships within and between strains and to optimize the contribution of each strain. In this study it has been shown that strains harboring low levels of genetic diversity should contribute less individuals in order to maximize the global diversity of the base population.

In the creation of base populations, one important point to determine beforehand is if the objective is to maximize the genetic variance for a particular trait (i.e., the trait in the breeding goal) or to maintain the highest global diversity. For the former objective, Bennewitz and Meuwissen (2005) showed that the optimal strategy is what they called maximum variance total (MVT), which gives more weight to the variance between strains than within strains. However, although the profitability of the breeding program depends on the performance for the target trait, diversity must be also maintained for other traits that are likely to be included in future breeding objectives and for fitness related traits. Following this logic the methodology should be minimizing the global coancestry which poses the same weight to within and between strains diversity. The latter method is similar to minimizing the long-term inbreeding of the population as demonstrated by Eding and Meuwissen (2001) and was the chosen strategy for the present study. Accordingly, in our results for three of the simulated scenarios, strategy E yielded the highest V_A for the target trait in the base population. The lower levels of V_A observed under strategies MC and IC were due to the fact that the objective was to maximize the global genetic diversity measured as H_e across all the genome. Hayes et al. (2006) compared random with marker-based optimized selection of breeders from a single population of Atlantic salmon in terms of the genetic variance captured for three different traits (growth and two disease traits). They followed an equivalent methodology to strategy IC presented in this study and found higher additive variances in the breeders for the disease resistance traits but a lower variance for growth when optimizing the selection than when breeders were chosen at random. The explanation for these contrasting results was the different genetic architecture of the traits. It must be noticed that the simulated trait in the present study was controlled by a large number of additive loci and had an intermediate heritability typical for growth. This is the reason for similar performance (i.e., highest levels under E than IC strategy) observed

in Hayes et al. (2006) for growth and in the present study (at least for three of the simulated scenarios). Another problem for the interpretation of the results in Hayes et al. (2006) is that they employed 237 AFLPs and this number may be not enough for obtaining a high correlation between diversity at markers and at loci controlling growth.

In concordance with the previous considerations, in this study the highest levels of global diversity (H_e measured at the non-marker loci) were captured when optimizing the creation of the base population using individual genotypes (strategy IC) with a large number of SNPs, although differences with the strategy equalizing proportions (strategy E) were small. Surprisingly, scenarios with a limited number of markers (i.e., 100) implied only a loss in H_e of 1% when comparing with results from using large numbers. In any case, when using few markers the H_e maintained under IC was sometimes lower than under the E strategy because of the lack of correlation between diversity at markers and diversity in the rest of the genome.

When individual information was absent (i.e., strategy MC) there was a reduction in the ability to capture diversity respect IC strategy whatever the number of markers used. Notwithstanding, values of H_e obtained when relying on strain averages were less than 2% lower than those observed when individual genotypes of candidates were available. This is an appealing result for cases where the budget is low and no all candidates can be genotyped.

Beyond all considerations about the genetic diversity, we must remember that the short-term profitability of a breeding program depends on actual mean levels of the phenotypic value for the trait of interest (as long as the breeding goal does not change and no fitness troubles arise in the population). The present study has shown that the mean phenotype of the selected individuals should be also accounted for when constructing the base population if that information is available. The loss of profit resulting from including low performance individuals in the base population may be not economically compensated in a reasonable period of time even if the response to selection is high due to a wider genetic variance for the trait. In fact our results showed that superiority of individuals selected under strategy IP last for the 10 generations of selection.

The presented results showed that, when reliable information is available for a fixed set of strains, a compromise solution between diversity and performance can be found when creating base populations. Having as a reference point the strategy randomly sampling the same number of individuals from each strain (strategy E, equivalent to the methodology used in Holtmark et al., 2006, 2008a,b) up to 7% higher levels of phenotypic performance can be achieved under strategy IP at the same level of global diversity (H_e measured at the non-marker loci) when using a large panel of SNPs to genotype all candidates. Depending on the market value for the increase of one unit of the target trait this could translate into a large economic gain. Moreover, results showed that phenotypic values remain higher during all generations of artificial selection that were simulated from the base population under the IP strategy. Therefore, there was a clear superiority of fishes obtained using this strategy.

It must be realized that giving a large weight to the phenotypic value for a particular trait will have consequences on other

correlated traits. Special attention should be paid for traits negatively correlated with the target trait which could be of potential interest. In such situations a useful strategy would be to use an index that includes several traits in the objective function to be maximized or to include an additional restriction in the optimization to ensure a minimum acceptable phenotypic level for the secondary traits.

It must be stressed that genotyping for a limited number of markers may give undesirable results because diversity at those markers will be loosely related to diversity at non-marker loci and to diversity at loci controlling the trait. Consequently, individuals with a high phenotypic performance may actually maintain little global diversity but still exhibit by chance high levels of diversity at the markers. Our results show that, when relying in few markers, lower H_e levels were found when optimizing contributions using the IP strategy than using the E strategy even when a restriction was imposed to keep the same level of diversity (Table 3). However, results from our simulations suggest that, although it will depend on the particular characteristics of the species under management and the genetic architecture of the available strains, about 1000 SNPs could be enough to efficiently create base populations in aquaculture as no relevant improvements are obtained by increasing further the number of markers.

For all scenarios, strategies accounting for the phenotypic level of the founders (i.e., MP and IP) started the selection program from lower values of V_A for the target trait. This fact did not preclude these strategies to maintain the initial advantage in performance during the 10 generations of selection.

Results of this study suggest that if “healthy” commercial strains (i.e., where diversity is not exhausted and no problems of inbreeding exist) are available they should be used to form the base population because they provide a higher starting performance level. This is even clearer when the objective is to complement the breeders’ population in an ongoing selection program. Contrarily, the use of wild adapted strains with low performance would be only recommended if we suspect that unique information for other traits of interest is present in them. This could be the situation for strains naturally resistant to a particular disease. Otherwise, the general diversity that they could provide will not compensate for the lower trait phenotypic mean.

The small differences in H_e observed under different strategies in our simulations could be due to the large number of individuals selected to form the base population (200) making strategy E to perform so well that the other strategies have difficulties in improving H_e . In an extra scenario run with a smaller set of candidates (only four strains with 20 individuals each), harboring lower levels of diversity (mutation-drift equilibrium reached for a population of 100 individuals) and selecting a lower number of breeders (24), strategies implying optimized proportions showed still only slightly greater advantages over strategy E (2% increase). In any case, the differences observed in phenotypic values make worthy to optimize the construction of base populations in aquaculture and levels of diversity should be also accounted for in that task to get an appropriate balance.

Another advantage of using molecular information, beyond balancing phenotypic values and diversity when optimizing the construction of a base population, is that it provides us with the

possibility of estimating the actual relationships between breeders in the base population itself. These relationships can be used for calculating EBVs through BLUP methodology and also for controlling the rate of inbreeding through Optimal Contribution strategies. Holtsmark et al. (2008b) studied the effects on the performance of the breeding program across generations of assuming unrelated and non-inbred founders when they are not. They concluded that an incorrect estimation of the relationships between and within strains and individuals leads to sub-optimal use of subpopulations with an increased risk of loss of alleles of direct and strategic relevance to the breeding program.

If genotypes for dense panel of markers and phenotypes are available for the same individuals (the candidates to be part of the base population or related individuals), the additive effect of each SNP can be calculated in the same way as in the Genomic Selection methodology (Meuwissen et al., 2001). Thus, genomic value of candidates can be calculated and used to take decisions instead of their phenotypic value.

In the present study the selected trait was simulated with an additive gene action both within and across loci. However, traits with commercial interest may have an important dominant component. Dominant effects can be also estimated and then used to design the mating scheme between breeders, at least to form the families from which the selection program will start. This way, the effects of heterosis can be accounted for and extra responses can be obtained in the first round of selection (Toro and Varona, 2010).

In short life species, where individuals may be not reproductively active by the time their phenotypic records and genotypes are available, it would be difficult to implement strategies based on individual information (i.e., IC and IP). However, even in such situations strain information can be used to optimize the creation of the base population, as proved in the present study. The impossibility of controlling the specific matings with mass spawning species does not interfere with the optimization of the base population either. Phenotypic levels and diversity can be optimized at the start of the program although responses in subsequent generations of selection may differ to those shown here given that selected breeders would need to be mated in groups.

In livestock terrestrial species, breeding programs have been running for many years and, thus, it is not very likely that there is a need for creating new base populations. Notwithstanding, our conclusions go beyond the scope of creating base populations. For instance, they may help to take decisions when creating a gene bank for any species. When the aim of such a bank is to store the genetic diversity from the available strains, the same strategies can be applied as when creating the base population of breeders. When, in the future, the stored material will be used for creating a live population or for complementing a breeding program the same results and consequences than those presented here are expected. If the phenotypic values of the candidates are not taken into consideration when determining the sampling scheme the starting population will show low levels for the trait of interest. Finally, the methodology presented in this study is also useful when the objective is to create a “core” live population within an ex-situ conservation program that aim at collecting the genetic

diversity that exist in all available populations. This scenario is common for local breeds of livestock species.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (KBBE.2013.1.2-10) under grant agreement n° 613611.

REFERENCES

- Bennewitz, J., and Meuwissen, T. H. E. (2005). A novel method for the estimation of the relative importance of breeds in order to conserve the total genetic variance. *Genet. Sel. Evol.* 37, 315–337. doi: 10.1186/1297-9686-37-4-315
- Bürger, R., Wagner, G. P., and Stettinger, F. (1989). How much heritable variation can be maintained in finite populations by mutation-selection balance? *Evolution* 43, 1748–1766. doi: 10.2307/2409390
- Caballero, A., and Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 3, 289–299. doi: 10.1023/A:1019956205473
- Eding, H., and Meuwissen, T. H. E. (2001). Marker based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.* 118, 141–159. doi: 10.1046/j.1439-0388.2001.00290.x
- Eding, J. H., Crooijmans, R. P. M. A., Groenen, M. A. M., and Meuwissen, T. H. E. (2002). Assessing the contribution of breeds to genetic diversity in conservation schemes. *Genet. Sel. Evol.* 34, 613–633. doi: 10.1186/1297-9686-34-5-613
- Ekmath, A. E., Dey, M. M., Rye, M., Gjerde, B., Abella, T. A., Sevilleja, R., et al. (1998). "Selective breeding for Nile tilapia for Asia," in *Proceedings of the 6th World Congress on Genetics Applied in Livestock Production*, Vol. 27 (Guelph, ON), 89–96.
- Ekmath, A. E., Tayamen, M. M., Palada-de Vera, M. S., Danting, J. C., Reyes, R. A., Dionisio, E. E., et al. (1993). Genetic improvement of farmed tilapia: the growth performance of eight strains of *Oreochromis niloticus* tested in different environments. *Aquaculture* 111, 171–188. doi: 10.1016/0044-8486(93)90035-W
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th Edn. Essex: Longman.
- Gjedrem, T., Gjøen, H. M., and Gjerde, B. (1991). Genetic origin of Norwegian farmed Atlantic salmon. *Aquaculture* 98, 41–50. doi: 10.1016/0044-8486(91)90369-I
- Gómez-Romano, F., Villanueva, B., de Cara, M. A. R., and Fernández, J. (2013). The benefit of using molecular coancestry in the management of populations under conservation and its dependency on effective population size and marker density. *Genet. Sel. Evol.* 45:38. doi: 10.1186/1297-9686-45-38
- Hayes, B., He, J., Moen, T., and Bennewitz, J. (2006). Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. *Aquaculture* 255, 573–578. doi: 10.1016/j.aquaculture.2005.11.038
- Holtmark, M., Klemetsdal, G., Sonesson, A. K., and Woolliams, J. A. (2008a). Establishing a base population for a breeding program in aquaculture, from multiple subpopulations, differentiated by genetic drift: I. Effects of the number of subpopulations, heritability and mating strategies using optimum contribution selection. *Aquaculture* 274, 232–240. doi: 10.1016/j.aquaculture.2007.11.042
- Holtmark, M., Klemetsdal, G., Sonesson, A. K., and Woolliams, J. A. (2008b). Establishing a base population for a breeding program in aquaculture, from multiple subpopulations, differentiated by genetic drift: II. Sensitivity to assumptions on the additive genetic relationships of base animals. *Aquaculture* 274, 241–246. doi: 10.1016/j.aquaculture.2007.11.043
- Holtmark, M., Sonesson, A. K., Gjerde, B., and Klemetsdal, G. (2006). Number of contributing subpopulations and mating design in the base population when establishing a selective breeding program for fish. *Aquaculture* 258, 241–249. doi: 10.1016/j.aquaculture.2006.03.039
- Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., Downing, A., et al. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15:90. doi: 10.1186/1471-2164-15-90
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* 220, 671–680. doi: 10.1126/science.220.4598.671
- Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., et al. (2014). Development of the catfish 250K SNP array for genome-wide association studies. *BMC Res. Notes* 7:135. doi: 10.1186/1756-0500-7-135
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome wide dense marker maps. *Genetics* 157, 1819–1829.
- Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., et al. (2014). The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol. Ecol. Res.* doi: 10.1111/1755-0998.12337. [Epub ahead of print].
- Toro, M. A., and Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.* 42:33. doi: 10.1186/1297-9686-42-33
- Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.* 25, 138–193. doi: 10.1016/0040-5809(84)90017-0
- Woolliams, J. A., and Bijma, P. (2000). Predicting rates of inbreeding in populations undergoing selection. *Genetics* 154, 1851–1864.
- Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., et al. (2014). Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics* 15:307. doi: 10.1186/1471-2164-15-307

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 July 2014; accepted: 06 November 2014; published online: 25 November 2014.

Citation: Fernández J, Toro MÁ, Sonesson AK and Villanueva B (2014) Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Front. Genet.* 5:414. doi: 10.3389/fgene.2014.00414

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Fernández, Toro, Sonesson and Villanueva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes

Ali Ali^{1,2}, Caird E. Rexroad³, Gary H. Thorgaard⁴, Jianbo Yao⁵ and Mohamed Salem^{1,5*}

¹ Department of Biology, Middle Tennessee State University, Murfreesboro, TN, USA

² Department of Zoology, Faculty of Science, Benha University, Benha, Egypt

³ The National Center for Cool and Cold Water Aquaculture, United States Department of Agriculture Agricultural Research Service, Leetown, WV USA

⁴ School of Biological Sciences, Washington State University, Pullman, WA, USA

⁵ Division of Animal and Nutritional Science, West Virginia University, Morgantown, WV, USA

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Alexandre Rodrigues Caetano, Embrapa Recursos Genéticos e Biotecnologia, Brazil

Ed Smith, Virginia Polytechnic Institute and State University, USA

*Correspondence:

Mohamed Salem, Department of Biology, Box 60, 2055 SCI, 1500 Greenland Dr., Murfreesboro, TN 37132, USA
e-mail: mohamed.salem@mtsu.edu

Resistance against diseases affects profitability of rainbow trout. Limited information is available about functions and mechanisms of teleost immune pathways. Immunogenomics provides powerful tools to determine disease resistance genes/gene pathways and develop genetic markers for genomic selection. RNA-Seq sequencing of the rainbow trout spleen yielded 93,532,200 reads (100 bp). High quality reads were assembled into 43,047 contigs. 26,333 (61.17%) of the contigs had hits to the NR protein database and 7024 (16.32%) had hits to the KEGG database. Gene ontology showed significant percentages of transcripts assigned to binding (51%), signaling (7%), response to stimuli (9%) and receptor activity (4%) suggesting existence of many immune-related genes. KEGG annotation revealed 2825 sequences belonging to “organismal systems” with the highest number of sequences, 842 (29.81%), assigned to immune system. A number of sequences were identified for the first time in rainbow trout belonging to Toll-like receptor signaling (35), B cell receptor signaling pathway (44), T cell receptor signaling pathway (56), chemokine signaling pathway (73), Fc gamma R-mediated phagocytosis (52), leukocyte transendothelial migration (60) and NK cell mediated cytotoxicity (42). In addition, 51 transcripts were identified as spleen-specific genes. The list includes 277 full-length cDNAs. The presence of a large number of immune-related genes and pathways similar to other vertebrates suggests that innate and adaptive immunity in fish are conserved. This study provides deep-sequence data of rainbow trout spleen transcriptome and identifies many new immune-related genes and full-length cDNAs. This data will help identify allelic variations suitable for genomic selection and genetic manipulation in aquaculture.

Keywords: spleen transcriptome, annotation, KEGG, immune-related genes, spleen-specific genes, full-length cDNA

INTRODUCTION

Teleost fish are the first class of vertebrates that have the elements of both innate and adaptive immune responses (Whyte, 2007). Innate immunity is more important in teleosts as the first line of defense due to the restrictions on adaptive immunity in suboptimal environments (Ullal et al., 2008). Teleost fish have no bone marrow or lymph nodes. Immunogenomics can be used in clarifying the origin and evolution of the immune systems (Kaiser et al., 2008).

The ability of fish to combat viral, bacterial or parasitic pathogen is affected by genetic factors. Thus, genetic selection can improve disease resistance and provide prolonged protection against pathogens (Skamene and Pietrangeli, 1991; Wiegertjes et al., 1996; Van Muiswinkel et al., 1999; Leeds et al., 2010; Wiens et al., 2013a). In addition, investigations on immune reactions in fish could aid in development of vaccines (Raida and Buchmann, 2009).

Rainbow trout (*Oncorhynchus mykiss*) are widely distributed and cultured aquaculture species used as a food and sport fish (Thorgaard et al., 2002). Additionally, this organism is used as a model species in many different fields of research such as cancer biology (Tilton et al., 2005), toxicology (Köllner et al., 2002), nutrition (Wong et al., 2013), evolutionary biology (Taylor et al., 2011), and immunology (Nya and Austin, 2011). Rainbow trout are larger than other fish model species, making them a source of large quantities of specific tissues and cells for immunological, biochemical and molecular analyses. Several genomic resources have been developed for research to help improve rainbow trout commercial production traits including disease resistance. The list includes clonal lines (Young et al., 1996; Thorgaard et al., 2002), BAC libraries (Palti et al., 2004), genetic linkage maps (Young et al., 1998; Sakamoto et al., 2000; Rexroad et al., 2008; Palti et al., 2011; Guyomard et al., 2012), microarrays (Salem et al., 2008), expressed sequence tags (ESTs) (Rexroad et al., 2003;

Sánchez et al., 2009; Boussaha et al., 2012), single nucleotide polymorphisms (SNPs) (Sánchez et al., 2009; Amish et al., 2012; Boussaha et al., 2012; Houston et al., 2012; Salem et al., 2012; Christensen et al., 2013; Colussi et al., 2014; Palti et al., 2014), next-generation sequence read archives (SRA) and genome as well as transcriptome reference assemblies (Salem et al., 2010; Sanchez et al., 2011; Berthelot et al., 2014; Fox et al., 2014). Several studies have identified many immune-related key genes and gene networks (Thorgaard et al., 2002; Cerdà et al., 2008; Chiu et al., 2010; Zhang et al., 2011). Immunogenomic studies of rainbow trout have established an agricultural importance due to their direct and immediate contributions to the aquaculture industry (Thorgaard et al., 2002).

The spleen is a primary hematopoietic and peripheral lymphoid organ (Zapata et al., 2006; Mahabady et al., 2012). This organ has melano-macrophage centers for breakdown of aged erythrocytes, and T-like as well as B-like cells for antigen trapping. In addition, spleen has a role in antigen presentation and initiation of the adaptive immune response (Espenes et al., 1995; Zapata et al., 1997; Chaves-Pozo et al., 2005; Whyte, 2007; Alvarez-Pellitero, 2008). Positive genetic correlation exists between bacterial cold water disease resistance and spleen index in domesticated rainbow trout (Wiens et al., 2013b).

Transcriptomic profiling is useful in revealing spleen genes that are engaged in the innate and adaptive immune responses and expressed as a result of the presence of toxicants or infection (Pereiro et al., 2012). RNA-Seq (whole transcriptome sequencing) is an effective tool for identifying the functional complexity of transcriptomes, alternative splicing, non-coding RNAs, new transcription units and assembly of full-length genes (Xiang et al., 2010; Huang et al., 2011; Salem et al., 2012; Djari et al., 2013). This deep sequencing technology gives low background noise, aids in identifying allele-specific expression and reveals weakly expressed transcripts. Bioinformatics algorithms were developed facilitating transcriptomic profiling (Yang et al., 2012). Therefore, RNA-Seq is a valuable tool in studying functional complexity of the spleen transcriptome and identifying immune-relevant genes and signaling networks (Nie et al., 2012).

In this study, we aimed to (1) characterize the transcriptome of rainbow trout spleen and (2) identify spleen-specific and immune-relevant genes (including full-length cDNAs) that could be used to develop genetic markers for disease resistance. Identifying the networks associated with such genes will be helpful in generating new technologies to improve aquaculture (Takano et al., 2011).

RESULTS AND DISCUSSION

The spleen transcriptome was sequenced from an apparently healthy single homozygous doubled-haploid fish from the Swanson clonal line, the same line used for BAC library construction (Palti et al., 2004) and sequencing both of the whole transcriptome (Salem et al., 2010) and the whole genome reference (Berthelot et al., 2014). A single doubled-haploid fish was used to help overcome the assembly complications associated with the tetraploid genome of the rainbow trout (Allendorf and Thorgaard, 1984). Spleen transcriptome RNA-Seq data were *de novo* assembled into contigs. Assembled contigs were

analyzed and annotated to identify genes that are predominantly expressed in the spleen and genes that are involved in immune signaling pathways. Spleen sequencing data yielded a total of 93,532,200 reads with a read length of at least 100 bp. After filtration to remove the adaptors, low complexity reads and duplicates, 58,013,135 (62%) high quality reads (Q values $> 33\%$) were obtained and assembled into 43,047 contigs with an average contig length of 1154 nt and N50 equal to 1306 nt.

FUNCTIONAL ANNOTATION

Contigs were searched against the NCBI's non-redundant protein (NR) database using the BLASTX program with E -value of $1.0E-3$. There were 26,333 (61.17%) contigs with hits to the NR database (Table S1). The contigs which had no hits [16,714 (38.83%)] may be attributed to non-coding RNAs, contig mis-assemblies (Grabherr et al., 2011), limited information about protein sequences of related fish in the NCBI database or diverged sequences of rainbow trout due to partial genome duplication (Ravi and Venkatesh, 2008; Lee et al., 2011). Further work toward characterization of the non-coding RNAs is still needed.

Data statistics of the sequencing, assembly and annotations are presented in **Table 1**. A total of 13,780 (88.40%) of the contigs of more than 1000 bp in length had BLAST matches, whereas only 12,543 (51.72%) of contigs shorter than 1000 bp had BLAST hits (**Figure 1**). Short sequences may give false-negative results because they are not long enough to show sequence matches or may lack a representative protein domain (Wang et al., 2012). The identity distribution revealed that 70% of the contigs have greater than 80% similarity and 24% possess identity similarities between 60 and 80%. The E -value distribution of the top

Table 1 | Statistical summary of rainbow trout spleen sequencing, assembly and annotation.

Data generation and filtration

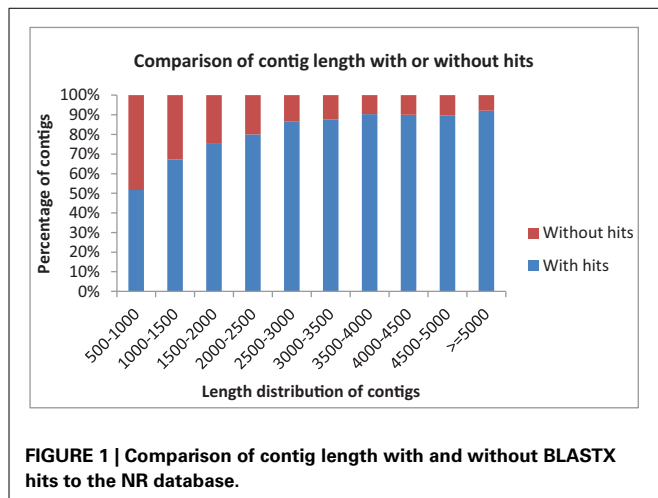
Total number of reads	93,532,200
Total number of duplicates	35,519,065 (38%)
Number of high quality reads	58,013,135 (62%)

Assembly statistics

Number of bases	49,669,063
Number of long contigs (≥ 500)	43,047
Largest contig length (nt)	14,276
Smallest contig length (nt)	500
Average contig length (nt)	1154
N50 (nt)	1306

Annotation

Total number of annotated contigs	43,047
Number of contigs with NR hits	26,333 (61.17%)
Number of contigs without NR hits	16,714 (38.83%)
Number of contigs with KEGG hits	7024 (16.32%)
Number of contigs without KEGG hits	36,023 (83.68%)



hits to the NR database showed that 28% of the assembled contigs showed significant homology to previously deposited sequences (less than $1.0E-50$), and 72% ranged from $1.0E-50$ to 0. The assembled contigs have been submitted to the NAGRP Aquaculture Genome Projects (<http://www.animalgenome.org/repository/pub/MTSU2014.0811/>).

Gene ontology (GO) categorizes gene products and standardizes their representation across species (Consortium, 2008). Contigs with lengths of 500 nucleotides or more (43,047) were annotated using the Blast2GO suite (Conesa et al., 2005; Götz et al., 2008); contigs were assigned to appropriate molecular function, biological process and cellular component GO terms (Ashburner et al., 2000). **Figure 2** shows summary of the GO assignments at the second level in the three areas of gene ontology. GO term distribution was compared to a transcriptome reference that we previously assembled from 13 tissues (Salem et al., 2010) (**Figure 3**). In the biological process category, the most represented terms were related to cellular process (17%), followed by metabolic process (15%) and biological regulation (12%) (**Figure 2A**). These percentages are lower than the corresponding categories at the whole transcriptome reference; 25, 18, and 25%, respectively (**Figure 3A**). Conversely, some immune-relevant sub-categories of metabolic processes exist in higher percentages compared to counterparts at sub-categories at the whole transcriptome reference; response to stimuli (9%) and signaling (7%) compared to 3% response to stimuli and 1% immune system process, respectively (**Figure 3A**). These percentages suggest that we identified a larger number of immune-related genes in this study.

Within the molecular function category, 51, 30, and 4% of the spleen transcripts were assigned to binding, catalytic activity and receptor activity, respectively. The whole transcriptome reference showed 46 and 32% and unlisted percentage, respectively (**Figure 3B**). Pereiro and co-workers suggested several immune-related processes were represented in the binding and catalytic activity categories (Pereiro et al., 2012). In the cellular component categories, a significant percentage of clusters assigned to cell (42%), organelle (27%), macromolecular complex (13%), membrane enclosed lumen (8%) and membrane

7% (**Figure 3C**) compared to their corresponding categories at the whole transcriptome reference; 59, 24, 9, 3% and unlisted percentage, respectively (**Figure 3C**). Discrepancies in the GO distribution profiles may be attributed to differences in the nature of the cDNA libraries, the numbers of sequences used to retrieve GO terms, sequencing technology and the assembly approaches. Information about GO terms is supplied in additional Table S2.

KEGG pathway analysis was carried out to categorize and annotate the assembled contigs (Kanehisa and Goto, 2000; Kanehisa et al., 2012). Searching contigs against the KEGG database yielded 7,024 KEGG hits (16.32% of the total number of transcripts), with 4779 unique hits (11.10% of total number of transcripts). KEGG Orthology (KO) numbers were used to assign sequences to different metabolic pathways (**Table 2**). In total, 2236 (31.83% of the total number of KEGG hits) KEGG annotated sequences were assigned to metabolism that was further classified into carbohydrate metabolism (453 sequences, 20.26%), lipid metabolism (349 sequences, 15.61%) and amino acid metabolism (343 sequences, 15.34%). In addition, 1,394 (19.85%) annotated sequences were assigned to genetic information processing which includes folding, sorting and degradation (484 sequences, 34.72%), translation (459 sequences, 32.93%), replication and repair (255 sequences, 18.29%), and transcription (196 sequences, 14.06%). Further, 1830 sequences (26.05%) were classified as environmental information processing assigning 1508 sequences (82.40%) to signal transduction and 283 sequences (15.46%) to signaling molecules and interaction. The Cellular processes group contained 1421 (20.23%) KEGG-annotated sequences.

Remarkably, the KEGG organismal systems category contained 2825 (40.22%) annotated sequences with the highest number of sequences assigned to immune system (842 sequences, 29.81%) followed by endocrine system (619 sequences, 21.91%), nervous system (526 sequences, 18.62%), digestive system (268 sequences, 9.49%) and development (210 sequences, 7.43%). Assignments of the organismal systems to the last four categories support the previously reported relationships between function of the spleen and other systems of the body. For example, a subset of genes with functions relevant to neurodevelopment was identified in the spleen transcriptome of the house finch (*Haemorrhous mexicanus*) (Backström et al., 2013). Regarding the endocrine functions, it was thought that spleen secretes a hormone-like substance under the control of pituitary gland and adrenal cortex in case of emergencies (Ungar, 1945). Recently, the spleen endocrine function has been confirmed after in-depth studies of its function (Wu, 1998; Horiguchi et al., 2004; Tarantino et al., 2011). A cytokine known as Lymphotoxin was reported to keep the immunological balance of the gastrointestinal tract through regulation of the immune system of the digestive tract which is represented by immune cells, immunoglobulins and intestinal bacteria (Kruglov et al., 2013). In addition, hormones of the gastrointestinal tract activate the immune system in case of gut inflammation (Khan and Ghia, 2010).

TAXONOMIC ANALYSIS

A BLASTX top-hit species distribution of gene annotations showed highest homology to *Salmo salar* (4,833 BLAST hits;

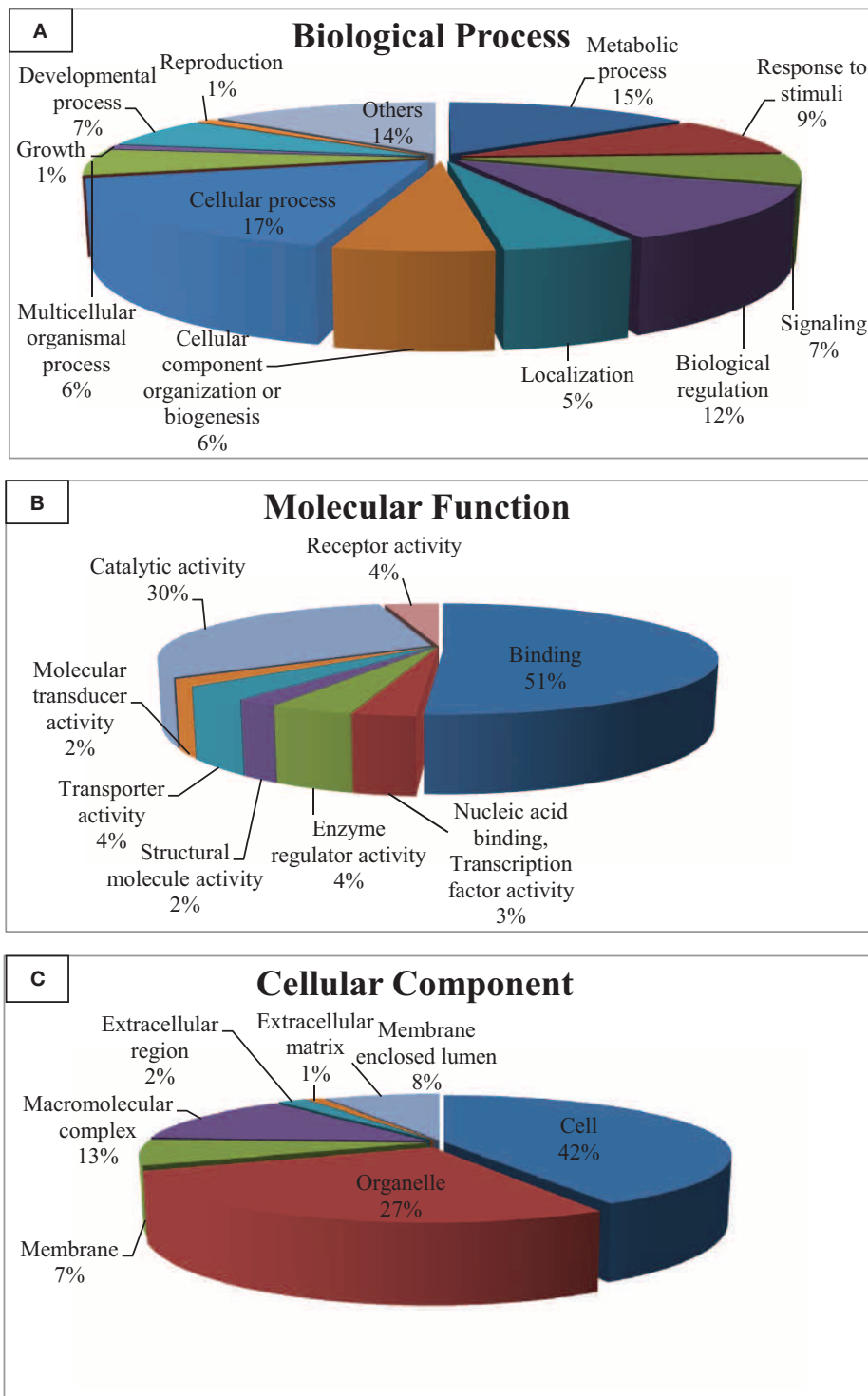
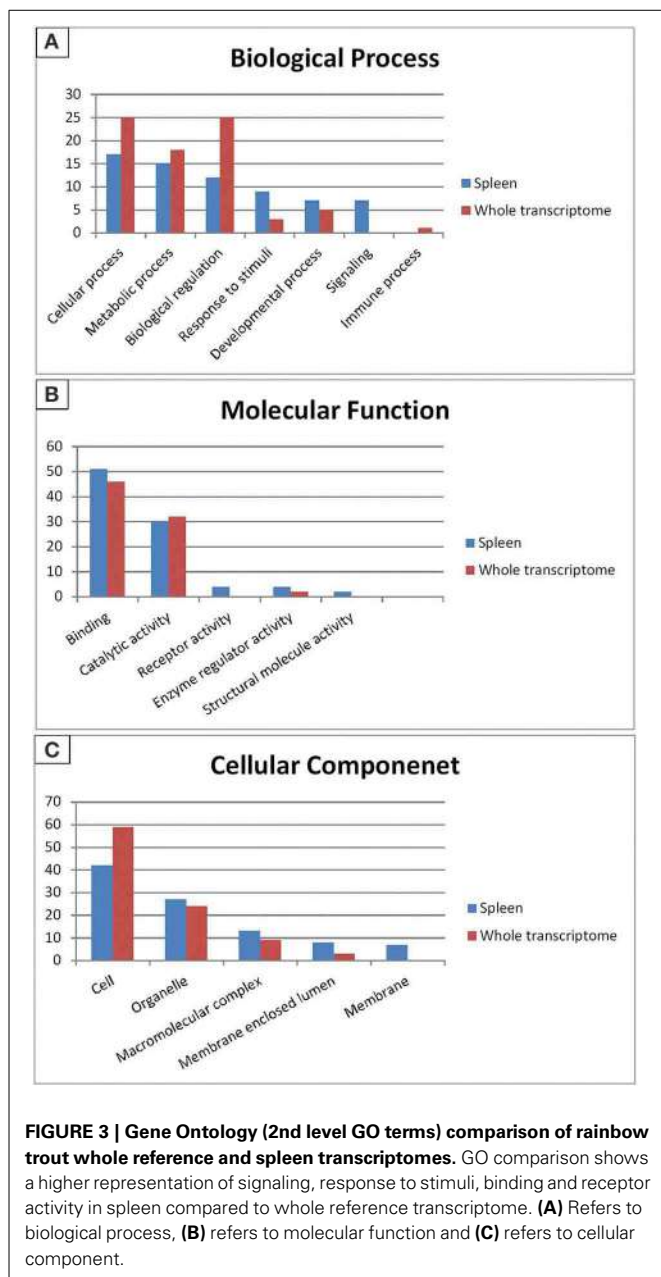


FIGURE 2 | Gene ontology (GO) assignments (2nd level GO terms) for the spleen transcriptomic sequences of rainbow trout. (A) Refers to biological process, **(B)** refers to molecular function and **(C)** refers to cellular component.

18.35%), followed by *Oreochromis niloticus* (16.64%), *Maylandia zebra* (16.47%) and *Danio rerio* (16.12%) (Figure 4). Other fish species in the BLASTX top-hit were *Takifugu rubripes* (5.98%) and *Oryzias latipes* (5.17%). Rainbow trout itself (983 BLAST hits;

3.73%) fell in the seventh position of the top-hit species distribution. This may be explained by identification of a large number of new genes in this study and/or and existence of a limited number of rainbow trout proteins (6965 proteins) that currently available



at NCBI database. The model fish species in the list, *D. rerio*, *T. rubripes* and *O. latipes*, have large number of proteins available in the NCBI database. All first nine species were fish, starting with *S. salar* from the family Salmonidae to which rainbow trout belongs. Therefore, these results support the high quality and high level of phylogenetic conservation of the assembled spleen transcriptome. Other species with known genome sequences appearing in the BLASTX top-hit species distribution were mammals including *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*, *Gallus* (chicken) and the amphibian *Xenopus laevis*.

IMMUNE TRANSCRIPTOME ANALYSIS

Our transcriptome analysis identified 842 immune-related transcripts in 15 KEGG immune pathways, (Table S3). Many of

Table 2 | KEGG biochemical mappings for rainbow trout.

KEGG categories	No. of annotated sequences (%)	Number of KOs
Metabolism	2236	1496
Carbohydrate metabolism	453 (20.26)	281
Lipid metabolism	349 (15.61)	220
Amino acid metabolism	343 (15.34)	239
Glycan biosynthesis and metabolism	273 (12.21)	187
Nucleotide metabolism	220 (9.84)	152
Energy metabolism	189 (8.45)	132
Metabolism of cofactors and vitamins	129 (5.77)	99
Metabolism of other amino acids	105 (4.70)	63
Xenobiotic biodegradation and metabolism	99 (4.43)	69
Biosynthesis of secondary metabolites	38 (1.70)	25
Metabolism of terpenoids and polyketides	38 (1.70)	29
Genetic information processing	1394	920
Folding, sorting and degradation	484 (34.72)	311
Translation	459 (32.93)	283
Replication and repair	255 (18.29)	196
Transcription	196 (14.06)	130
Environmental information processing	1830	1169
Signal transduction	1508 (82.40)	909
Signaling molecules and interaction	283 (15.46)	227
Membrane transport	39 (2.13)	33
Cellular processes	1421	905
Cell growth and death	485 (34.13)	306
Transport and catabolism	434 (30.54)	295
Cell communication	358 (25.19)	215
Cell motility	144 (10.13)	89
Organismal systems	2825	1777
Immune system	842 (29.81)	541
Endocrine system	619 (21.91)	378
Nervous system	526 (18.62)	334
Digestive system	268 (9.49)	166
Development	210 (7.43)	132
Excretory system	126 (4.46)	73
Circulatory system	103 (3.65)	69
Environmental adaptation	88 (3.12)	56
Sensory system	43 (1.52)	28
Total	9706	6267

these transcripts are represented by complete cDNA sequences that were identified for the first time in rainbow trout. The immune-related transcripts were mapped to a newly assembled genome reference (Berthelot et al., 2014). The coordinate genome reference IDs and complete/incomplete ORF conditions are provided in Tables S4–S10. The immune-related transcripts were clustered according to their KEGG assigned pathways (Figure 5).

Toll-like receptor signaling pathway

Toll-like receptors (TLRs) activate the innate immune response through recognition of pathogen associated molecular patterns (PAMPs) including lipopolysaccharides or peptidoglycan in bacterial cell wall, β -1,3-glucan on fungal cell wall and dsRNA from viruses (Medzhitov and Janeway, 2000; Janeway and Medzhitov, 2002). This leads to the activation of nuclear factor- κ B (NF- κ B)

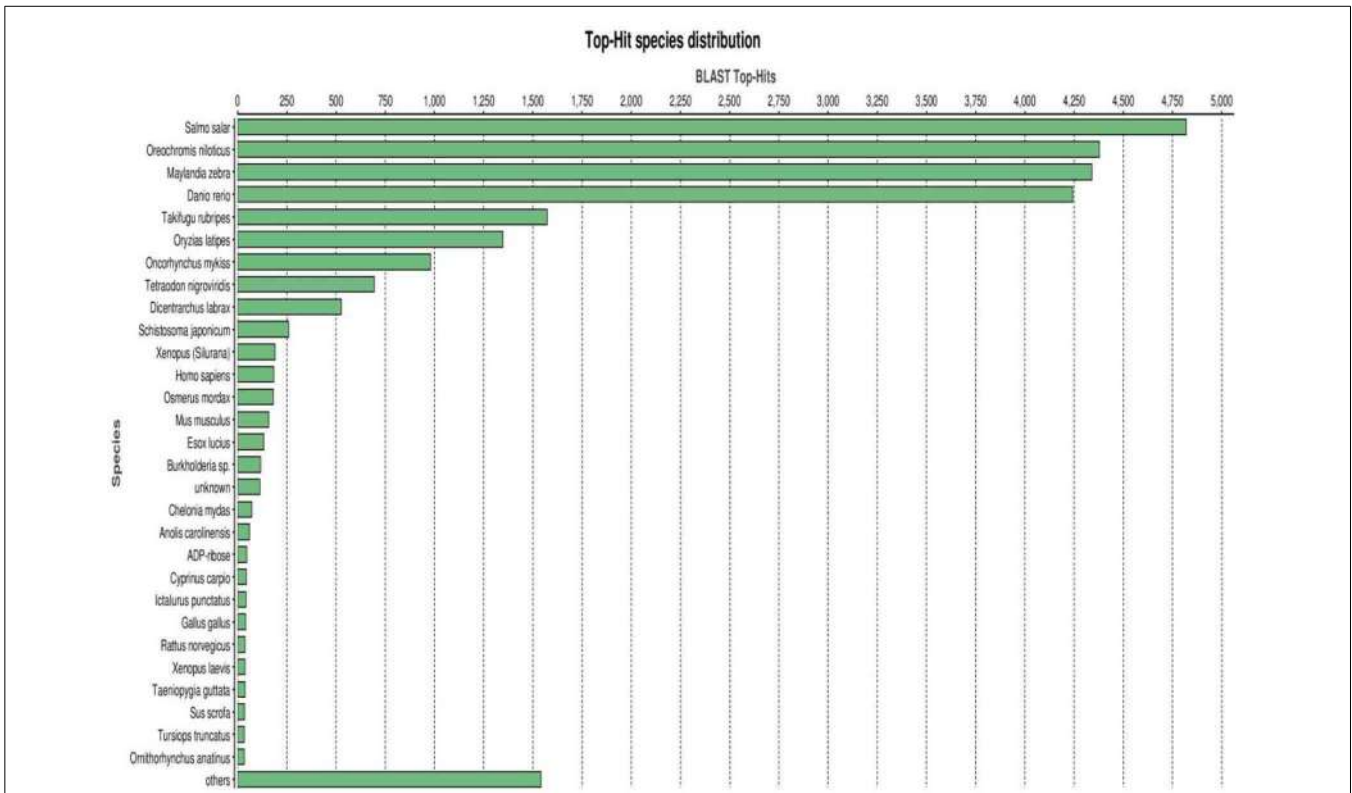


FIGURE 4 | BLASTX top-hit species distribution of gene annotations showing high homology to fish species with known genome sequences.

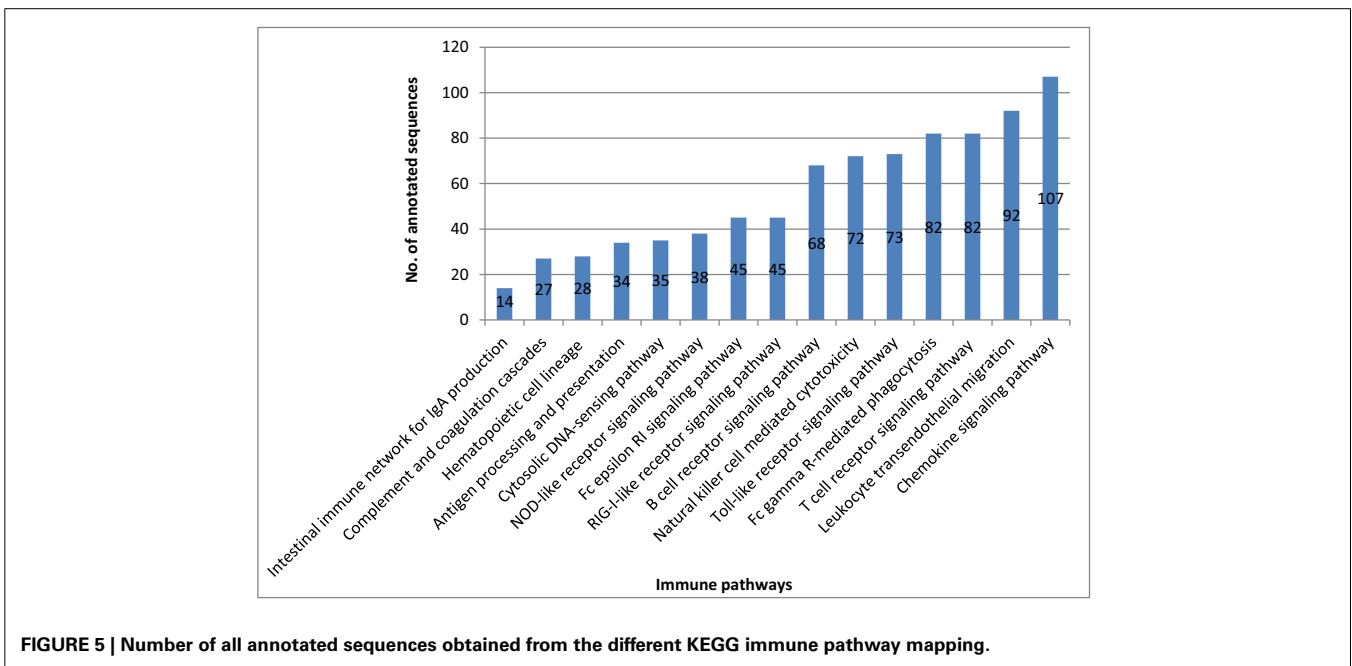


FIGURE 5 | Number of all annotated sequences obtained from the different KEGG immune pathway mapping.

that in turn induces proinflammatory cytokines (Barton and Medzhitov, 2003; Takeda and Akira, 2004).

More than 10 TLRs have been identified in teleost fish including zebrafish (Jault et al., 2004; Meijer et al., 2004), rainbow

trout (Tsujita et al., 2004; Palti et al., 2010a), common carp (Kongchum et al., 2010), pufferfish (Oshiumi et al., 2003), channel catfish (Bilodeau and Waldbieser, 2005; Baoprasertkul et al., 2007a,b) and Atlantic salmon (Rebl et al., 2007). In our spleen

transcriptomic data, there were 7 transcripts, as represented in Table S4, showing high similarity to TLR1, TLR2, TLR3, TLR5, TLR7/8, and TLR9. The coordinate genome reference IDs were identified and five transcripts matching TLR1, TLR3, TLR7, TLR8, and TLR9 had complete cDNA sequences. Several TLRs were previously identified in rainbow trout (Tsujita et al., 2004; Rodriguez et al., 2005; Ortega-Villaizan et al., 2009; Palti et al., 2010a,b). TLR4 was not reported in teleosts except zebrafish whereas TLR6 is totally missing in teleost fish (Takano et al., 2011). This study supports the notion of absence of both TLR4 and TLR6 in rainbow trout.

There were three transcripts matching the NF-κB complex. Two transcripts matched each of MKK4/7 and IFN-αβR complexes. Each of MKK3/6 and MEK1/2 complexes has one transcript whereas MAP2K3 and MAP2K1 have no matches. Additionally, AP-1 which is composed of JUN and FOS (Zenz et al., 2008) had only one transcript matching JUN. The remaining 56 transcripts, out of the 73 total transcripts, showed high similarity to other members of the TLR signaling pathway of higher vertebrates (Figure 6). A total of 38 transcripts have complete cDNA sequences. To our knowledge, 26 different proteins were annotated for the first time in rainbow trout in the current study. Information about transcripts that showed homology to molecules involved in Toll-like receptor signaling pathway is included in additional Table S4.

B cell receptor signaling pathway

B-lymphocytes are involved in antigen specific defense. They are activated through binding of antigen to B cell receptors. These cells produce specific antibodies to neutralize the foreign particles

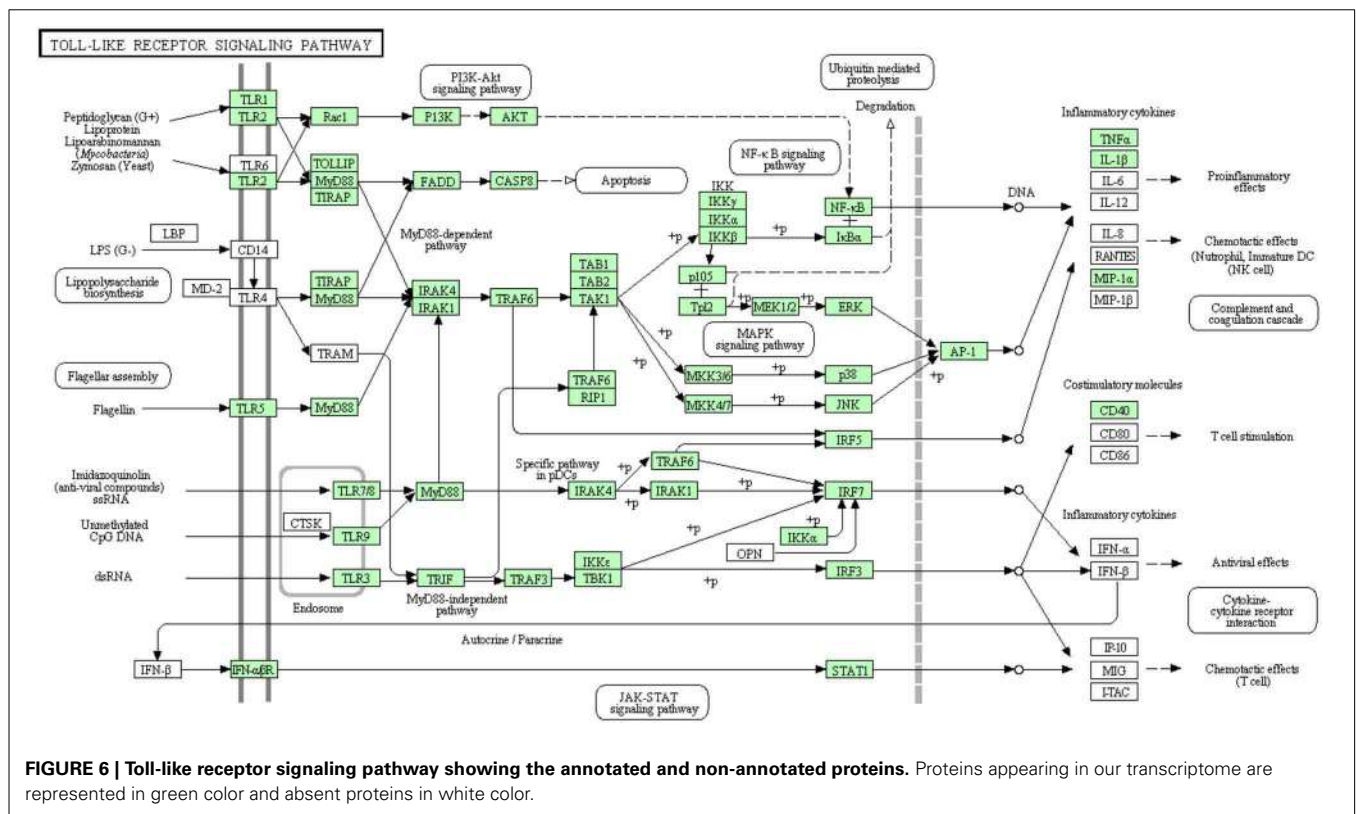
(Kurosaki et al., 2010). The binding of antigen to B-cell receptor activates B lymphocytes (Batista and Neuberger, 1998). Figure 7 shows all B cell signaling pathway annotated and non-annotated proteins in the current study. A total of 68 sequences were assigned to the B cell signaling pathway, 44 have been identified for the first time in this transcriptomic study. Information about transcripts that showed homology to molecules involved in B cell receptor signaling pathway is included in additional Table S5.

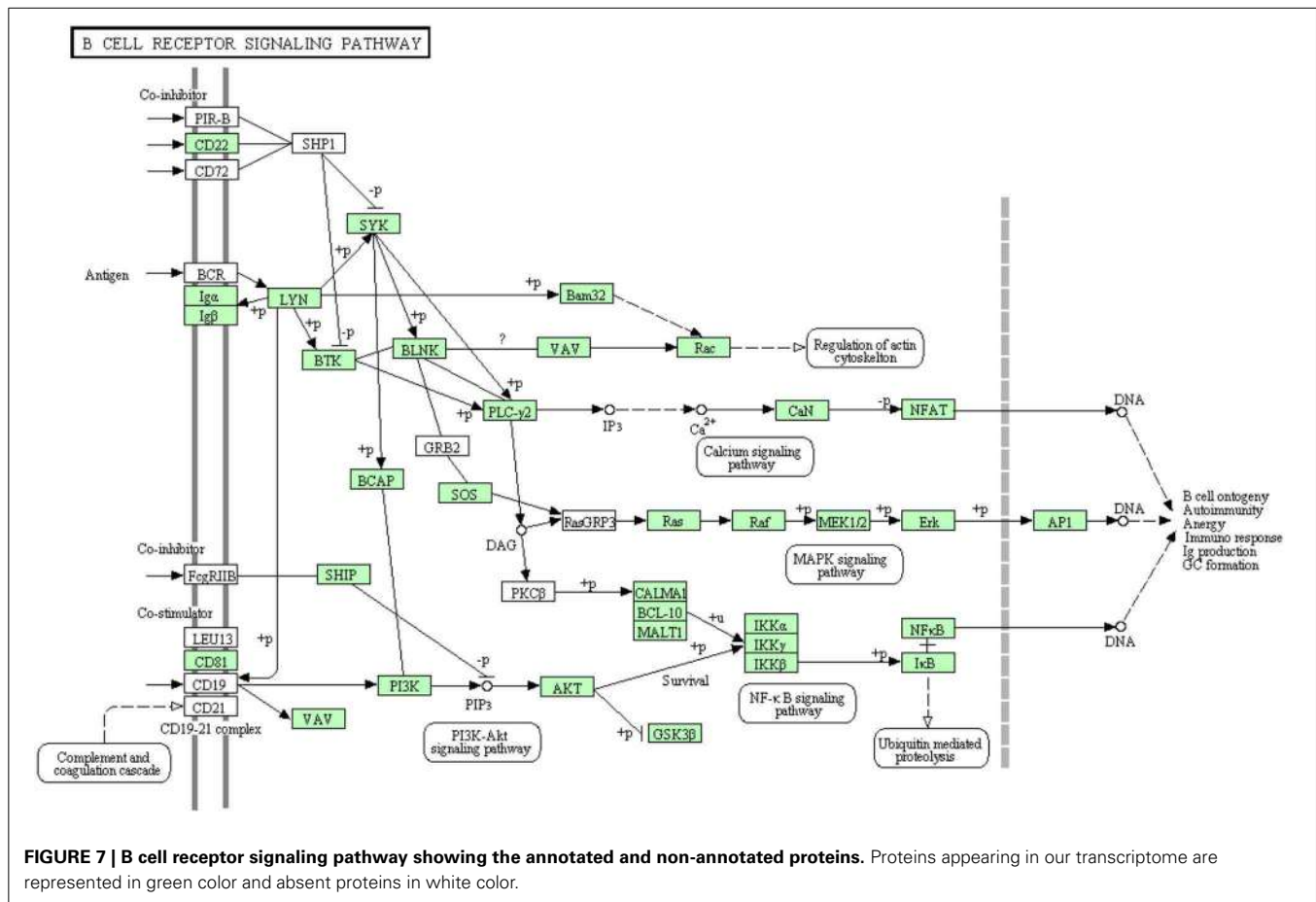
T cell receptor signaling pathway

Like B cells, T lymphocytes are involved in the antigen specific defense. Both T cell receptors (TCR) and costimulatory molecules such as CD28 are required for T cell activation. The cytotoxicity of Cytotoxic T lymphocytes in fish is obscure due to lack of suitable experimental systems even though few studies have depicted the lysis of virus-infected cells by NK-like cells in rainbow trout (Moody et al., 1985; Yoshinaga et al., 1994) and channel catfish (Hogan et al., 1996). Information about proteins that are involved in this cascade was very limited in rainbow trout. The annotated transcripts showed high similarity to many members of the T cell receptor signaling pathway of higher vertebrates as shown in Figure 8. In this study, many transcripts (56 out of 82) that are included in T cell receptor signaling pathway were identified for the first time. Information about transcripts that showed homology to molecules involved in T cell receptor signaling pathway is included in additional Table S6.

Chemokine Signaling Pathway

Chemokines have a major role in trafficking and activation of leukocytes toward the site of inflammation by the aid of





C-terminal domain of their receptors (chemotaxis) (Pease and Williams, 2006). Few chemokine-related genes have been cloned in rainbow trout including CCR9 (Dixon et al., 2013), CXCR4, CCR7 (Daniels et al., 1999) and CXCL14 (Bobe et al., 2006). Many of chemokine receptors haven't been reported to date in rainbow trout (Dixon et al., 2013). In the present study, most of the proteins in the chemokine signaling pathway have been identified (Figure 9). Out of 107 annotated transcripts, 73 sequences matching 49 proteins have been identified for the first time in rainbow trout. Information about transcripts that showed homology to molecules involved in chemokine signaling pathway was included in additional Table S7.

Fc gamma R-mediated phagocytosis

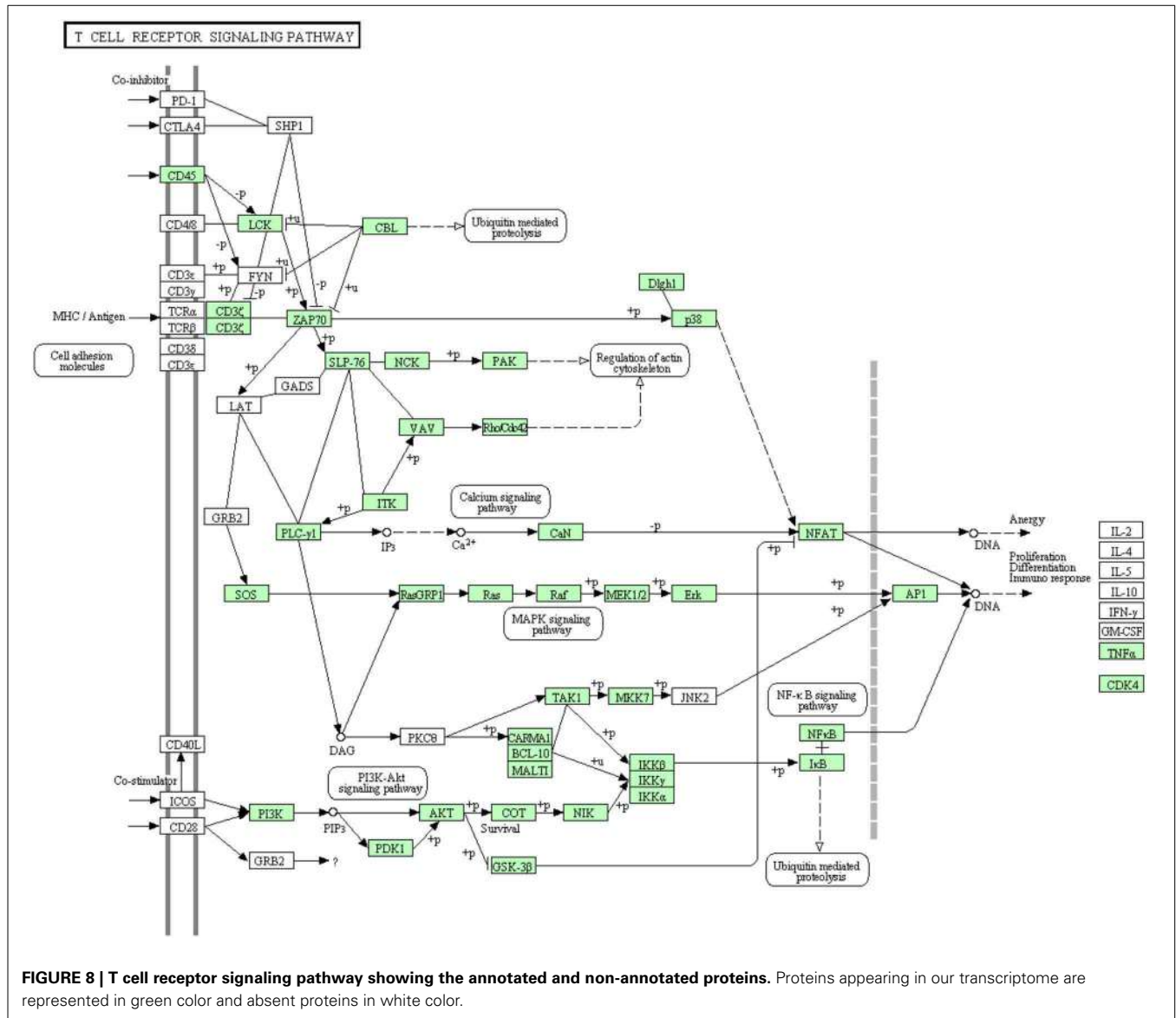
Clusters of IgG coat the foreign particles in a process termed opsonization. Leukocytes and tissue macrophages phagocytose the opsonized pathogens through Fc gamma receptors (Pacheco et al., 2013). Before the present study, some proteins were known to be involved in Fc gamma R-mediated phagocytosis as listed in Table S8. In the current transcriptome analysis, all annotated proteins belonging to the Fc gamma R-mediated phagocytosis are shown in Figure 10. There were 52 sequences out of 82 annotated sequences matching 30 proteins identified for the first time in rainbow trout. Information about transcripts that showed homology to molecules involved in the Fc gamma R-mediated phagocytosis signaling pathway is included in additional Table S8.

Leukocyte transendothelial migration

White blood cells migrate in an amoeboid fashion through the endothelium lining of the blood vessels to drive the immune response to the site of infection (Muller, 2011, 2013). Most of the proteins belonging to this pathway have been identified in this transcriptome analysis (Figure 11). There were 92 transcripts showing high similarity to members of the leukocyte transendothelial migration cascade of higher vertebrates. Prior to the current transcriptome sequencing, some proteins were previously annotated in rainbow trout (Table S9). To our knowledge, 36 proteins belonging to this pathway were reported for the first time in the current study. Information about transcripts that showed homology to molecules involved in leukocyte transendothelial migration is included in additional Table S9.

Natural killer (NK) cell mediated cytotoxicity

NK cells are lymphocytes working as a part of the innate immune system. Although they don't have classical antigen receptors like T and B lymphocytes, their receptors can discriminate between self and non-self-cells (Lanier, 2003). In the current transcriptome analysis, many but not all proteins that are involved in the NK cell mediated cytotoxicity pathway were annotated (Figure 12). Many proteins involved in this pathway were reported before the current study. In this cascade, 42 transcripts out of 72 identified sequences have been annotated for the first time in rainbow trout. The newly annotated transcripts matched 24 proteins. Information

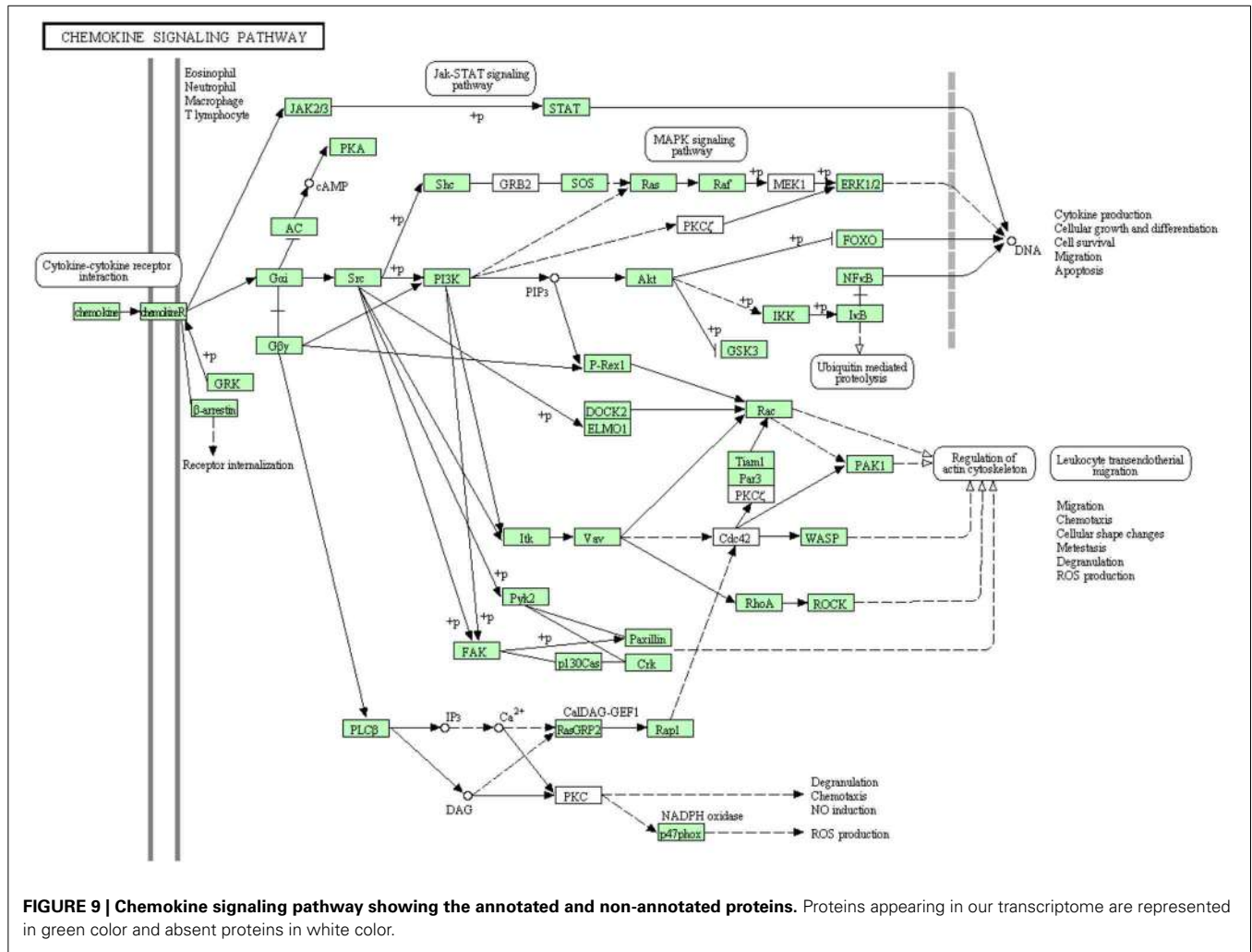


about transcripts that showed homology to molecules involved in NK cell mediated cytotoxicity signaling pathway is included in additional Table S10.

SPLEEN-SPECIFIC GENES

Recently, a total of 51 spleen-specific have been identified in our lab (data will be published elsewhere). The assembled contigs were submitted to the NAGRP Aquaculture Genome Projects (<http://www.animalgenome.org/repository/pub/MTSU2014.0811/>). The coordinate IDs of the spleen-specific transcripts were determined using BLASTN (cut off *E*-value of 1.00E-10) against our spleen transcriptome (Table S11). As shown in Table S11, the level of gene expression was at least 20 fold higher in the spleen compared to 12 other tissues, with statistical false discovery rate (FDR) less than 5%. The spleen-specific genes were mapped to the newly assembled genome reference (Berthelot et al., 2014). The coordinate genome reference IDs and complete/incomplete ORF conditions are provided in additional Table S11.

The list of spleen-specific genes includes: (1) Immune proteins include 5 transcripts such as Fc receptor-like protein 3-like, thrombopoietin receptor precursor, P-selectin precursor, nuclear factor, interleukin 3 regulated (NFIL3), and lectin precursor. (2) Respiratory gas transport proteins include 4 most highly expressed transcripts assigned to hemoglobin and one transcript assigned to carbonic anhydrase II. A large representation of iron-binding proteins was reported in spleen of virus infected Turbot (Pereiro et al., 2012). (3) Coagulation cascade and adhesion proteins include 10 transcripts; five out of the ten transcripts were assigned to platelet glycoproteins whilst the other 5 transcripts were assigned to coagulation factor XIII A chain precursor, thrombopoietin receptor precursor, integrin alpha 2b, integrin beta-3-like and Von Willebrand factor. Thrombocytes appear in spleen during the first week postfertilization before appearing in blood because they participate in body defense (Tavares-Dias and Oliveira, 2009). (4) Development proteins involve 5 transcripts including ectodysplasin receptor, T-box



transcription factor TBX6, homeobox proteins Nkx-2.6-like, T-cell leukemia homeobox protein 1 and Zinc finger protein Gfi1b. The participation of the house finch (*H. mexicanus*) spleen transcriptome in neurodevelopment through a subset of genes has been already reported (Backström et al., 2013). (5) Transporter proteins include one transcript for band 3 anion exchange protein. Both of the biosynthetic and membrane incorporation processes of Band 3 protein have been studied *in vivo* in erythroid spleen cells (Sabban et al., 1981). Among the other identified spleen-specific genes was 5-aminolevulinic synthase erythroid specific mitochondrial precursor which is necessary for heme biosynthesis. Other spleen-specific genes such as rhomboid-like protease 4, Methionine aminopeptidase 2, zinc finger protein 143-like, GATA binding factor 1, N-acetyltransferase 6-like, RING finger protein 151 and cytosolic purine 5'-nucleotidase were also expressed. In mouse and rat, 39 spleen-specific genes are found in tissue-specific database (Xiao et al., 2010). Moreover, 168 Refseq are preferentially expressed in human spleen based on ESTs (Liu et al., 2008). Further work is still needed to validate spleen-specific genes obtained from our high-throughput spleen transcriptome analysis.

Next to spleen, the spleen-specific genes showed relatively higher expression in kidney and fat compared to the rest of the tissues. The mean RPKM values of the spleen-specific genes were 5491, 973, and 400 in spleen, kidney and fat, respectively (**Figure 13**). Kidney is one of the large lymphoid organs in teleosts containing macrophages and lymphoid cells (Zapata et al., 2006; Uribe et al., 2011). The relatively high level of expression of the spleen-specific genes in adipose tissue may be attributed to presence of many populations of immune cells in fat tissues (Ferrante, 2013).

Most of the 51 spleen-specific genes have functions not related to immunity. Interestingly, upon testing a few of these genes after infection with flavobacterium, three transcripts showed differential expression associated with infection, these are contig C23964_c2_seq1_P-selectin_precursor, contig C13_c2_seq1_Hemoglobin_subunit_beta-1 and contig C83628_c0_seq1_RING_finger_protein_151 (Detailed data will be published elsewhere). These data indicate that many of the spleen-specific genes may have immune functions. Further research is still needed to test if the other spleen-specific gene functions are related to immunity.

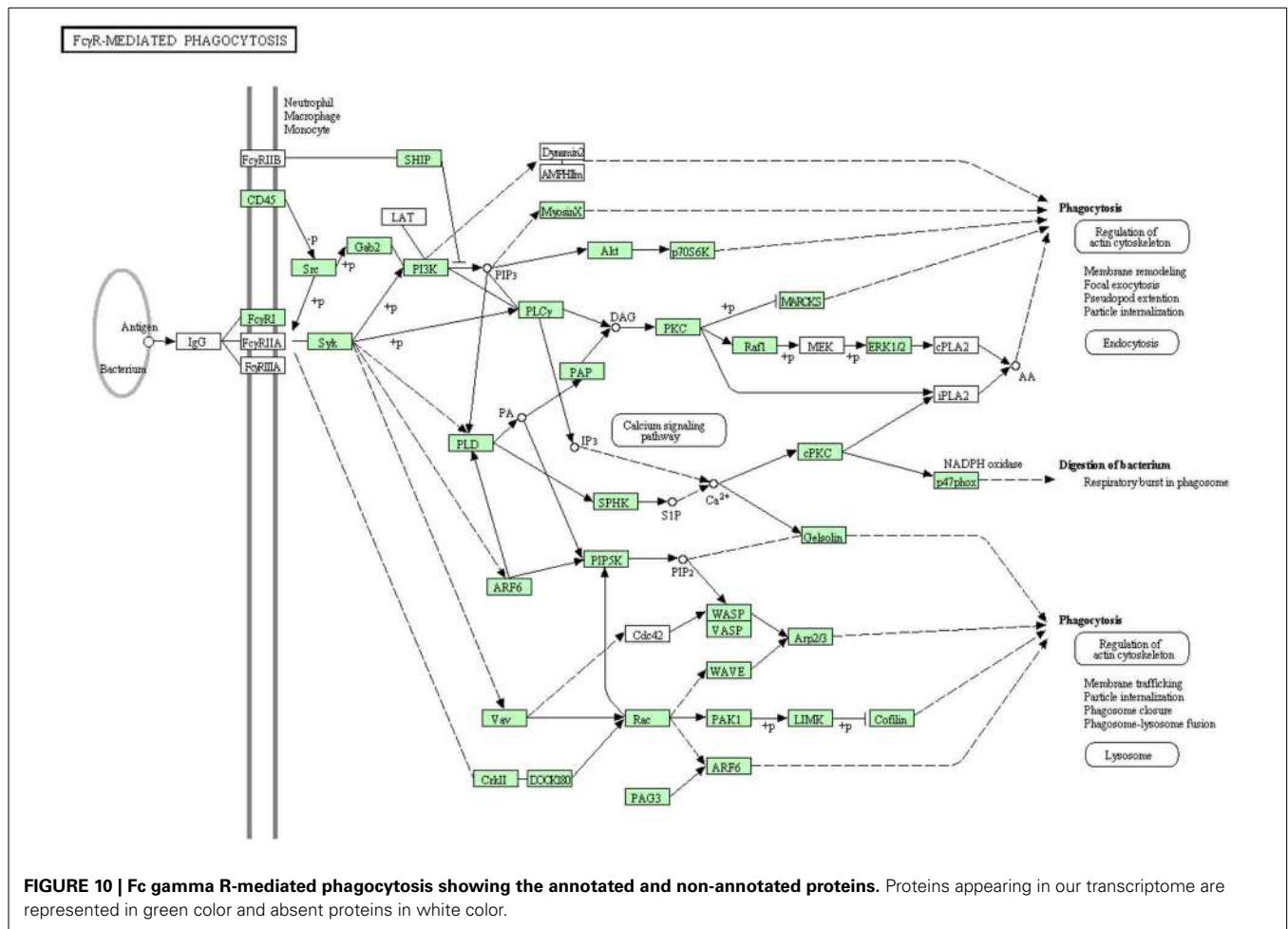


FIGURE 10 | Fc gamma R-mediated phagocytosis showing the annotated and non-annotated proteins. Proteins appearing in our transcriptome are represented in green color and absent proteins in white color.

ORF/ FULL-LENGTH cDNA PREDICTION

Full-length cDNAs are a crucial tool for many genetic and genomic studies including alternative splicing and characterization of gene duplications or pseudogenes (Xin et al., 2008). To identify full-length cDNAs in the above mentioned seven immune pathways, contigs were analyzed by the TargetIdentifier server (Min et al., 2005). A total of 38, 29, 30, 49, 31, 37, and 26 full-length cDNAs were identified in the Toll-like receptor signaling pathway, B cell receptor signaling pathway, T cell receptor signaling pathway, chemokine signaling pathway, Fc gamma R-mediated phagocytosis, Leukocyte transendothelial migration and NK cell mediated cytotoxicity, respectively (Tables S4–S10). Out of the total number of full-length cDNAs, there were 25, 15, 18, 30, 24, 24, and 19 sequences with completely sequenced ORF identified in all studied immune pathways, respectively. Likewise, spleen-specific genes were analyzed by the online TargetIdentifier server (Min et al., 2005) to identify full-length cDNAs. A total of 37 full-length sequences including 24 with completely sequenced ORF were identified among the spleen-specific transcripts. Many of these transcripts were annotated for the first time in rainbow trout. Further work is needed to validate these full-length cDNAs and examine their genomic characteristics (UTR length, Kozak sequence, and conserved motifs) in detail.

METHODS

TISSUE SAMPLING AND RNA ISOLATION

Homozygous doubled haploid rainbow trout fish from the Swanson clonal line were produced at Washington State University (WSU) by androgenesis (Scheerer et al., 1986, 1991; Young et al., 1996; Robison et al., 1999). The fish, approximately 300 g in weight, was reared in recirculating water systems at 12–16°C and had sexually matured as a male prior to sampling. Spleen tissues were collected and frozen in liquid nitrogen, then shipped on ice. All samples were preserved at –80°C until RNA isolation to reduce autocatalytic degradation. Total RNA was isolated from spleen tissues with TRIzol (Invitrogen, Carlsbad, CA) and purified according to the manufacturer’s guidelines. Quantity of total RNA was assessed by measuring the absorbance at A260/A280 using a Nanodrop™ ND-1000 spectrophotometer (Thermo Scientific). RNA quality was checked by electrophoresis through a 1% (w/v) agarose gel. Moreover, RNA integrity was tested using the bioanalyzer 2100 (Agilent, CA).

cDNA LIBRARY PREPARATION AND ILLUMINA SEQUENCING

RNA-Seq library preparation and sequencing were carried out by the University of Illinois at Urbana-Champaign (UIUC), 901 West Illinois street Urbana, IL 61801 USA. A RiboMinus™

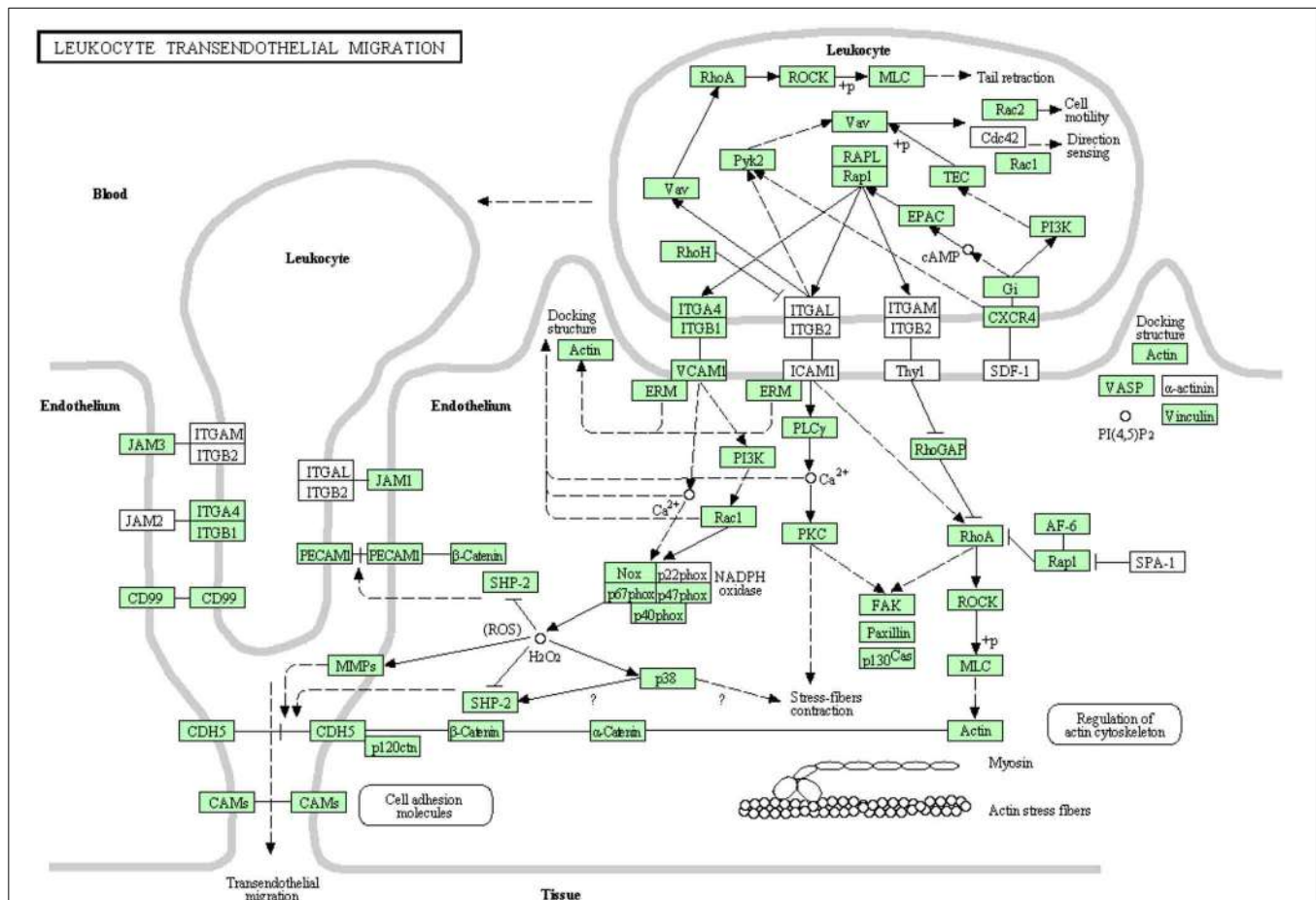


FIGURE 11 | Leukocyte transendothelial migration showing the annotated and non-annotated proteins. Proteins appearing in our transcriptome are represented in green color and absent proteins in white color.

Eukaryote Kit V2 (Invitrogen, Carlsbad, CA) was used to deplete rRNA from the total RNA. cDNA libraries were constructed using ~1 µg of rRNA depleted RNA following the protocol of the Illumina TruSeq RNA sample preparation Kit (Illumina). The resulting double-stranded manufactured cDNA was used in the preparation of the Illumina library. The standard end-repair step was carried out first, followed by the standard ligation reaction where the end-repaired DNA along with a single A base overhang were ligated to the adaptors using T4-DNA Ligase (TrueSeq RNA Sample Prep Kit v2, Illumina, San Diego, CA). The products of the ligation reaction were purified and exposed to size selection of the target length (400–450 bp) from the gel for carrying out the ligation-mediated PCR. Cluster generation and sequencing were carried out following the cluster generation and sequencing manual from Illumina (Cluster Station User Guide and Genome Analyzer Operations Guide). All sequenced raw data were first exported in FASTQ format and are currently being uploaded to the NCBI short read archive (SRA).

CLC GENOMICS DE NOVO ASSEMBLY

De novo assembly of the expressed short reads was carried out by CLC Genomics Workbench (version 6.0; CLC bio, Aarhus,

Denmark; <http://www.clcbio.com/products/clc-genomics-workbench/>). The raw data were filtered by removing short, duplicated and low quality reads. CLC was run using the default settings for all parameters including a minimum contig length of 500 bp.

FUNCTIONAL ANNOTATION AND GENE ONTOLOGY ANALYSIS

Blast2Go version 2.6.5 (<http://www.blast2go.com/b2ghome>) was used for the functional annotation and analysis of the assembled contigs according to molecular function, biological process and cellular component ontologies. BLASTX search for sequence homology (*E*-value of 1.0E-3, maximum 20 hits) was carried out against NCBI's non-redundant protein database (NR). GO terms related to the established hits were extracted and modulated. The functional annotations were analyzed and statistical analysis of GO distributions was performed.

IDENTIFICATION OF IMMUNE-RELATED PROTEINS

Assembled consensus sequences were uploaded to the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007) Ver. 1.67x (<http://www.genome.jp/tools/kaas/>). The functional annotation of genes was carried out by searching local BLAST against KEGG database.

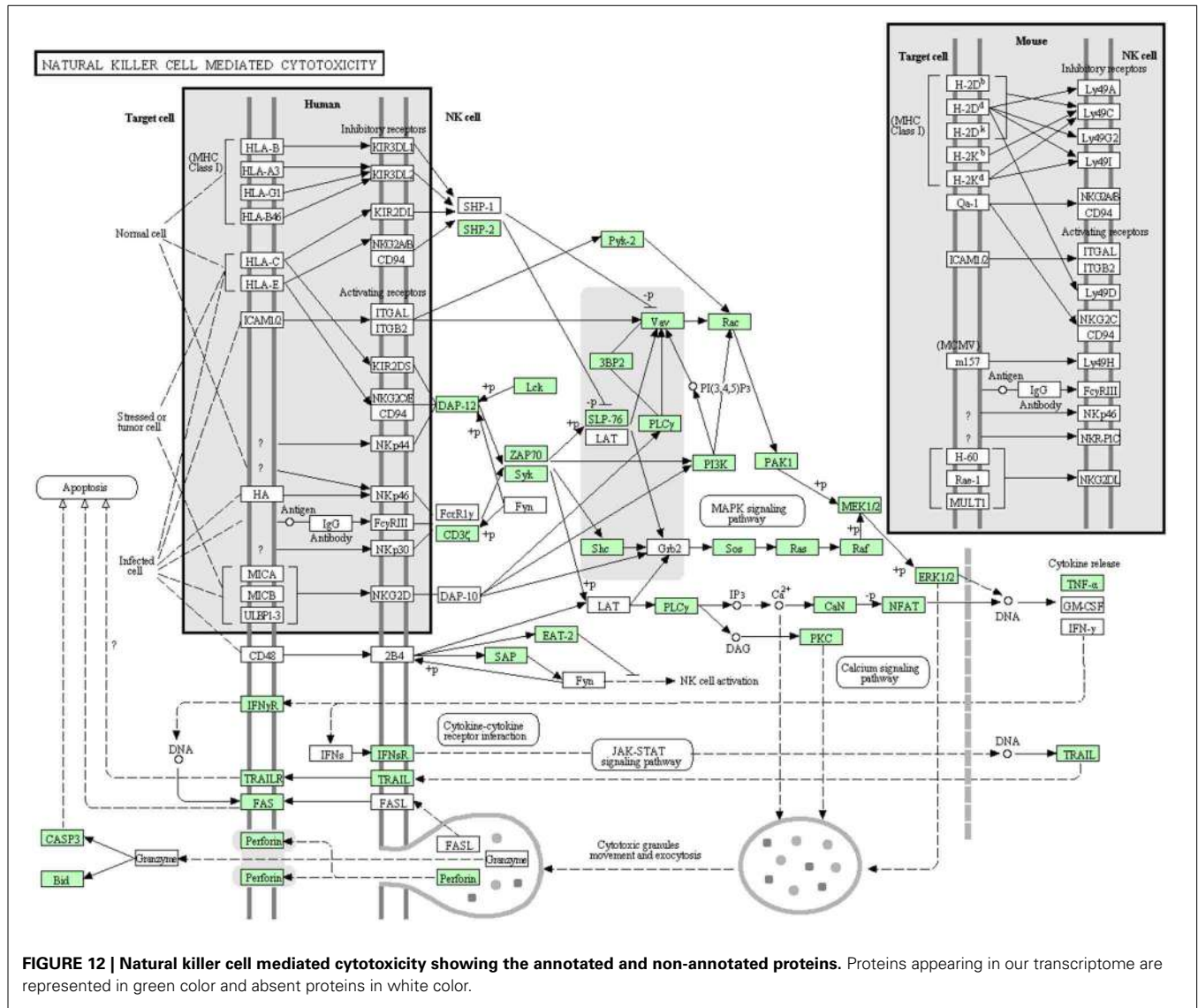


FIGURE 12 | Natural killer cell mediated cytotoxicity showing the annotated and non-annotated proteins. Proteins appearing in our transcriptome are represented in green color and absent proteins in white color.

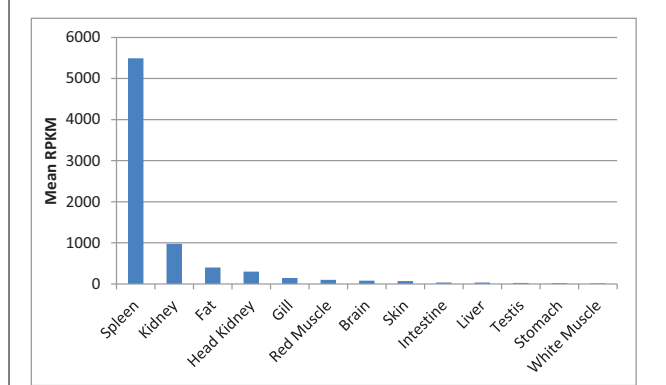


FIGURE 13 | Mean of RPKM of the spleen-specific genes in 13 tissues of rainbow trout. The mean RPKM values of the spleen-specific genes were 5491, 973, and 400 in spleen, kidney and fat, respectively.

The Bi-directional best hit (BBH) method was used to analyze and identify the immune molecules that were present and absent in the seven immune pathways containing the highest number of transcripts that showed high similarity to different members of each pathway. Transcripts were mapped to a newly assembled genome reference. The coordinate genome reference IDs of the immune-related transcripts were determined using BLASTX (cut off *E*-value of 1.00E-10) against the genome protein dataset (Berthelot et al., 2014).

IDENTIFICATION OF SPLEEN-SPECIFIC GENES

Tissue-specific genes in spleen were identified using CLC genomics workbench Baggeley’s test in which expression level of a gene in spleen was compared to its expression level in 12 tissues (brain, white muscle, red muscle, fat, gill, head kidney, kidney, intestine, skin, stomach, liver and testis). For distinction of tissue-specific genes, FDR value was set as 5% and fold change

was at least 20 fold higher in spleen relative to the other 12 tissues. The spleen-specific genes were mapped to the newly assembled genome reference (Berthelot et al., 2014) to determine the coordinate genome reference IDs with a cutoff *E*-value of 1.00E-10.

ORF/ FULL-LENGTH cDNA PREDICTION

All contigs annotated in the interesting KEGG immune pathways, including Toll-like receptor signaling pathway, T-cell receptor signaling pathway, B-cell receptor signaling pathway, chemokine signaling pathway, Fc gamma R-mediated phagocytosis, Leukocyte transendothelial migration and NK cell mediated cytotoxicity, in addition to spleen-specific genes were analyzed using the online TargetIdentifier server (Min et al., 2005) to look for open reading frames and putative full-length cDNAs. A BLASTX output file including the BLASTX results for all cDNA sequences in the "FASTA" file with a cutoff *E*-value of 1.00E-3 was uploaded to the TargetIdentifier program to work properly. cDNA was considered as full-length if the sequence has a 5' stop codon followed by a start codon or the sequence does not have a 5' stop codon but there is an in-frame start codon present prior to the 10th codon of the subject sequence. Based on the BLASTX results, TargetIdentifier predicts existence of an open reading frame (ORF) completely sequenced or not.

ACKNOWLEDGMENTS

This study was supported by the USDA ARS Cooperative Agreement No. 58-1930-0-059. We thank Paul Wheeler for providing tissues from the Swanson doubled haploid trout. Mention of trade names of commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00348/abstract>

REFERENCES

- Allendorf, F. W., and Thorgaard, G. H. (1984). "Tetraploidy and the evolution of salmonid fishes," in *Evolutionary Genetics of Fishes*, ed B. J. Turner (New York, NY: Plenum Press), 1–53.
- Alvarez-Pellitero, P. (2008). Fish immunity and parasite infections: from innate immunity to immunoprophylactic prospects. *Vet. Immunol. Immunopathol.* 126, 171–198. doi: 10.1016/j.vetimm.2008.07.013
- Amish, S. J., Hohenlohe, P. A., Painter, S., Leary, R. F., Muhlfeld, C., Allendorf, F. W., et al. (2012). Rad sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic snp assays. *Mol. Ecol. Resour.* 12, 653–660. doi: 10.1111/j.1755-0998.2012.03157.x
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Backström, N., Zhang, Q., and Edwards, S. V. (2013). Evidence from a house finch (*Haemorrhous mexicanus*) spleen transcriptome for adaptive evolution and biased gene conversion in passerine birds. *Mol. Biol. Evol.* 30, 1046–1050. doi: 10.1093/molbev/mst033
- Baoprasertkul, P., Peatman, E., Abernathy, J., and Liu, Z. (2007a). Structural characterization and expression analysis of toll-like receptor 2 gene from catfish. *Fish Shellfish Immunol.* 22, 418–426. doi: 10.1016/j.fsi.2006.04.005
- Baoprasertkul, P., Xu, P., Peatman, E., Kucuktas, H., and Liu, Z. (2007b). Divergent toll-like receptors in catfish (*Ictalurus punctatus*): tlr5s, tlr20, tlr21. *Fish Shellfish Immunol.* 23, 1218–1230. doi: 10.1016/j.fsi.2007.06.002
- Barton, G. M., and Medzhitov, R. (2003). Toll-like receptor signaling pathways. *Science* 300, 1524–1525. doi: 10.1126/science.1085536
- Batista, F. D., and Neuberger, M. S. (1998). Affinity dependence of the b cell response to antigen: a threshold, a ceiling, and the importance of off-rate. *Immunity* 8, 751–759. doi: 10.1016/S1074-7613(00)80580-4
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5:3657. doi: 10.1038/ncomms4657
- Bilodeau, A. L., and Waldbieser, G. C. (2005). Activation of tlr3 and tlr5 in channel catfish exposed to virulent edwardsiella ictaluri. *Dev. Comp. Immunol.* 29, 713–721. doi: 10.1016/j.dci.2004.12.002
- Bobe, J., Montfort, J., Nguyen, T., and Fostier, A. (2006). Identification of new participants in the rainbow trout (*Oncorhynchus mykiss*) oocyte maturation and ovulation processes using cdna microarrays. *Reprod. Biol. Endocrinol.* 4:39. doi: 10.1186/1477-7827-4-39
- Boussaha, M., Guyomard, R., Cabau, C., Esquerré, D., and Quillet, E. (2012). Development and characterisation of an expressed sequence tags (est)-derived single nucleotide polymorphisms (snps) resource in rainbow trout. *BMC Genomics* 13:238. doi: 10.1186/1471-2164-13-238
- Cerdà, J., Mercadé, J., Lozano, J. J., Machado, M., Tingaud-Sequeira, A., Astola, A., et al. (2008). Genomic resources for a commercial flatfish, the senegalese sole (*solea senegalensis*): est sequencing, oligo microarray design, and development of the soleamold bioinformatic platform. *BMC Genomics* 9:508. doi: 10.1186/1471-2164-9-508
- Chaves-Pozo, E., Mulero, V., Meseguer, J., and García Ayala, A. (2005). Professional phagocytic granulocytes of the bony fish gilthead seabream display functional adaptation to testicular microenvironment. *J. Leukoc. Biol.* 78, 345–351. doi: 10.1189/jlb.0205120
- Chiu, S. C., Tsao, S. W., Hwang, P. I., Vanisree, S., Chen, Y. A., and Yang, N. S. (2010). Differential functional genomic effects of anti-inflammatory phytochemicals on immune signaling. *BMC Genomics* 11:513. doi: 10.1186/1471-2164-11-513
- Christensen, K. A., Brunelli, J. P., Lambert, M. J., Dekoning, J., Phillips, R. B., and Thorgaard, G. H. (2013). Identification of single nucleotide polymorphisms from the transcriptome of an organism with a whole genome duplication. *BMC Bioinformatics* 14:325. doi: 10.1186/1471-2105-14-325
- Colussi, S., Prearo, M., Bertuzzi, S. A., Scanzio, T., Peletto, S., Favaro, L., et al. (2014). Association of a specific major histocompatibility complex class IIβ single nucleotide polymorphism with resistance to lactococcosis in rainbow trout, *Oncorhynchus mykiss* (walbaum). *J. Fish Dis.* doi: 10.1111/jfd.12193. [Epub ahead of print].
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Consortium, G. O. (2008). The gene ontology project in 2008. *Nucleic Acids Res.* 36, D440–D444. doi: 10.1093/nar/gkm883
- Daniels, G. D., Zou, J., Charlemagne, J., Partula, S., Cunningham, C., and Secombes, C. J. (1999). Cloning of two chemokine receptor homologs (cxcr4 and cc-r7) in rainbow trout *Oncorhynchus mykiss*. *J. Leukoc. Biol.* 65, 684–690.
- Dixon, B., Luque, A., Abós, B., Castro, R., González-Torres, L., and Tafalla, C. (2013). Molecular characterization of three novel chemokine receptors in rainbow trout (*Oncorhynchus mykiss*). *Fish Shellfish Immunol.* 34, 641–651. doi: 10.1016/j.fsi.2012.12.003
- Djari, A., Esquerré, D., Weiss, B., Martins, F., Meersseman, C., Boussaha, M., et al. (2013). Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics* 14:307. doi: 10.1186/1471-2164-14-307
- Espenes, A., Press, C., Dannevig, B., and Landsverk, T. (1995). Immune-complex trapping in the splenic ellipsoids of rainbow trout (*Oncorhynchus mykiss*). *Cell Tissue Res.* 282, 41–48.
- Ferrante, A. W. (2013). The immune cells in adipose tissue. *Diabetes Obes. Metab.* 15(Suppl. 3), 34–38. doi: 10.1111/dom.12154

- Fox, S. E., Christie, M. R., Marine, M., Priest, H. D., Mockler, T. C., and Blouin, M. S. (2014). Sequencing and characterization of the anadromous steelhead (*Oncorhynchus mykiss*) transcriptome. *Mar. Genomics* 15, 13–15. doi: 10.1016/j.margen.2013.12.001
- Götz, S., Garcia-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the blast2go suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from rna-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Guyomard, R., Boussaha, M., Krieg, F., Hervet, C., and Quillet, E. (2012). A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of syntenic among teleosts. *BMC Genet.* 13:15. doi: 10.1186/1471-2156-13-15
- Hogan, R. J., Stuge, T. B., Clem, L. W., Miller, N. W., and Chinchar, V. G. (1996). Anti-viral cytotoxic cells in the channel catfish (*Ictalurus punctatus*). *Dev. Comp. Immunol.* 20, 115–127. doi: 10.1016/0145-305X(95)00043-S
- Horiguchi, K., Yagi, S., Ono, K., Nishiura, Y., Tanaka, M., Ishida, M., et al. (2004). Prolactin gene expression in mouse spleen helper t cells. *J. Endocrinol.* 183, 639–646. doi: 10.1677/joe.1.05744
- Houston, D. D., Elzinga, D. B., Maughan, P. J., Smith, S. M., Kauwe, J. S., Evans, R. P., et al. (2012). Single nucleotide polymorphism discovery in cutthroat trout subspecies using genome reduction, barcoding, and 454 pyro-sequencing. *BMC Genomics* 13:724. doi: 10.1186/1471-2164-13-724
- Huang, Y., Huang, X., Yan, Y., Cai, J., Ouyang, Z., Cui, H., et al. (2011). Transcriptome analysis of orange-spotted grouper (*Epinephelus coioides*) spleen in response to Singapore grouper iridovirus. *BMC Genomics* 12:556. doi: 10.1186/1471-2164-12-556
- Janeway, C. A., and Medzhitov, R. (2002). Innate immune recognition. *Annu. Rev. Immunol.* 20, 197–216. doi: 10.1146/annurev.immunol.20.083001.084359
- Jault, C., Pichon, L., and Chluba, J. (2004). Toll-like receptor gene family and tir-domain adaptors in *Danio rerio*. *Mol. Immunol.* 40, 759–771. doi: 10.1016/j.molimm.2003.10.001
- Kaiser, P., Howell, J., Fife, M., Sadeyen, J. R., Salmon, N., Rothwell, L., et al. (2008). Integrated immunogenomics in the chicken: deciphering the immune response to identify disease resistance genes. *Dev. Biol. (Basel)* 132, 57–66. doi: 10.1159/000317144
- Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109–D114. doi: 10.1093/nar/gkr988
- Khan, W. I., and Ghia, J. E. (2010). Gut hormones: emerging role in immune activation and inflammation. *Clin. Exp. Immunol.* 161, 19–27. doi: 10.1111/j.1365-2249.2010.04150.x
- Köllner, B., Wasserrab, B., Kotterba, G., and Fischer, U. (2002). Evaluation of immune functions of rainbow trout (*Oncorhynchus mykiss*)—how can environmental influences be detected? *Toxicol. Lett.* 131, 83–95. doi: 10.1016/S0378-4274(02)00044-9
- Kongchum, P., Palti, Y., Hallerman, E. M., Hulata, G., and David, L. (2010). Snp discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). *Fish Shellfish Immunol.* 29, 356–361. doi: 10.1016/j.fsi.2010.04.013
- Kruglov, A. A., Grivennikov, S. I., Kuprash, D. V., Winsauer, C., Prepens, S., Seleznik, G. M., et al. (2013). Nonredundant function of soluble *Ita3* produced by innate lymphoid cells in intestinal homeostasis. *Science* 342, 1243–1246. doi: 10.1126/science.1243364
- Kurosaki, T., Shinohara, H., and Baba, Y. (2010). B cell signaling and fate decision. *Annu. Rev. Immunol.* 28, 21–55. doi: 10.1146/annurev.immunol.021908.132541
- Lanier, L. L. (2003). Natural killer cell receptor signaling. *Curr. Opin. Immunol.* 15, 308–314. doi: 10.1016/S0952-7915(03)00039-6
- Lee, A. P., Kerk, S. Y., Tan, Y. Y., Brenner, S., and Venkatesh, B. (2011). Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. *Mol. Biol. Evol.* 28, 1205–1215. doi: 10.1093/molbev/msq304
- Leeds, T. D., Silverstein, J. T., Weber, G. M., Vallejo, R. L., Palti, Y., Rexroad, C. E. 3rd., et al. (2010). Response to selection for bacterial cold water disease resistance in rainbow trout. *J. Anim. Sci.* 88, 1936–1946. doi: 10.2527/jas.2009-2538
- Liu, X., Yu, X., Zack, D. J., Zhu, H., and Qian, J. (2008). Tiger: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* 9:271. doi: 10.1186/1471-2105-9-271
- Mahabady, M., Morovvati, H., Arefi, A., and Karamifar, M. (2012). Anatomical and histomorphological study of spleen and pancreas in barzem (*barbus peccoralis*). *World J. Fish Mar. Sci.* 4, 263–267. doi: 10.5829/idosi.wjfm.2012.04.03.61283
- Medzhitov, R., and Janeway, J. (2000). Innate immune recognition: mechanisms and pathways. *Immunol. Rev.* 173, 89–97. doi: 10.1034/j.1600-065X.2000.917309.x
- Meijer, A. H., Gabby Krens, S. F., Medina Rodriguez, I. A., He, S., Bitter, W., Ewa Snaar-Jagalaska, B., et al. (2004). Expression analysis of the toll-like receptor and tir domain adaptor families of zebrafish. *Mol. Immunol.* 40, 773–783. doi: 10.1016/j.molimm.2003.10.003
- Min, X. J., Butler, G., Storms, R., and Tsang, A. (2005). TargetIdentifier: a web-server for identifying full-length cdnas from est sequences. *Nucleic Acids Res.* 33, W669–W672. doi: 10.1093/nar/gki436
- Moody, C., Serreze, D., and Reno, P. (1985). Non-specific cytotoxicity activity of teleost leukocytes. *Dev. Comp. Immunol.* 9, 51–64. doi: 10.1016/0145-305X(85)90059-X
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). Kaas: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Muller, W. A. (2011). Mechanisms of leukocyte transendothelial migration. *Annu. Rev. Pathol.* 6, 323–344. doi: 10.1146/annurev-pathol-011110-130224
- Muller, W. A. (2013). Getting leukocytes to the site of inflammation. *Vet. Pathol.* 50, 7–22. doi: 10.1177/0300985812469883
- Nie, Q., Sandford, E. E., Zhang, X., Nolan, L. K., and Lamont, S. J. (2012). Deep sequencing-based transcriptome analysis of chicken spleen in response to avian pathogenic *Escherichia coli* (apec) infection. *PLoS ONE* 7:e41645. doi: 10.1371/journal.pone.0041645
- Nya, E. J., and Austin, B. (2011). Development of immunity in rainbow trout (*Oncorhynchus mykiss*, walbaum) to *Aeromonas hydrophila* after the dietary application of garlic. *Fish Shellfish Immunol.* 30, 845–850. doi: 10.1016/j.fsi.2011.01.008
- Ortega-Villaizán, M., Chico, V., Falco, A., Perez, L., Coll, J. M., and Estepa, A. (2009). The rainbow trout *tlr9* gene and its role in the immune responses elicited by a plasmid encoding the glycoprotein g of the viral haemorrhagic septicaemia rhabdovirus (vhsv). *Mol. Immunol.* 46, 1710–1717. doi: 10.1016/j.molimm.2009.02.006
- Oshiumi, H., Tsujita, T., Shida, K., Matsumoto, M., Ikeo, K., and Seya, T. (2003). Prediction of the prototype of the human toll-like receptor gene family from the pufferfish, *Fugu rubripes*, genome. *Immunogenetics* 54, 791–800. doi: 10.1007/s00251-002-0519-8
- Pacheco, P., White, D., and Sulchek, T. (2013). Effects of microparticle size and fc density on macrophage phagocytosis. *PLoS ONE* 8:e60989. doi: 10.1371/journal.pone.0060989
- Palti, Y., Gahr, S. A., Hansen, J. D., and Rexroad, C. E. 3rd. (2004). Characterization of a new bac library for rainbow trout: evidence for multi-locus duplication. *Anim. Genet.* 35, 130–133. doi: 10.1111/j.1365-2052.2004.01112.x
- Palti, Y., Gahr, S. A., Purcell, M. K., Hadidi, S., Rexroad, C. E., and Wiens, G. D. (2010a). Identification, characterization and genetic mapping of *tlr7*, *tlr8a1* and *tlr8a2* genes in rainbow trout (*Oncorhynchus mykiss*). *Dev. Comp. Immunol.* 34, 219–233. doi: 10.1016/j.dci.2009.10.002
- Palti, Y., Gao, G., Miller, M. R., Vallejo, R. L., Wheeler, P. A., Quillet, E., et al. (2014). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Mol. Ecol. Resour.* 14, 588–596. doi: 10.1111/1755-0998.12204
- Palti, Y., Genet, C., Luo, M. C., Charlet, A., Gao, G., Hu, Y., et al. (2011). A first generation integrated map of the rainbow trout genome. *BMC Genomics* 12, 180–180. doi: 10.1186/1471-2164-12-180
- Palti, Y., Rodriguez, M. F., Gahr, S. A., Purcell, M. K., Rexroad, C. E., and Wiens, G. D. (2010b). Identification, characterization and genetic mapping of *tlr1* loci in rainbow trout (*Oncorhynchus mykiss*). *Fish Shellfish Immunol.* 28, 918–926. doi: 10.1016/j.fsi.2010.02.002
- Pease, J. E., and Williams, T. J. (2006). The attraction of chemokines as a target for specific anti-inflammatory therapy. *Br. J. Pharmacol.* 147(Suppl. 1), S212–S221. doi: 10.1038/sj.bjp.0706475

- Pereiro, P., Balseiro, P., Romero, A., Dios, S., Forn-Cuni, G., Fuste, B., et al. (2012). High-throughput sequence analysis of turbot (*scophthalmus maximus*) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes. *PLoS ONE* 7:e35369. doi: 10.1371/journal.pone.0035369
- Raida, M. K., and Buchmann, K. (2009). Innate immune response in rainbow trout (*Oncorhynchus mykiss*) against primary and secondary infections with yersinia ruckeri o1. *Dev. Comp. Immunol.* 33, 35–45. doi: 10.1016/j.dci.2008.07.001
- Ravi, V., and Venkatesh, B. (2008). Rapidly evolving fish genomes and teleost diversity. *Curr. Opin. Genet. Dev.* 18, 544–550. doi: 10.1016/j.gde.2008.11.001
- Rebl, A., Siegl, E., Köllner, B., Fischer, U., and Seyfert, H. M. (2007). Characterization of twin toll-like receptors from rainbow trout (*Oncorhynchus mykiss*): evolutionary relationship and induced expression by aeromonas salmonicida salmonicida. *Dev. Comp. Immunol.* 31, 499–510. doi: 10.1016/j.dci.2006.08.007
- Rexroad, C. E., Lee, Y., Keele, J. W., Karamycheva, S., Brown, G., Koop, B., et al. (2003). Sequence analysis of a rainbow trout cDNA library and creation of a gene index. *Cytogenet. Genome Res.* 102, 347–354. doi: 10.1159/000075773
- Rexroad, C. E., Palti, Y., Gahr, S. A., and Vallejo, R. L. (2008). A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* 9:74. doi: 10.1186/1471-2156-9-74
- Robison, B., Wheeler, P., and Thorgaard, G. (1999). Variation in development rate among clonal lines of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173, 131–141. doi: 10.1016/S0044-8486(98)00481-5
- Rodriguez, M. F., Wiens, G. D., Purcell, M. K., and Palti, Y. (2005). Characterization of toll-like receptor 3 gene in rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics* 57, 510–519. doi: 10.1007/s00251-005-0013-1
- Sabban, E., Marchesi, V., Adesnik, M., and Sabatini, D. D. (1981). Erythrocyte membrane protein band 3: its biosynthesis and incorporation into membranes. *J. Cell Biol.* 91, 637–646. doi: 10.1083/jcb.91.3.637
- Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., et al. (2000). A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155, 1331–1345.
- Salem, M., Kenney, P. B., Rexroad, C. E., and Yao, J. (2008). Development of a 37 k high-density oligonucleotide microarray: a new tool for functional genome research in rainbow trout. *J. Fish Biol.* 72, 2187–2206. doi: 10.1111/j.1095-8649.2008.01860.x
- Salem, M., Rexroad, C. E., Wang, J., Thorgaard, G. H., and Yao, J. (2010). Characterization of the rainbow trout transcriptome using sanger and 454-pyrosequencing approaches. *BMC Genomics* 11:564. doi: 10.1186/1471-2164-11-564
- Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264
- Sánchez, C. C., Smith, T. P., Wiedmann, R. T., Vallejo, R. L., Salem, M., Yao, J., et al. (2009). Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* 10:559. doi: 10.1186/1471-2164-10-559
- Sanchez, C. C., Weber, G. M., Gao, G., Cleveland, B. M., Yao, J., and Rexroad, C. E. 3rd. (2011). Generation of a reference transcriptome for evaluating rainbow trout responses to various stressors. *BMC Genomics* 12:626. doi: 10.1186/1471-2164-12-626
- Scheerer, P., Allendorf, F., and Knudsen, K. (1986). Androgenetic rainbow trout produced from inbred and outbred sperm show similar survival. *Aquaculture* 57, 289–298. doi: 10.1016/0044-8486(86)90207-3
- Scheerer, P., Thorgaard, G., and Allendorf, F. (1991). Genetic analysis of androgenetic rainbow trout. *J. Exp. Zool.* 260, 382–390. doi: 10.1002/jez.1402600312
- Skamene, E., and Pietrangeli, C. E. (1991). Genetics of the immune response to infectious pathogens. *Curr. Opin. Immunol.* 3, 511–517. doi: 10.1016/0952-7915(91)90013-Q
- Takano, T., Kondo, H., Hirono, I., Endo, M., Saito-Taki, T., and Aoki, T. (2011). *Toll-Like Receptors in Teleosts*. Selangor: Fish Health Section, Asian Fisheries Society.
- Takeda, K., and Akira, S. (2004). TLR signaling pathways. *Semin. Immunol.* 16, 3–9. doi: 10.1016/j.smim.2003.10.003
- Tarantino, G., Savastano, S., Capone, D., and Colao, A. (2011). Spleen: a new role for an old player? *World J. Gastroenterol.* 17, 3776–3784. doi: 10.3748/wjg.v17.i33.3776
- Tavares-Dias, M., and Oliveira, S. R. (2009). A review of the blood coagulation system of fish. *Braz. J. Biosci.* 7, 205–224.
- Taylor, E. B., Tamkee, P., Keeley, E. R., and Parkinson, E. A. (2011). Conservation prioritization in widespread species: the use of genetic and morphological data to assess population distinctiveness in rainbow trout (*Oncorhynchus mykiss*) from british columbia, Canada. *Evol. Appl.* 4, 100–115. doi: 10.1111/j.1752-4571.2010.00136.x
- Thorgaard, G. H., Bailey, G. S., Williams, D., Buhler, D. R., Kaattari, S. L., Ristow, S. S., et al. (2002). Status and opportunities for genomics research with rainbow trout. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 133, 609–646. doi: 10.1016/S1096-4959(02)00167-7
- Tilton, S. C., Gerwick, L. G., Hendricks, J. D., Rosato, C. S., Corley-Smith, G., Givan, S. A., et al. (2005). Use of a rainbow trout oligonucleotide microarray to determine transcriptional patterns in aflatoxin b1-induced hepatocellular carcinoma compared to adjacent liver. *Toxicol. Sci.* 88, 319–330. doi: 10.1093/toxsci/kfi309
- Tsujita, T., Tsukada, H., Nakao, M., Oshiumi, H., Matsumoto, M., and Seya, T. (2004). Sensing bacterial flagellin by membrane and soluble orthologs of toll-like receptor 5 in rainbow trout (*Oncorhynchus mykiss*). *J. Biol. Chem.* 279, 48588–48597. doi: 10.1074/jbc.M407634200
- Ullal, A. J., Litaker, R. W., and Noga, E. J. (2008). Antimicrobial peptides derived from hemoglobin are expressed in epithelium of channel catfish (*Ictalurus punctatus*, rafinesque). *Dev. Comp. Immunol.* 32, 1301–1312. doi: 10.1016/j.dci.2008.04.005
- Ungar, G. (1945). Endocrine function of the spleen and its participation in the pituitary-adrenal response to stress. *Endocrinology* 37, 329–340. doi: 10.1210/endo-37-5-329
- Uribe, C., Folch, H., Enriquez, R., and Moran, G. (2011). Innate and adaptive immunity in teleost fish: a review. *Vet. Med.* 56, 486–503.
- Van Muiswinkel, W. B., Wiegertjes, G. F., and Stet, R. J. M. (1999). The influence of environmental and genetic factors on the disease resistance of fish. *Aquaculture* 172, 103–110. doi: 10.1016/S0044-8486(98)00444-X
- Wang, X., Xu, R., Wang, R., and Liu, A. (2012). Transcriptome analysis of *sacha inchi* (*Plukenetia volubilis* L.) seeds at two developmental stages. *BMC Genomics* 13:716. doi: 10.1186/1471-2164-13-716
- Whyte, S. K. (2007). The innate immune response of finfish—a review of current knowledge. *Fish Shellfish Immunol.* 23, 1127–1151. doi: 10.1016/j.fsi.2007.06.005
- Wiegertjes, G. F., Stet, R. J., Parmentier, H. K., and Van Muiswinkel, W. B. (1996). Immunogenetics of disease resistance in fish: a comparative approach. *Dev. Comp. Immunol.* 20, 365–381. doi: 10.1016/S0145-305X(96)0032-8
- Wiens, G. D., Lapatra, S. E., Welch, T. J., Evenhuis, J. P., Rexroad, C. E., and Leeds, T. D. (2013a). On-farm performance of rainbow trout (*Oncorhynchus mykiss*) selectively bred for resistance to bacterial cold water disease: effect of rearing environment on survival phenotype. *Aquaculture* 388–391, 128–136. doi: 10.1016/j.aquaculture.2013.01.018
- Wiens, G. D., Vallejo, R. L., Leeds, T. D., Palti, Y., Hadidi, S., Liu, S., et al. (2013b). Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout: identification of QTL on chromosome omy19. *PLoS ONE* 8:e75749. doi: 10.1371/journal.pone.0075749
- Wong, S., Waldrop, T., Summerfelt, S., Davidson, J., Barrows, F., Kenney, P. B., et al. (2013). Aquacultured rainbow trout (*Oncorhynchus mykiss*) possess a large core intestinal microbiota that is resistant to variation in diet and rearing density. *Appl. Environ. Microbiol.* 79, 4974–4984. doi: 10.1128/AEM.00924-13
- Wu, X. N. (1998). Current concept of spleen-stomach theory and spleen deficiency syndrome in tcm. *World J. Gastroenterol.* 4, 2–6.
- Xiang, L. X., He, D., Dong, W. R., Zhang, Y. W., and Shao, J. Z. (2010). Deep sequencing-based transcriptome profiling analysis of bacteria-challenged lateolabrax japonicus reveals insight into the immune-relevant genes in marine fish. *BMC Genomics* 11:472. doi: 10.1186/1471-2164-11-472
- Xiao, S. J., Zhang, C., Zou, Q., and Ji, Z. L. (2010). TisGED: a database for tissue-specific genes. *Bioinformatics* 26, 1273–1275. doi: 10.1093/bioinformatics/btq109
- Xin, D., Hu, L., and Kong, X. (2008). Alternative promoters influence alternative splicing at the genomic level. *PLoS ONE* 3:e2377. doi: 10.1371/journal.pone.0002377

- Yang, D., Liu, Q., Yang, M., Wu, H., Wang, Q., Xiao, J., et al. (2012). Rna-seq liver transcriptome analysis reveals an activated mhc-i pathway and an inhibited mhc-ii pathway at the early stage of vaccine immunization in zebrafish. *BMC Genomics* 13:319. doi: 10.1186/1471-2164-13-319
- Yoshinaga, K., Okamoto, N., Kurata, O., and Ikeda, Y. (1994). Individual variations of natural killer activity of rainbow trout leukocytes against ipn virus-infected and uninfected rtg-2 cells. *Fish Pathol.* 29, 1–4. doi: 10.3147/jsfp.29.1
- Young, W. P., Wheeler, P. A., Coryell, V. H., Keim, P., and Thorgaard, G. H. (1998). A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics* 148, 839–850.
- Young, W. P., Wheeler, P. A., Fields, R. D., and Thorgaard, G. H. (1996). DNA fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *J. Hered.* 87, 77–80. doi: 10.1093/oxfordjournals.jhered.a022960
- Zapata, A., Chibá, A., and Varas, A. (1997). “Cells and tissues of the immune system of fish,” in *The Fish Immune System: Organism, Pathogen, and Environment*, eds G. Iwama and T. Nakanishi (San Diego, CA: Academic Press), 1–62.
- Zapata, A., Diez, B., Cejalvo, T., Gutiérrez-De Frías, C., and Cortés, A. (2006). Ontogeny of the immune system of fish. *Fish Shellfish Immunol.* 20, 126–136. doi: 10.1016/j.fsi.2004.09.005
- Zenz, R., Eferl, R., Scheinecker, C., Redlich, K., Smolen, J., Schonhaler, H. B., et al. (2008). Activator protein 1 (fos/jun) functions in inflammatory bone and skin disease. *Arthritis Res. Ther.* 10:201. doi: 10.1186/ar2338
- Zhang, J., Yang, Y., Wang, Y., Wang, Z., Yin, M., and Shen, X. (2011). Identification of hub genes related to the recovery phase of irradiation injury by microarray and integrated gene network analysis. *PLoS ONE* 6:e24680. doi: 10.1371/journal.pone.0024680

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 July 2014; accepted: 16 September 2014; published online: 14 October 2014.

Citation: Ali A, Rexroad CE, Thorgaard GH, Yao J and Salem M (2014) Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes. *Front. Genet.* 5:348. doi: 10.3389/fgene.2014.00348

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Ali, Rexroad, Thorgaard, Yao and Salem. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Zebrafish as animal model for aquaculture nutrition research

Pilar E. Ulloa^{1*}, Juan F. Medrano² and Carmen G. Feijoo¹

¹ Departamento de Ciencias Biológicas, Facultad de Ciencias Biológicas, Universidad Andres Bello, Santiago, Chile

² Department of Animal Science, University of California, Davis, Davis, CA, USA

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Nelson Colihueque, Universidad de Los Lagos, Chile
Gen Hua Yue, National University of Singapore, Singapore

*Correspondence:

Pilar E. Ulloa, Departamento de Ciencias Biológicas, Facultad de Ciencias Biológicas, Universidad Andres Bello, República 217, Santiago 8370146, Chile
e-mail: pilarelizabeth@gmail.com

The aquaculture industry continues to promote the diversification of ingredients used in aquafeed in order to achieve a more sustainable aquaculture production system. The evaluation of large numbers of diets in aquaculture species is costly and requires time-consuming trials in some species. In contrast, zebrafish (*Danio rerio*) can solve these drawbacks as an experimental model, and represents an ideal organism to carry out preliminary evaluation of diets. In addition, zebrafish has a sequenced genome allowing the efficient utilization of new technologies, such as RNA-sequencing and genotyping platforms to study the molecular mechanisms that underlie the organism's response to nutrients. Also, biotechnological tools like transgenic lines with fluorescently labeled neutrophils that allow the evaluation of the immune response *in vivo*, are readily available in this species. Thus, zebrafish provides an attractive platform for testing many ingredients to select those with the highest potential of success in aquaculture. In this perspective article aspects related to diet evaluation in which zebrafish can make important contributions to nutritional genomics and nutritional immunity are discussed.

Keywords: zebrafish, aquaculture nutrition, diet evaluation, nutritional genomics, nutritional immunity

AQUACULTURE NUTRITION: TRENDS AND FUTURE PERSPECTIVE

Worldwide fish production is increasing at 8.8% per year (Food and Agriculture Organization of the United Nations [FAO], 2012). However, the decreasing availability of fishmeal because of a reduction in the supply of important sources of fish limits its use as the primary protein in fish diets. In fact, harvest volumes of these species decreased from 10.7 million tons in 2004 to 4.2 million tons in 2010 (Food and Agriculture Organization of the United Nations [FAO], 2012). This outcome, apart to generate a negative ecological impact, has led to an increase in the value of fishmeal, affecting the profitability of many aquaculture enterprises (Rosamond et al., 2000). Therefore, the utilization of plant protein meals has emerged as an alternative to replace of fish meal in aquaculture feeds (Hardy, 2010).

Within the wide range of available vegetable protein sources (peas, lupine, maize, and wheat), soybean has been the most commonly used due to its wide availability in the market, low cost, high content of digestible protein and balanced amino acid profile (Naylor et al., 2009). Increasing dietary levels of soybean meal and other vegetable proteins has been tested in a variety of fish species, with inclusion levels ranging from 20 to 100% of fishmeal replacement. Unfortunately, results have shown several adverse effects such as reduction of growth and intestinal inflammation, even at low levels of inclusion (Gómez-Requeni et al., 2004; Mundheim et al., 2004; Knudsen et al., 2007; Urán et al., 2008).

Different approaches have been taken to make the utilization of plant protein by different fish more efficient, including the improvement of genetic selection in plants aiming to reduce the effect of anti-nutritional components and the stimulation

of the intestinal microbiota of fish (Bakke-McKellep et al., 2007; Froystad-Saugen et al., 2009; Merrifield et al., 2009). Additionally, the diversification of new protein ingredients (from animals and plants), and identification of additives (natural or synthetic) with “intestinal protective” activity to promote growth and health have been made a priority (Refstie et al., 2010; Sicuro et al., 2010; Czubinski et al., 2014). Thus, the effects of nutrition on genomics and immunity are being addressed (Montero et al., 2010; Hernández et al., 2013). The implementation of new technologies such as RNA-sequencing, together with progress in sequencing genomes of different fish, can identify the genes affected by nutrition and also identify the genetic variants that influence the organism's response to nutrients.

To fully understand the repercussions of new diets on fish physiology, a shift in approach is required to determine the molecular and cellular pathways that regulate responses to different diets. The evaluation of a large numbers of diets directly in aquaculture species results in high costs and long-term assays, so it is necessary to implement new strategies in order to accelerate and make this experimental process cost-effective. It is also essential to determine the molecular mechanisms by which physiological processes are perturbed in response to diet. This will provide insights on how to solve existing problems resulting from nutrition interventions in the aquaculture industry.

As an alternative approach to address the aforementioned issues, we propose the use of zebrafish (*Danio rerio*) as an animal model for high throughput testing of experimental diets in aquaculture. This teleost fish has numerous advantages related to its fast life cycle, ease of handling and very well-known biology, besides allowing *in vivo* analysis with large numbers of fish

(Kimmel et al., 1995). Here, we highlight the important contribution that zebrafish can make in the field of nutritional genomics and nutritional immunity. Both lines of investigation provide useful contributions to the evaluation of diets.

ADVANTAGES OF ZEBRAFISH AS AN ANIMAL MODEL FOR AQUACULTURE NUTRITION RESEARCH

Zebrafish is a well-established animal model for a wide range of research areas, from biomedicine to toxicology (Roush, 1996; Bergeron et al., 2008). The use of this model fish for improving production process of aquaculture has emerged as an important research field (Ulloa et al., 2011; Ribas and Piferrer, 2013). In particular, research improving husbandry and survival, immune response, nutrition and growth have been carried out in zebrafish, and are expected to provide results applicable to important commercial fish (Oyarbide et al., 2012; Hedrera et al., 2013; Ulloa et al., 2013).

Among the advantages of this model are its ease of handling in breeding and experimentation, short generation intervals (~3 months) and large numbers of eggs per brood (100–200 eggs/clutch), which allow performing all analyses with a high number of specimens per data point (Kimmel et al., 1995). Embryos hatch at 2 days post-fertilization and larvae can live for 5 days without feeding due to the consumption of the yolk (Lawrence, 2007). During the larval period all organs and systems are functional, making these individuals physiologically equivalent to adult animals. In fact, both larvae and adult zebrafish can eat a wide variety of foods including live feed (*Paramecium* and *Artemia* nauplii) and different commercial fish diets, as well as experimental plant protein-based diets (Hedrera et al., 2013; Ulloa et al., 2013). The availability of a sequenced genome (assembly ZV9) allows evaluating the effect of diet on molecular mechanisms using genomic tools such as RNA-sequencing (RNA-seq; Morozova and Marra, 2008). This technology has been recently used in some aquaculture species and also in zebrafish (Li et al., 2013; Long et al., 2013; Xu et al., 2013; Cui et al., 2014; Liu et al., 2014). On the other hand, a wide diversity of approaches in genetic manipulation are readily available in zebrafish. The availability of a large number of transgenic lines, which carry specific promoters coupled to GFP (green fluorescent protein), is widespread. For example, the use of the Tg(Bacmpx:GFP) line allows to follow specific innate immune cells such as neutrophils *in vivo* due to its fluorescent mark (Renshaw et al., 2006). Moreover, it has been demonstrated that neutrophils output from hematopoietic tissue toward affected territories correlates with pro-inflammatory cytokine production, thus making transgenic lines “live indicators” of inflammatory process (Barros-Becker et al., 2012). All these assays can be carried out with embryos and larvae, which are distributed individually or in small groups in micro well plates in small volumes (0.5–2 ml) allowing sufficient biological replicas in each experiment.

Directly related to the evaluation of diets are two aspects in which zebrafish can make important contributions: nutritional genomics and nutritional immunity.

NUTRITIONAL GENOMICS CONTRIBUTIONS

Nutritional genomics is a discipline that investigates the relationship between the genome and diets. Two approaches are essentially

used: “Nutrigenomics,” which studies how dietary ingredients affect the gene expression and “Nutrigenetics,” which aims to understand how the genetic makeup of an individual coordinates the response to diet (Mutch et al., 2005). Both approaches attempt to clarify the effect of dietary components that contribute to phenotype by altering gene expression and individual genetic variants.

Since one effect of plant diets on fish phenotypes is growth, for more than one decade two approaches have been used to understand the genomics associated with fish growth: (1) global evaluation of genes by the creation of microarray platforms based on EST libraries, and (2) evaluation of candidate genes involved in growth. These studies have generated a list of genes that are over/under expressed in response to vegetable diets in different development stages of fish (Douglas, 2006; Panserat et al., 2008; Von Schalburg et al., 2008; Alami-Durante et al., 2010; Tacchi et al., 2011). However, the results obtained in the many nutritional studies are commonly difficult to compare among each other. This is due to the use of different origins of the same ingredient, feed formulation, genetic background of fish and experimental design. Moreover it is a common practice that, the biological samples used for transcription analysis are randomly selected. This experimental background makes interpretation of data difficult in order to dissect the relationship between genotype and phenotype, as well as the effects of diet on gene expression. Thus, in order develop a better understanding of molecular mechanism modulated by nutrition, it is necessary to select fish according to genotype, phenotype and/or ideally based on genetically improved populations (Kolditz et al., 2008; Morais et al., 2011; Salem et al., 2012).

Advanced technologies such as RNA-seq and genotyping platforms allow accelerated research in the nutritional genomics field that can be projected to aquaculture species (Houston et al., 2014; Qian et al., 2014). These technologies have been widely used to increase the genomic understanding of phenotypes in other livestock species (Wickramasinghe et al., 2014). A recent aquaculture study analyzed genotype-diet interaction in the transcriptome analysis of Atlantic salmon fed vegetable oil. The authors identified metabolic pathways and key regulators that may respond differently to alternative plant-based feeds depending on genotype (Morais et al., 2011). Salem et al. (2012) using RNA-seq, identified 23 single nucleotide polymorphisms (SNPs) markers in rainbow trout associate with growth response to commercial fish meal-based diet. However, despite these efforts, the identification of genetic differences (gene expression and SNPs) among fish in relation to growth rates in response to plant protein diets has not been reported.

In order to address this subject, a new approach using zebrafish was developed (Ulloa et al., 2013). Briefly, a population of 24 experimental families was generated to examine growth response in zebrafish fed with a plant protein-based diet. At 30 days post-fertilization (dpf) each family was split to generate two replicates (40 fish per family replicate), which created two populations of fish with similar genetic backgrounds. The fish were fed from larval transition (35 dpf) to sexual maturity (98 dpf). The first replicate of 24 families was fed a diet containing 100% plant protein as the only protein source (experimental diet) and the second

replicate was fed a diet containing 100% animal protein as the only protein source (control diet). The results showed decreased growth in fish fed a plant protein-based diet compared to fish fed a fish meal-based diet, as has been documented in farmed fish (Gómez-Requeni et al., 2004; Mundheim et al., 2004), and very large growth variations from juvenile to adult stages among families (Ulloa et al., 2013). In order to evaluate the effect of a plant protein-based diet on the expression of growth-related genes in the muscle of zebrafish, individuals from three similar families representative of the mean weight in both populations were selected. To understand the effect of family variation on gene expression, these families were evaluated separately. The results showed that the effect of plant diet and family variation on gene expression were significantly different, and clearly suggested that gene expression is influenced not only by nutrition but also by genetic differences in each family; such as been described by Gjedrem (2000). Thus, it was demonstrated that to understand the effect of diet on transcriptome analysis, it is important to homogenize the phenotype and genetic components to avoid conflicts in the interpretation of results (Ulloa et al., 2013).

To measure gene expression and identify SNPs to evaluate growth associations in zebrafish fed a plant protein diet, a new approach was developed using RNA-seq. Samples of muscle collected from low and high growth fish were analyzed to identify SNP in differentially expressed genes. One hundred twenty-four genes were differentially expressed between phenotypes. From these genes 164 SNP were selected and genotyped in 240 fish samples. Marker-trait analysis revealed five SNP in key genes directly related with growth response (Unpublished data). This study provided new candidate genes associated with growth that could be evaluated in farmed fish through comparative genomics. Additionally, this strategy promises to be useful to identify SNP and characterize individuals with the highest performance for growth in response to a plant protein-based diet.

NUTRITIONAL IMMUNITY CONTRIBUTIONS

Diverse studies in cultured fish have shown that soybean meal induces intestinal inflammation, a pathology called enteritis (Bakke-Mckellep et al., 2000; Knudsen et al., 2007). The hallmark of innate immunity is inflammation, this process is triggered in response to different insults, including pathogens, injury or irritants (Chen and Nuñez, 2010). When inflammation occurs, influx, accumulation, and activation of leukocytes (predominantly neutrophils) are triggered at the affected site during the early stages of the response (Witko-Sarsat et al., 2000). One of the first cytokine to be released when inflammation occurs is the pro-inflammatory cytokines Interleukin-1 β (IL-1 β ; Rodríguez-Tovar et al., 2011). Other essential proteins for neutrophils chemoattraction and migration are the chemokine Cxcl8 and some metalloproteinase enzymes (MMPs). Cxcl8 promotes neutrophils recruitment to the sites of insult; meanwhile MMPs are involved in the degradation of the extracellular matrix in order to promote granulocytes migration (Oliveira et al., 2013). Once neutrophils reach the affected site, they destroy the *insulting agent* through the production of non-specific toxins (Fialkow et al., 2007).

On the opposite side are the anti-inflammatory cytokines, such as transforming growth factor beta (TGF- β) and Interleukin 10

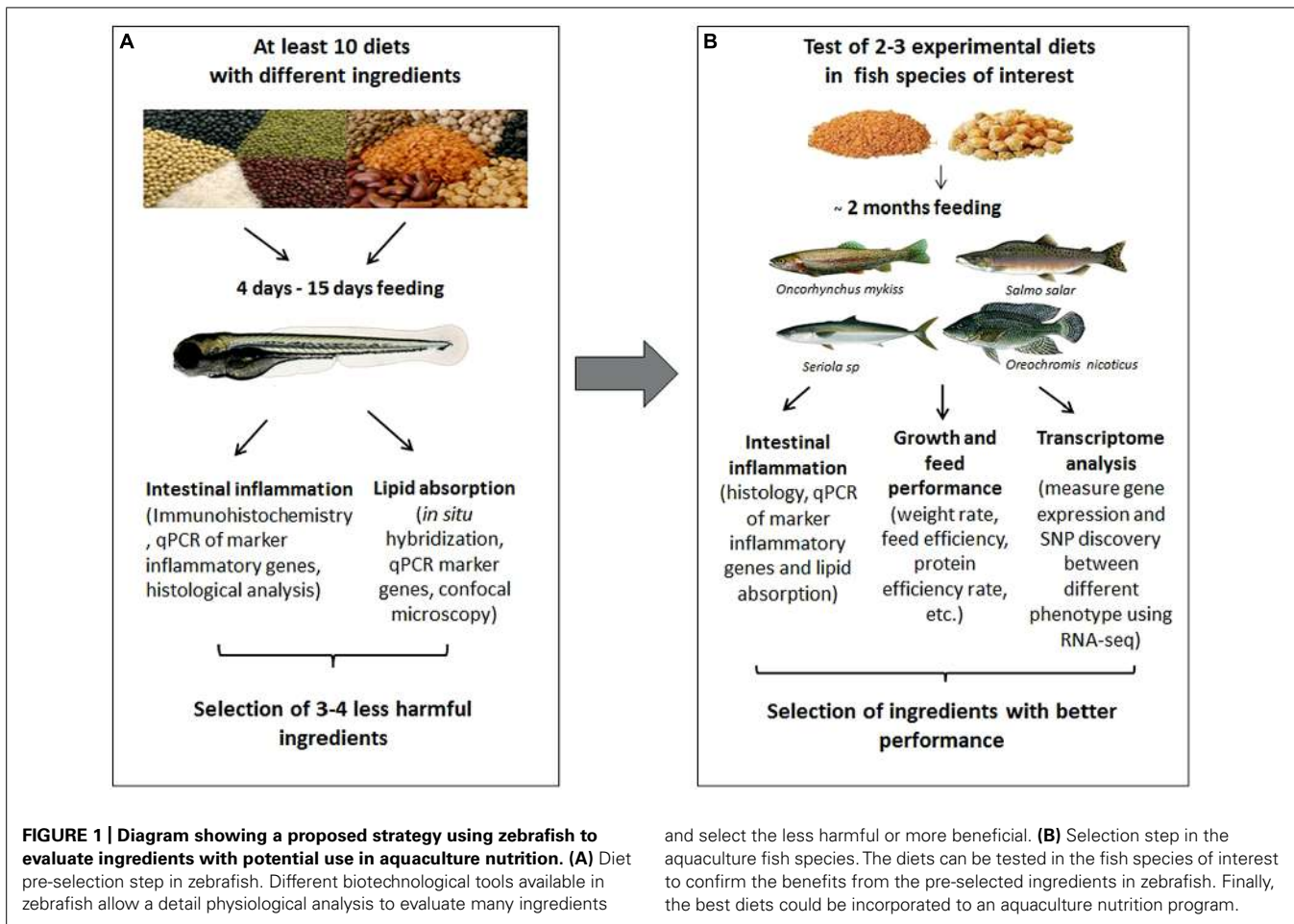
(IL-10), which are mainly secreted by macrophages when the inflammatory agent is removed, promoting the end of the inflammatory process (Engelsma et al., 2002; Ouyang et al., 2011). It is noteworthy that an inflamed intestine of fish is characterized by shorter mucosal folds, loss of vacuole absorptive cells in the intestinal epithelium and large infiltration of neutrophils, macrophages and, eosinophils in the *lamina propria*, among others (Baeverfjord and Krogh, 1996).

The severity of inflammation differs between species and depends on the percentage of plant feeds inclusion to the diets. In salmonids, the most affected by the inclusion of plant protein appears to be Atlantic salmon (*Salmo salar*) and to a lesser extent rainbow trout (*Oncorhynchus mykiss*; Morris et al., 2005; Bakke-McKellep et al., 2007). However, the effect of inflammation has also been described in omnivorous fish such as carp (*Cyprinus carpio*) and zebrafish (Urán et al., 2008; Hedrera et al., 2013). This situation affects the cellular and humoral immunological processes, with negative consequences in food intake and growth (Gómez-Requeni et al., 2004; Mundheim et al., 2004; Montero et al., 2010).

In recent years, additives such as prebiotics mannoooligosaccharides (MOS) and fructooligosaccharides (FOS), probiotics (bacteria); immunostimulants (β -glucans), and nucleotides have been incorporated into fish diets in order to control diseases, improve health, and the immune status against acute stress (Piaget et al., 2007; Staykov et al., 2007; Tahmasebi-Kohyani et al., 2012). In the case of MOS, the supplementation of 0.2% into diets with 14% inclusion of soybean meal decreased the intestinal inflammation in Atlantic salmon (Refstie et al., 2010). In sea bream, the effect of supplementation of 0.4% MOS into diets with 31% inclusion of soybean meal revealed an increase of microvilli density and length of intestinal folds (Dimitroglou et al., 2010). These results showed that MOS have a protective effect on intestinal inflammation triggered by soybean meal. However, multiple factors were involved, such as interspecific variation, inclusion of soybean meal and percentage of additive used in the supplementation. Thus, further research is needed to compare efficiencies between new “intestinal protector” additives.

To develop research in zebrafish to find solutions to the issues mentioned above requires first the corroboration that intestinal inflammation triggered by soybean meal in zebrafish recapitulates what is seen on farmed fish. This approach has been addressed exactly by Hedrera et al. (2013). They present a new strategy to analyze the potential intestinal impact that the intake of different food ingredients can trigger. Specifically, they analyzed the effects of the ingestion of soybean meal and two of its components, soy protein and soy saponin in zebrafish. Demonstrating that larvae fed with soybean meal developed an intestinal inflammation as early as 2 days after start feeding. Moreover, it was observed that saponin but not soy protein extract was responsible for the inflammatory response.

These findings support the use of zebrafish screening assays to identify novel ingredients/additives that would lead to improved current fish diets or to the formulation of new ones. **Figure 1** illustrates the two steps in the new proposed strategy. The first step is a “pre-screening” developed in zebrafish. The aim of this step is



to evaluate a large number and wide range of ingredients or additives in order to select the more beneficial or less harmful. The second step considers the determination of the intestinal effects generated by the selected ingredients on the target fish specie. This method eliminates the need to evaluate all diets directly on commercial fish, reducing high costs and time consuming experimentation.

Besides zebrafish has potential in nutritional studies, it is important to highlight that assays in this fish cannot replace analysis in farmed fish, as well as results cannot be direct extrapolated to other fish species. For example, results regarding the level of enteritis triggered by soybean meal in Atlantic salmon are different from those detected in rainbow trout. What is important is that in both species soybean causes intestinal inflammation that is mainly due to the saponin, which also occurs in zebrafish. Moreover, in all these species the level of proinflammatory cytokines are upregulated, suggesting that the molecular mechanisms are conserved. In a similar way, several studies have shown that the intake of a diet based on soybean meal decreases the growth rate of salmon, rainbow trout, carp, tilapia, sea bream, and also zebrafish (Pongmaneerat et al., 1993; Médale et al., 1998; Fontainhas-Fernandes et al., 1999; Gómez-Requeni et al., 2004; Mundheim et al., 2004; Ulloa et al., 2013). These results suggest that the biological processes and molecular mechanisms that

underlie the growth response to nutrients overlap among different fish, regardless of evolutionary distance or environmental conditions. Understanding how signaling cascades are coordinated and their effects on physiological response, such as growth and inflammation, may be unraveled in zebrafish. Thus, investigations undertaken in zebrafish nutrition could make important contributions to aquaculture nutrition research.

FUTURE DIRECTIONS

The current challenge is to apply the knowledge obtained in zebrafish to benefit the aquaculture industry. In the future, one of the principal challenges will be to cultivate carnivorous fish that can tolerate higher levels of plant protein in their diet. New technologies such as RNA-seq and genotyping platforms will be key in our ability to select fish with increased tolerance to a vegetal protein diet. Identification of more friendly vegetal ingredients should also be examined. Thus, it is not hard to imagine that in a near future fish diets will be formulated with ingredients and/or additives according to the genetic background of the strain of interest instead of depending solely on the species.

AUTHOR CONTRIBUTIONS

Pilar E. Ulloa: drafting of the manuscript; Carmen G. Feijoo: drafting of the manuscript and critical revision of the intellectual

content; Juan F. Medrano: critical revisions of the intellectual content.

ACKNOWLEDGMENT

Grant of Fondecyt 3130664 to Pilar E. Ulloa.

REFERENCES

- Alami-Durante, H., Médale, F., Cluzeaud, M., and Kaushik, S. J. (2010). Skeletal muscle growth dynamics and expression of related genes in white and red muscles of rainbow trout fed diets with graded levels of a mixture of plant protein sources as substitutes for fishmeal. *Aquaculture* 303, 50–58. doi: 10.1016/j.aquaculture.2010.03.012
- Baeverfjord, G., and Krogdahl, A. (1996). Development and regression of soybean meal induced enteritis in Atlantic salmon, *Salmo salar* L., distal intestine: a comparison with the intestines of fasted fish. *J. Fish Dis.* 19, 375–387. doi: 10.1046/j.1365-2761.1996.d01-92.x
- Bakke-Mckellep, M., McL Press, C., Baeverfjord, G., Krogdahl, A., and Landsverk, T. (2000). Changes in immune and enzyme histochemical phenotypes of cell in the intestinal mucosa of Atlantic salmon, *Salmo salar* L., with soybean meal-induced enteritis. *J. Fish Dis.* 23, 115–127. doi: 10.1046/j.1365-2761.2000.00218.x
- Bakke-Mckellep, M., Penn, H., Salas, M., Refstie, S., Sperstad, S., Landsverk, T., et al. (2007). Effects of dietary soybean meal, inulin and oxytetracycline on intestinal microbiota and epithelial cell stress, apoptosis and proliferation in the teleost Atlantic salmon (*Salmo salar* L.). *Br. J. Nutr.* 97, 699–713. doi: 10.1017/S0007114507381397
- Barros-Becker, F., Romero, J., Pulgar, A., and Feijoo, C. G. (2012). Persistent oxytetracycline exposure induces an inflammatory process that improves regenerative capacity in zebrafish larvae. *PLoS ONE* 5:e36827. doi: 10.1371/journal.pone.0036827
- Bergeron, S., Milla, L., Villegas, R., Shen, M.-C., Burgess, S., Allende, M., et al. (2008). Expression profiling identifies novel Hh/Gli regulated genes in developing zebrafish embryos. *Genomics* 91, 165–177. doi: 10.1016/j.ygeno.2007.09.001
- Chen, G. Y., and Nuñez, G. (2010). Sterile inflammation: sensing and reacting to damage. *Nat. Rev. Immunol.* 10, 826–837. doi: 10.1038/nri2873
- Cui, J., Liu, S., Zahng, B., Wang, H., Sun, H., Song, S., et al. (2014). Transcriptome analysis of the gill and swimbladder of *Takifugu rubripes* by RNA-seq. *PLoS ONE* 9:e85505. doi: 10.1371/journal.pone.0085505
- Czubinski, J., Dwiecki, K., Siger, A., Neunert, G., and Lampart-Szczapa, E. (2014). Characterisation of different digestion susceptibility of lupin seed Globulins. *Food Chem.* 143, 418–426. doi: 10.1016/j.foodchem.2013.08.015
- Dimitroglou, A., Marrifield, D., Spring, P., Sweetman, J., Moate, R., and Davies, S. (2010). Effects of mannan oligosaccharide (MOS) supplementation on growth performance, feed utilization, intestinal histology and gut microbiota of gilthead sea bream (*Sparus aurata*). *Aquaculture* 300, 182–188. doi: 10.1016/j.aquaculture.2010.01.015
- Douglas, S. (2006). Microarray studies of genes expression in fish. *OMICS* 10, 474–489. doi: 10.1089/omi.2006.10.474
- Engelsma, M., Huising, M., Van Muiswinkel, W., Flik, G., Kwang, G., Savelkoul, H., et al. (2002). Neuroendocrine-immune interaction in fish: a role for interleukin-1. *Vet. Immunol. Immunopathol.* 87, 467–479. doi: 10.1016/S0165-2427(02)00077-6
- Fialkow, L., Wang, Y., and Downey, G. P. (2007). Reactive oxygen and nitrogen species signaling molecules regulating neutrophil function. *Free Radic. Biol. Med.* 42, 153–164. doi: 10.1016/j.freeradbiomed.2006.09.030
- Fontainhas-Fernandes, A., Gomes, E., Reis-Henriques, M. A., and Coimbra, J. (1999). Replacement of fish meal by plant proteins in the diet of Nile tilapia: digestibility and growth performance. *Aquac. Int.* 7, 57–67. doi: 10.1023/A:1009296818443
- Food and Agriculture Organization of the United Nations [FAO]. (2012). El Estado Mundial de la Pesca y la Acuicultura. Organización de las Naciones Unidas para la Alimentación y la Agricultura. Rome: FAO.
- Froystad-Saugen, M., Lilleeng, E., Bakke-Mckellep, A., Vekterud, K., Valen, E., Hemre, G., et al. (2009). Distal intestinal gene expression in Atlantic salmon (*Salmo salar* L.) fed genetically modified maize. *Aquac. Nutr.* 15, 104–115. doi: 10.1111/j.1365-2095.2008.00572.x
- Gjedrem, T. (2000). Genetic improvement of cold-water fish species. *Aquac. Res.* 31, 25–33. doi: 10.1046/j.1365-2109.2000.00389.x
- Gómez-Requeni, P., Mingarro, M., Caldach-Ginera, J. A., Médale, F., Martin, S. A. M., Houlihan, D. F., et al. (2004). Protein growth performance, amino acid utilisation and somatotropic axis responsiveness to fish meal replacement by plant protein sources in gilthead sea bream (*Sparus aurata*). *Aquaculture* 232, 493–510. doi: 10.1016/S0044-8486(03)00532-5
- Hardy, R. (2010). Utilization of plant proteins in fish diets: effects of global demand and supplies of fishmeal. *Aquac. Res.* 41, 770–776. doi: 10.1111/j.1365-2109.2009.02349.x
- Hedra, M., Galdames, J., Gimenez-Reyes, M., Reyes, A., Avandano-Herrera, R., Romero J., et al. (2013). Soybean meal induces intestinal inflammation in zebrafish larvae. *PLoS ONE* 8:e69983. doi: 10.1371/journal.pone.0069983
- Hernández, A. J., Roman, D., Hooft, J., Cofre, C., Cepeda, V., and Vidal, R. (2013). Growth performance and expression of immune-regulatory gene in rainbow trout (*Oncorhynchus mykiss*) juveniles fed extruded diets with varying levels of lupin (*Lupinus albus*), peas (*Pisum sativum*) and rapeseed (*Brassica napus*). *Aquac. Nutr.* 19, 321–332. doi: 10.1111/j.1365-2095.2012.00961.x
- Houston, R. D., Taggart, J. B., Cezard, T., Bekaert, M., Lowe, N. R., Downing, A., et al. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15:90. doi: 10.1186/1471-2164-15-90
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203, 253–310. doi: 10.1002/aja.1002030302
- Knudsen, D., Uran, P., Arnous, A., Koppe, W., and Frokiaer, H. (2007). Saponin-containing subfractions of soybean molasses induce enteritis in the distal intestine of Atlantic salmon. *J. Agric. Food Chem.* 55, 2261–2267. doi: 10.1021/jf0626967
- Kolditz, C. I., Pabouef, G., Borthaire, M., Esquerré, D., SanCristobal, M., Lefèvre, F., et al. (2008). Changes induced by dietary energy intake and divergent selection for muscle fat content in rainbow trout (*Oncorhynchus mykiss*), assessed by transcriptome and proteome analysis of the liver. *BMC Genomics* 9:506. doi: 10.1186/1471-2164-9-506
- Lawrence, C. (2007). The husbandry of zebrafish (*Danio rerio*): a review. *Aquaculture* 269, 1–20. doi: 10.1016/j.aquaculture.2007.04.077
- Li, Z.-H., Xu, H., Zheng, W., Lam, S. H., and Gong, Z. (2013). RNA-sequencing analysis of TCDD-induced responses in zebrafish liver reveals high relatedness to in vivo mammalian models and conserved biological pathways. *PLoS ONE* 8:e77292. doi: 10.1371/journal.pone.0077292
- Liu, S., Gao, G., Palti, Y., Cleveland, B. M., Weber, G. M., and Rexroad, C. E. III. (2014). RNA-seq analysis of early hepatic response to handling and confinement stress in rainbow trout. *PLoS ONE* 2:e88492. doi: 10.1371/journal.pone.0088492
- Long, Y., Song, G., Yan, J., He, X., Li, Q., and Cui, Z. (2013). Transcriptomic characterization of cold acclimation in larval zebrafish. *BMC Genomics* 14:612. doi: 10.1186/1471-2164-14-612
- Médale, F., Boujard, T., Vallee, F., Blanc, D., Mambrini, M., Roem, R., et al. (1998). Voluntary feed intake, nitrogen and phosphorus losses in rainbow trout (*Oncorhynchus mykiss*) fed increasing dietary levels of soy protein concentrate. *Aquat. Living Resour.* 11, 239–246. doi: 10.1016/S0990-7440(98)89006-2
- Merrifield, D., Dimitroglou, A., Bradley, G., Baker, R., and Davies, S. (2009). Soybean meal alters autochthonous microbial populations, microvilli morphology and compromises intestinal enterocyte integrity of rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J. Fish Dis.* 32, 755–766. doi: 10.1111/j.1365-2761.2009.01052.x
- Montero, D., Mathlouthi, F., Tort, L., Afonso, J. M., Torrecillas, S., Fernández-Vaquero, A., et al. (2010). Replacement of dietary fish oil by vegetable oils affects humoral immunity and expression of pro-inflammatory cytokines genes in gilthead sea bream *Sparus aurata*. *Fish Shellfish Immunol.* 29, 1073–1081. doi: 10.1016/j.fsi.2010.08.024
- Morais, S., Pratoomyot, J., Taggart, J., Bron, J., Guy, D., Bell, J., et al. (2011). Genotype-specific responses in Atlantic salmon (*Salmo salar*) subject to dietary. *BMC Genomics* 12:255. doi: 10.1186/1471-2164-12-255
- Morozova, O., and Marra, M. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264. doi: 10.1016/j.ygeno.2008.07.001
- Morris, P., Gallimore, P., Handley, J., Hide, G., Haughton, P., and Black, A. (2005). Full fat soya for rainbow trout (*Oncorhynchus mykiss*) in freshwater: effects on performance, composition and flesh fatty acid profile in absence of

- hind-gut enteritis. *Aquaculture* 248, 147–161. doi: 10.1016/j.aquaculture.2005.04.021
- Mundheim, H., Aksnes, A., and Hope, B. (2004). Growth, feed efficiency and digestibility in salmon (*Salmo salar* L.) fed different dietary proportions of vegetable protein sources in combination with two fish meal qualities. *Aquaculture* 237, 315–331. doi: 10.1016/j.aquaculture.2004.03.011
- Mutch, M., Wahli, W., and Williamson, G. (2005). Nutrigenomics and nutrigenetics: the emerging faces of nutrition. *FASEB J.* 19, 1602–1615. doi: 10.1096/fj.05-3911rev
- Naylor, R. L., Hardy, R. W., Bureau, D. P., Chiu, A., Elliott, M., Farrell, A. P., et al. (2009). Feeding aquaculture in an era of finite resources. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15103–15110. doi: 10.1073/pnas.0905235106
- Oliveira, S., Reyes-Aldasoro, C. C., Candel, S., Renshaw, S. A., Mulero, V., and Calado, A. (2013). Cxcl8 (IL-8) mediates neutrophil recruitment and behavior in the zebrafish inflammatory response. *J. Immunol.* 190, 4349–4359. doi: 10.4049/jimmunol.1203266
- Ouyang, W., Rutz, S., Crellin, N. K., Valdez, P. A., and Hymowitz, S. G. (2011). Regulation and functions of the IL-10 family of cytokines in inflammation and disease. *Annu. Rev. Immunol.* 29, 71–109. doi: 10.1146/annurev-immunol-031210-101312
- Oyarbide, U., Rainieri, S., and Pardo, M. (2012). Zebrafish (*Danio rerio*) larvae as a system to test the efficacy of polysaccharides as immunostimulants. *Zebrafish* 9, 74–84. doi: 10.1089/zeb.2011.0724
- Panserat, S., Ducasse-Cabanot, S., Plagnes-Juan, E., Srivastava, P. P., Kolditz, C., Piumi, F., et al. (2008). Dietary fat level modifies the expression of hepatic genes in juvenile rainbow trout (*Oncorhynchus mykiss*) as revealed by microarray analysis. *Aquaculture* 275, 235–241. doi: 10.1016/j.aquaculture.2007.12.030
- Piaget, N., Vega, A., Silva, A., and Toledo, P. (2007). Effect of the application of β -glucans and mannanoligosaccharides (MOS) in an intensive larval rearing system of *Paralichthys adspersus* (Paralichthyidae). *Investig. Mar.* 35, 35–43. doi: 10.4067/S0717-71782007000200004
- Pongmaneerat, J., Watanabe, T., Takeuchi, T., and Satoh, T. (1993). Use of different protein meals as partial or total substitution for fish meal in carp diets. *Nippon Suisan Gakk* 59, 1249–1257. doi: 10.2331/suisan.59.1249
- Qian, X., Ba, Y., Zhuang, Q., and Zhong, G. (2014). RNA-seq technology and its application in fish transcriptomics. *OMICS: J. Integr. Biol.* 18, 98–110. doi: 10.1089/omi.2013.0110
- Refstie, S., Baeverfjord, G., Seim, R. R., and Elvebo, O. (2010). Effects old dietary yeast cell wall β -glucans and MOS on performance, gut health, and salmon lice resistance in Atlantic salmon (*Salmo salar*) fed sunflower and soybean meal. *Aquaculture* 305, 109–116. doi: 10.1016/j.aquaculture.2010.04.005
- Renshaw, S. A., Loynes, C. A., Trushell, D. M., Elworthy, S., Ingham, P. W., and Whyte, M. K. (2006). A transgenic zebrafish model of neutrophilic inflammation. *Blood* 108, 3976–3978. doi: 10.1182/blood-2006-05-024075
- Ribas, L., and Piferrer, F. (2013). The zebrafish (*Danio rerio*) as a model organism, with emphasis on applications for finfish aquaculture research. *Rev. Aquac.* 5, 1–32. doi: 10.1111/raq.12041
- Rodriguez-Tovar, L., Speare, D., and Markham, R. (2011). Fish microsporidia: immune response, immunomodulation and vaccination. *Fish Shellfish Immunol.* 30, 999–1006. doi: 10.1016/j.fsi.2011.02.011
- Rosamond, L. N., Goldberg, R. J., Primavera, J. H., Kautsky, N., Beveridge, M. C. M., Clay, J., et al. (2000). Effect of aquaculture on world fish supplies. *Nature* 405, 1017–1024. doi: 10.1038/35016500
- Roush, W. (1996). Zebrafish embryology builds a better model vertebrate. *Science* 272:1103. doi: 10.1126/science.272.5265.1103
- Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264
- Sicuro, B., Badino, P., Dapra, F., Gai, F., Galloni, M., Odore, R., et al. (2010). Physiological effects of natural olive oil antioxidants utilization in rainbow trout (*Oncorhynchus mykiss*) feeding. *Aquac. Int.* 18, 415–431. doi: 10.1007/s10499-009-9254-6
- Staykov, Y., Spring, P., Denev, S., and Sweetman, J. (2007). Effect of a mannan oligosaccharide on the growth performance and immune status of rainbow trout (*Oncorhynchus mykiss*). *Aquac. Int.* 15, 153–160. doi: 10.1007/s10499-007-9096-z
- Tacchi, L., Bickerdike, R., Douglas, A., Secombes, C., and Martin, S. (2011). Transcriptome responses to functional feeds in Atlantic salmon (*Salmo salar*). *Fish Shellfish Immunol.* 31, 704–715. doi: 10.1016/j.fsi.2011.02.023
- Tahmasebi-Kohyani, A., Keyvanshokoo, S., Nematollahi, A., Mahmoudi, N., and Pasha-Zanoosi, H. (2012). Effects of dietary nucleotides supplementation on rainbow trout (*Oncorhynchus mykiss*) performance and acute stress response. *Fish Physiol. Biochem.* 38, 431–440. doi: 10.1007/s10695-011-9524-x
- Ulloa, P. E., Iturra, P., Neira, R., and Araneda, C. (2011). Zebrafish as a model organism for nutrition and growth: towards comparative studies of nutritional genomics applied to aquacultured fishes. *Rev. Fish Biol. Fish.* 21, 649–666. doi: 10.1007/s11160-011-9203-0
- Ulloa, P. E., Peña, A., Lizama, C. D., Araneda, C., Iturra, P., Neira, R., et al. (2013). Growth response and expression of muscle growth-related candidate genes in adult zebrafish fed plant and fishmeal protein-based diets. *Zebrafish* 1, 99–109. doi: 10.1089/zeb.2012.0823
- Urán, P., Goncalves, A., Taverne-Thiele, J., Schrama, J., Verreth, J., and Rombout, J. (2008). Soybean meal induces intestinal inflammation in common carp (*Cyprinus carpio* L.). *Fish Shellfish Immunol.* 25, 751–760. doi: 10.1016/j.fsi.2008.02.013
- Von Schalburg, K. R., Cooper, G. A., Leong, J., Robb, A., Lieph, R., Rise, M. L., et al. (2008). Expansion of the genomics research on Atlantic salmon *Salmo salar* L. project (GRASP) microarray tools. *J. Fish Biol.* 72, 2051–2070. doi: 10.1111/j.1095-8649.2008.01910.x
- Wickramasinghe, S., Cánovas, A., Rincón, G., and Medrano, J. F. (2014). RNA-sequencing: a tool to explore new frontiers in animal genetics. *Livest. Sci.* 166, 206–216. doi: 10.1016/j.livsci.2014.06.015i
- Witko-Sarsat, V., Rieu, P., Descamps-Latscha, B., Lesavre, P., and Halbwachs-Mecarelli, L. (2000). Neutrophils: molecules, functions and pathophysiological aspects. *Lab. Invest.* 80, 617–653. doi: 10.1038/labinvest.3780067
- Xu, H., Lam, S. H., Shen, Y., and Gong, Z. (2013). Genome-wide identification of molecular pathways and biomarkers in response to arsenic exposure in zebrafish liver. *PLoS ONE* 8:e68737. doi: 10.1371/journal.pone.0068737

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 June 2014; accepted: 20 August 2014; published online: 10 September 2014.

Citation: Ulloa PE, Medrano JF and Feijoo CG (2014) Zebrafish as animal model for aquaculture nutrition research. *Front. Genet.* 5:313. doi: 10.3389/fgene.2014.00313
This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Ulloa, Medrano and Feijoo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals

Marc Vandeputte^{1,2*} and Pierrick Haffray³

¹ INRA UMR1313 Génétique Animale et Biologie Intégrative, Institut National de la Recherche Agronomique, Jouy en Josas, France

² Ifremer, Institut Français de Recherche pour l'Exploitation de la Mer, Palavas-les-Flots, France

³ Sysaaf, Syndicat des Sélectionneurs Avicoles et Aquacoles Français, Rennes, France

Edited by:

Ross Houston, University of Edinburgh, UK

Reviewed by:

Jeff Silverstein, United States Department of Agriculture, USA
Gen H. Yue, National University of Singapore, Singapore

*Correspondence:

Marc Vandeputte, Ifremer/INRA, Chemin de Maguelone, F-34250 Palavas-les-Flots, France
e-mail: marc.vandeputte@jouy.inra.fr

Since the middle of the 1990s, parentage assignment using microsatellite markers has been introduced as a tool in aquaculture breeding. It now allows close to 100% assignment success, and offered new ways to develop aquaculture breeding using mixed family designs in commercial conditions. Its main achievements are the knowledge and control of family representation and inbreeding, especially in mass spawning species, above all the capacity to estimate reliable genetic parameters in any species and rearing system with no prior investment in structures, and the development of new breeding programs in many species. Parentage assignment should not be seen as a way to replace physical tagging, but as a new way to conceive breeding programs, which have to be optimized with its specific constraints, one of the most important being to well define the number of individuals to genotype to limit costs, maximize genetic gain while minimizing inbreeding. The recent possible shift to (for the moment) more costly single nucleotide polymorphism markers should benefit from future developments in genomics and marker-assisted selection to combine parentage assignment and indirect prediction of breeding values.

Keywords: aquaculture, parentage assignment, selective breeding, microsatellites, SNPs

INTRODUCTION

Aquaculture is now the fastest growing animal production worldwide, and provides half of the fish for human consumption worldwide (FAO, 2014). Such an important sector would be expected to use the best knowledge-based improvement methods, amongst which selective breeding is of paramount importance. However, Gjedrem et al. (2012) estimated that only 10% of aquaculture production worldwide is based on genetically improved stocks. There may be several reasons for this, but one clear technical weakness of aquaculture regarding the development of optimized selective breeding schemes is the fact that pedigree information is difficult and costly to obtain.

The basic reason is rather straightforward: farmed aquatic animals are all too small at hatching (from a few micrograms in mollusks and crustaceans to ca. 100 mg in salmonids fishes) to be physically tagged.

Then, there were initially two ways for fish genetic studies and breeding programs to deal with the question of pedigrees. The first and simpler solution was not to use a pedigree, using individual selection. In this case, fish are selected solely based on their own individual phenotype (see review in Gjedrem and Thodesen, 2005). Although effective to obtain genetic gain, this method is very limiting for studying genetic variation as: (1) it provides results only after a minimum of two generations, (2) it

requires the maintenance of at least two fish lines, selected/control or divergent lines, (3) it limits the evaluation of genetic variation to one trait only, and (4) the precision of realized heritability estimates is low in reasonably-sized two generation experiments (Nicholas, 1980).

The second option to solve the pedigree problem is to use separate rearing of families until a size where tagging is possible, as in the Norwegian salmon breeding program, the first family-based selective breeding program in aquaculture, started in 1972 (Gjedrem, 2010). This was successfully extended to major aquaculture species such as salmonids, tilapias, oyster, or shrimps (Krishna et al., 2011; Thodesen et al., 2011; Gjedrem, 2012; Gjedrem et al., 2012; Zak et al., 2014). Although efficient, this method has three main drawbacks when it comes to estimating genetic parameters of traits. First, as families are reared separately, common environmental effects between tanks may inflate heritability estimates. The second point is that studying genetic variation with separate rearing of progenies requires the pre-existence of the family rearing units—i.e., of the infrastructure of the breeding program. Exploratory studies are then difficult to undertake. The third point is that the number of families is limited to the number of family rearing units used. Then, mating designs are constrained to those where the number of families produced is low for a given number of parents tested, like single pair mating

or nested designs, which, unlike factorial designs, do not allow the separation of additive, maternal, common environment, and dominance effects (Becker, 1967).

Therefore, the provision of a method to trace pedigrees in groups of mixed families, with any type of family structure, was expected to be of great interest to study genetic variation of quantitative traits in aquaculture species, and subsequently to set up new types of breeding programs.

The principles of parentage assignment were set up for livestock paternity testing with allozymes (Jamieson, 1965). In fish, the very first trials were done in the 1970s in Israel, also using allozymes, but the number of families that could be discriminated was very low (<10) and the use of the method was limited to carp in one research team (e.g., Brody et al., 1981). The real start of parentage assignment studies in fish was in the 1990s with the availability of microsatellite markers (Herbinger et al., 1995; Estoup et al., 1998).

TECHNICAL ASPECTS OF PARENTAGE ASSIGNMENT

PARENTAGE ASSIGNMENT METHODS

Basically, two computation methods are used for parentage assignment, exclusion-based methods and likelihood-based methods (see Jones et al., 2010 for a review). Exclusion is very simple and makes no hypotheses other than Mendelian segregation of alleles, but is very sensitive to genotyping errors. When error rates are moderate and theoretical assignment power is high, however, genotyping errors can be dealt with by allowing a limited number of allelic mismatches between an offspring and its parents alleles (Vandeputte et al., 2006), and exclusion remains the gold standard of parentage assignment (Yue and Xia, 2014). Exclusion programs used in aquaculture are PROBMAX (Danzmann, 1997), VITASSIGN (Vandeputte et al., 2006), and FAP (Taggart, 2007). Likelihood methods use a different approach, with probabilities. In this case, the most likely couple is chosen as the true one (eventually integrating a genotyping error rate), but the decision rules rely on hypotheses on allelic frequencies. Likelihood methods generally give higher assignment rates than exclusion with low power marker sets, but sometimes give inconsistent results (Herlin et al., 2007; Trong et al., 2013). Using sibship information in calculations can greatly improve the efficiency of the likelihood methods (Wang and Santure, 2009). Likelihood programs used in aquaculture are CERVUS (Kalinowski et al., 2007), PAPA (Duchesne et al., 2002), and PARENTE (Cercueil et al., 2002).

A specific question is also the assignment of polyploids, especially sturgeons (Rodzen et al., 2004) or induced polyploids (Miller et al., 2014), and specific packages have been developed for tetraploids (wHDP; Galli et al., 2011), diploids to octoploids (VITASSIGN-OCTO; Vandeputte, unpublished), as well as a general method to transform polyploid genotypes to pseudo-diploid dominant genotypes (Wang and Scribner, 2014).

A CRUCIAL ISSUE: THE ASSIGNMENT POWER OF MARKERS USED

However, whatever the method used, the first requirement to be able to use parentage assignment in practice is to obtain high levels of unique assignments, which primarily depends

on the assignment power of the marker set used. It depends on the exclusion probabilities of the markers used and on the size of the problem to be solved, the total number of putative parents having an exponential effect on the proportion of unassigned individuals (Vandeputte, 2012). Overestimation of the assignment power of markers is very frequent (Vandeputte et al., 2011), and can be explained by Hardy–Weinberg disequilibrium (Wang, 2007), sampling variance and relatedness of parents (Villanueva et al., 2002; Matson et al., 2008), incomplete genotypes, genotyping errors especially caused by stuttering or size-shift (Sutton et al., 2011; Yue and Xia, 2014), and null alleles (Christie, 2010). In some species groups like mollusks, null alleles may be extremely frequent and problematic (Hedgecock et al., 2004), but the main cause of overestimation of the theoretical assignment power is a widespread inappropriate calculation method (Vandeputte, 2012). Typically, assignment power >0.99 can generally be obtained by 8–15 microsatellite markers in fish crosses involving a few tens or hundreds of parents, and a reasonable option when designing a marker set is to include a few more markers than theoretically needed. This then spares a lot of time by providing easy assignment even if small problems of genotyping errors, inbreeding or null alleles appear. High quality genotyping is also essential, and a recent review by Yue and Xia (2014) gives very useful insights to this question.

MICROSATELLITES AND SNPs FOR PARENTAGE ASSIGNMENT

Microsatellites, due to their high number and high variability, are the markers that allowed the development of efficient parentage assignment methods. Today, however, SNPs (single nucleotide polymorphisms) use is growing exponentially (Guichoux et al., 2011), but not yet in parentage assignment. It was estimated that ~6 SNPs give the same assignment power as 1 microsatellite (Glaubitz et al., 2003). Empirical studies tend to suggest that the adequate number of SNP for an efficient panel would be in the 100–450 range (Trong et al., 2013; Lapègue et al., 2014; Nguyen et al., 2014; Sellars et al., 2014). With such numbers, the classical requirement of unlinked markers within a panel cannot be met, thus lowering the real assignment power. SNPs are individually less expensive to genotype than microsatellites, but multiplexing decreased the cost of microsatellites genotyping (Guichoux et al., 2011; Yue and Xia, 2014), and for the moment SNPs remain more expensive due to the number required, but technology is rapidly evolving for SNPs and not for microsatellites. Empirical studies also sometimes reveal quite a high number of genotyping error in SNPs (Trong et al., 2013) and the necessity to test a higher number of SNP markers than expected to select the appropriate ones (Lapègue et al., 2014; Nguyen et al., 2014). However, prospects for development of genomic selection with low-marker density may imply genotyping of a few hundred to several thousand SNPs per fish (Lillehammer et al., 2013), which in this case would be sufficient to provide parentage assignment at no additional cost. The recent shift to SNP markers was, however, efficient to improve assignment at least in some mollusks species which suffered from high numbers of null alleles with microsatellites (Lapègue et al., 2014; Nguyen et al., 2014).

IMPLEMENTATION OF PARENTAGE ASSIGNMENT IN AQUACULTURE

INBREEDING CONTROL

Mass selection is the simplest way to improve traits such as growth or morphology, but bears a high risk of rapid genetic loss, with highly unbalanced families, which was revealed by parentage assignment mostly in mass spawning species (Perez-Enriquez et al., 1999; Waldbieser and Wolters, 1999; Boudry et al., 2002; Brown et al., 2005; Fessehaye et al., 2006; Herlin et al., 2007; Wang et al., 2008), but also in controlled artificial reproduction systems (Saillant et al., 2002; Kaspar et al., 2008).

The impact of different factors (mating design, mating ratio, number of parents per generation, selection pressure, trait heritability, grading practices) were simulated to improve inbreeding control and optimize genetic progress (Gjerde et al., 1996; Dupont-Nivet et al., 2006; Loughnan et al., 2013; Domingos et al., 2014), including optimal contribution selection, which requires pedigree knowledge (Sonesson, 2005; Skaarud et al., 2011).

ESTIMATION OF GENETIC PARAMETERS

Estimation of heritability and genetic correlations allows to evaluate expected genetic gains and to design breeding programs. This is maybe where the possibility to access pedigree information by genotyping gave the most important and fruitful contribution to date to aquaculture genetics.

Optimization of mixed family designs for genetic parameters estimation was done by Vandeputte et al. (2001) for strain effects, Dupont-Nivet et al. (2002) for heritability, and Sae-Lim et al. (2010) for genotype by environment ($G \times E$) interaction. After several feasibility studies with few families (Herbinger et al., 1995, 1999; Saillant et al., 2002), heritabilities were estimated in a growing number of fishes species for growth (Wilson et al., 2003; Saillant et al., 2006; Dupont-Nivet et al., 2008; Pierce et al., 2008; Wang et al., 2008; Gheyas et al., 2009; Domingos et al., 2013; Whatmore et al., 2013), processing traits (Kocour et al., 2007; Navarro et al., 2009; Saillant et al., 2009; Haffray et al., 2012a), flesh color (Norris and Cunningham, 2004), muscle fiber diameter (Vieira et al., 2007), deformities (Bardon et al., 2009), disease resistance (Guy et al., 2006; Antonello et al., 2009) or sex ratio (Vandeputte et al., 2007), and in shrimps and mollusks for growth (Jerry et al., 2006; Lucas et al., 2006; Kong et al., 2013; Nguyen et al., 2014) or meat yield in mussel (Nguyen et al., 2014). Heritabilities obtained in mixed family rearing are often higher than those recorded in separate rearing, which may be linked to the general absence of between-family environmental variance due to family mixing, although non-genetic maternal effects may persist in mixed family rearing (Haffray et al., 2012b), with possible upward biases on heritability estimates. Explicit comparisons of the same families in mixed or separate rearing design concluded that separate family rearing induced much higher levels of between-families environmental effects (Herbinger et al., 1999; Ninh et al., 2011).

Mixed family rearing also allowed the estimation of $G \times E$ interactions between rearing systems (Dupont-Nivet et al., 2008, 2010c; Navarro et al., 2009; Kvingedal et al., 2010; Domingos et al., 2013; Sae-Lim et al., 2013; Vandeputte et al., 2014), density or rearing temperatures (Saillant et al., 2006), plant-based vs marine

feeds (Pierce et al., 2008; Le Boucher et al., 2013; Bestin et al., 2014), and even separated vs mixed fish family rearing designs (Ninh et al., 2011).

A limiting factor of such studies is that as fish are generally tagged to maximize individual information collection, individual performances are not available before physical tagging, thus limiting genetic studies on early stages. However, recent advances allow individual tagging at 200–400 mg (Ferrari et al., 2014), which should change this matter of fact.

IMPLEMENTATION OF BREEDING PROGRAMS

Concepts used and implementation

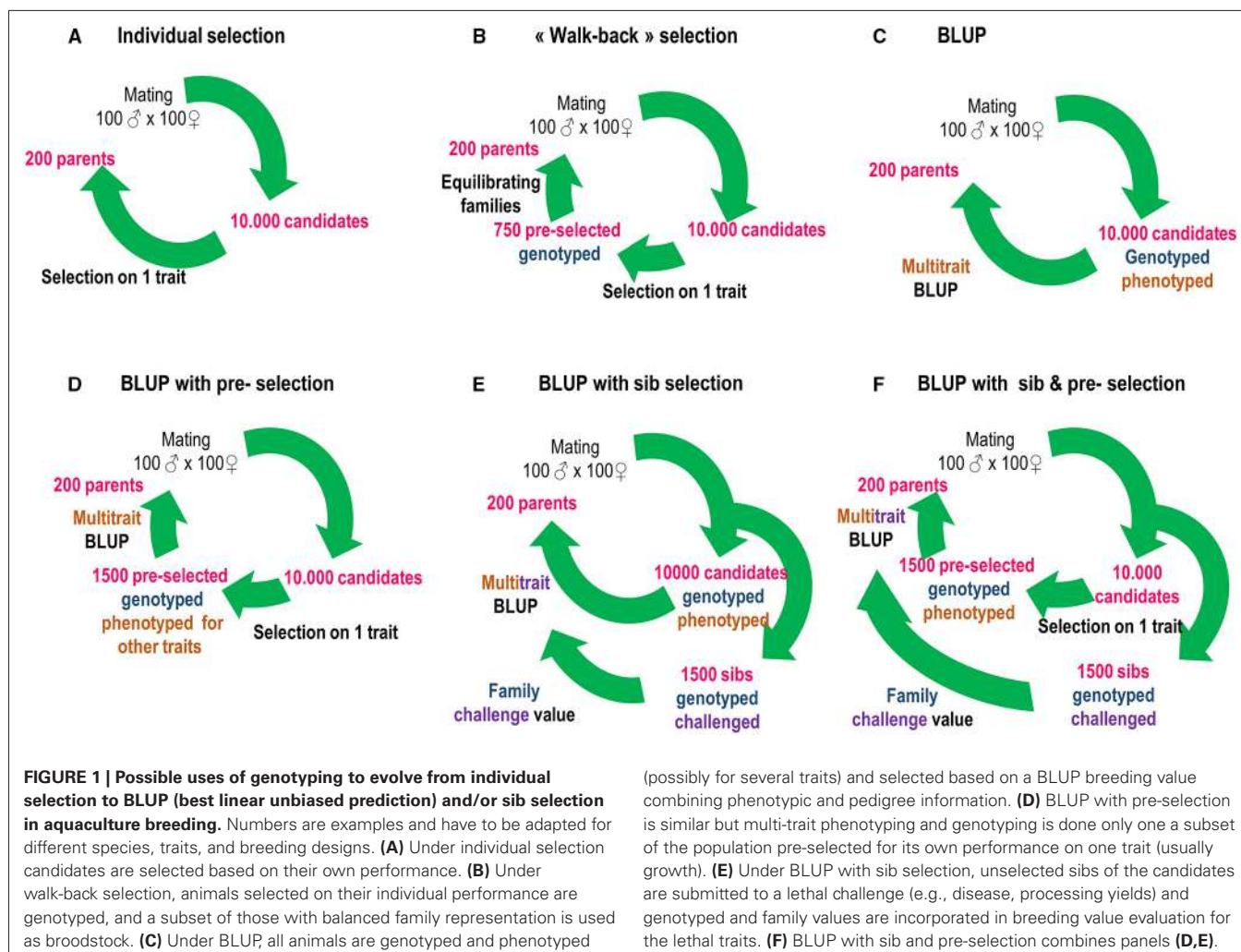
The first proposal to use parentage assignment in breeding at an acceptable cost was an improvement of within-family selection called “walk-back selection” (Doyle and Herbinger, 1995). A two-step process of assignment was suggested and tested to achieve a minimal number of selected candidates per family (Herbinger et al., 1995).

Since this date, public organizations and breeding companies initiated selection programs using parentage assignment in sturgeons (France, USA), Atlantic salmon (Ireland, Norway, Scotland), tilapia (Philippines), halibut (Norway and Scotland), rainbow trout (France), cod (Norway), gilthead sea bream (France, Greece, Spain), turbot (France), European sea bass (France, Greece), meagre and red drum (France), Asian sea bass (Singapore, Indonesia, Australia), and shrimps (Australia, Thailand, Mexico, Equator, Central and South America). This list may be incomplete and represents the present informal expert knowledge of the authors. Little information is publicly available in these programs but mass selection, family-based selection (often BLUP: best linear unbiased prediction) or a combination of both are used to improve growth, processing yields, quality traits and disease resistance according to different schemes (Figure 1).

Key parameters to choose to develop a breeding program using parentage assignment are not only the genotyping cost (12–20 Euros per individual), but also the capacity to produce a large number of families in one batch to avoid tank effects, the true assignment efficiency, as well as the availability of tools such as rapid mass genotyping capacities (specially for species with short generation interval), individual tagging to improve traceability and facilitate data collection, automated database systems to collect, store and link performances to tags, DNA samples and pedigrees, optimized genetic softwares to rank and mate candidates to maximize genetic progress and minimize inbreeding. Use of parentage assignment is not only “genetic tagging,” but requires a complete re-optimization of breeding programs.

Optimization of breeding schemes using parentage assignment

One main target for optimization has been the limitation of numbers genotyped, using two-way nested models for partial pedigrees (Li et al., 2003), or extreme phenotypes with family effect considered as a fixed effect (Morton and Howarth, 2005). BLUP selection normally requires the knowledge of performance and pedigree on all candidates, which is not the case in Figures 1D,F. In these cases, the loss of selection efficiency (compared to BLUP with pedigree known on all candidates) depends on selection intensity and genetic parameters (Chapuis et al., 2010;



Dupont-Nivet et al., 2010b; Sonesson et al., 2011). In addition, issues linked to mixing of families were studied, such as methods to limit non-genetic maternal effect in salmonids (Haffray et al.,

2012b), effect of grading practices to limit cannibalism on family contributions in barramundi (Loughnan et al., 2013), and the importance to consider male maturation status to estimate her-

Table 1 | SWOT analysis of parentage assignment with genomic markers for aquaculture breeding.

Strengths	Weaknesses
<ul style="list-style-type: none"> – Absence of common environment effects – Allows any type of mating design – No investment in structures and limited fish rearing costs (labor, consumables) – Allows family evaluation in industry conditions – Microsatellites available in most species – High flexibility 	<ul style="list-style-type: none"> – Each new trait measured on sibs may require additional genotyping to balance with the benefit expected – Biased BLUP estimates if pre-selection done – Unit cost of genotyping sometimes dissuasive – Ability to produce high numbers of families simultaneously needed for full benefits
Opportunities	Threats
<ul style="list-style-type: none"> – Ease to develop SNP or microsatellites at low cost in any species with next-generation sequencing technologies – Future use of (within family) genomic selection will decrease the cost of pedigree information and provide within-family relatedness estimates improving accuracy – Many research laboratories with appropriate knowledge in genotyping to initiate programs 	<ul style="list-style-type: none"> – Lack of maintenance of microsatellite genotyping equipment with the advent of SNPs – Limited number of professional genotyping companies high volume, service and automation capacities – Lack of efficiency when marker set assignment power overestimated or genotyping errors too numerous

itability of growth more accurately (Dupont-Nivet et al., 2010a). Ninh et al. (2011) and Sae-Lim et al. (2013) compared expected genetic gains with different systems of evaluation (mixed/separate families and impact of $G \times E$ interactions), while Haffray et al. (2013) proposed application of ultrasound tomography to predict processing yields on live candidates to limit the use of slaughtered sibs.

GLOBAL APPRAISAL AND PERSPECTIVES

The rapid increase of publications using parentage assignment in the last decade shows how powerful this method is to estimate genetic parameters in any species and rearing system. It avoids the initial investment in separate family rearing units and limits associated biases, even more in species with high larval mortality, small larval size, and initial live feeding. Applications are strongly driven by reproductive constraints linked to the need to simultaneously produce enough families (Table 1). The cost/information ratio has to be maximized with adequate management of variance sources (number of parents, initial representation of families, or groups of spawns), mating design, and number of individuals genotyped.

Optimal investment in parentage assignment is a balance between the reduction of investment and operational costs needed for the separate family rearing and the cost of genotyping, which presently limits the application of parentage assignment to mass selection and family-based selection on a limited number of traits. Moreover, any new trait that cannot be recorded on the live candidate and has to be measured on sibs then requires additional genotyping with a cost/benefit ratio to estimate case by case, and to compare with the possible use of indirect criteria.

A major benefit of parentage assignment is that it allows high selection pressure (<3%) to be applied in commercial conditions, while still controlling inbreeding. The knowledge of pedigree also allows an increase in selection accuracy (and then a higher selection gain) on all traits, as well as selection on lethal traits which cannot be done by individual selection. This technology also allows to easily combine sanitary protection of the breeding nucleus and sib testing in commercial environments. Parentage assignment offers simplicity and flexibility in the life of the breeding program that can be easily adapted to new traits, new mating schemes, different number of candidates. This is critical, especially at the initiation of domestication, for “niche” species or in developing countries, where the need for separate rearing system has often prevented any investment in selective breeding in the past, or has fixed the architecture of the breeding programs.

ACKNOWLEDGMENT

This work received funding from the European Union’s Seventh Framework Programme (FP7 2007–2013) under grant agreement no. 613611 (FISHBOOST).

REFERENCES

- Antonello, J., Massault, C., Franch, R., Haley, C., Pellizzari, C., Bovo, G., et al. (2009). Estimates of heritability and genetic correlation for body length and resistance to fish pasteurellosis in the gilthead sea bream (*Sparus aurata* L.). *Aquaculture* 298, 29–35. doi: 10.1016/j.aquaculture.2009.10.022
- Bardon, A., Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Haffray, P., Vergnet, A., et al. (2009). What is the heritable component of spinal deformities in the European sea bass (*Dicentrarchus labrax*)? *Aquaculture* 294, 194–201. doi: 10.1016/j.aquaculture.2009.06.018
- Becker, W. A. (1967). *Manual of Procedures in Quantitative Genetics*. Pullman, WA: Washington State University Press.
- Bestin, A., Dupont-Nivet, M., Haffray, P., Médale, F., Quillet, E., Vandeputte, M., et al. (2014). “Genotype by diet interactions on growth and processing traits in rainbow trout (*Oncorhynchus mykiss*), European sea bass (*Dicentrarchus labrax*), gilthead sea bream (*Sparus aurata*) and meagre (*Argyrosomus regius*) fed diets with almost complete substitution of both fish meal and fish oil by vegetal ingredients,” in *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*, Vancouver, Canada.
- Boudry, P., Collet, B., Cornette, F., Hervouet, V., and Bonhomme, F. (2002). High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* 204, 283–296. doi: 10.1016/S0044-8486(01)00841-9
- Brody, T., Wohlfarth, G., Hulata, G., and Moav, R. (1981). Application of electrophoretic genetic markers to fish breeding. IV. Assessment of breeding value of full-sib families. *Aquaculture* 24, 175–186. doi: 10.1016/0044-8486(81)90054-5
- Brown, C. R., Woolliams, J. A., and McAndrew, B. J. (2005). Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture* 247, 219–225. doi: 10.1016/j.aquaculture.2005.02.002
- Cercueil, A., Bellemain, E., and Manel, S. (2002). PARENTE: computer program for parentage analysis. *J. Hered.* 93, 458–459. doi: 10.1093/jhered/93.6.458
- Chapuis, H., Vandeputte, M., Dupont-Nivet, M., Haffray, P., and Quillet, E. (2010). “Selection for an improved disease resistance using factorial mating designs and molecular based pedigrees in fish: a simulation study,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, Germany, CDROM paper 210.
- Christie, M. R. (2010). Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Mol. Ecol. Res.* 10, 115–128. doi: 10.1111/j.1755-0998.2009.02687.x
- Danzmann, R. G. (1997). PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *J. Hered.* 88, 333. doi: 10.1093/oxfordjournals.jhered.a023112
- Domingos, J. A., Smith-Keune, C., and Jerry, D. R. (2014). Fate of genetic diversity within and between generations and implications for DNA parentage analysis in selective breeding of mass spawners: a case study of commercially farmed barramundi, *Lates calcarifer*. *Aquaculture* 424–425, 174–182. doi: 10.1016/j.aquaculture.2014.01.004
- Domingos, J. A., Smith-Keune, C., Robinson, N., Loughnan, S., Harrison, P., and Jerry, D. R. (2013). Heritability of harvest growth traits and genotype \times environment interactions in barramundi, *Lates calcarifer* (Bloch). *Aquaculture* 402–403, 66–75. doi: 10.1016/j.aquaculture.2014.01.004
- Doyle, R. W., and Herbinger, C. M. (1995). “The use of DNA fingerprinting for high-intensity, within family selection in fish breeding,” in *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Guelph, Ontario, Canada, Vol. 19, 364–371.
- Duchesne, P., Godbout, M. H., and Bernatchez, L. (2002). PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. *Mol. Ecol. Notes* 2, 191–193. doi: 10.1046/j.1471-8286.2002.00164.x
- Dupont-Nivet, M., Chevassus, B., Mauger, S., Haffray, P., and Vandeputte, M. (2010a). Side effects of sexual maturation on heritability estimates in rainbow trout (*Oncorhynchus mykiss*). *Aquac. Res.* 41, e878–e880. doi: 10.1111/j.1365-2109.2009.02448.x
- Dupont-Nivet, M., Haffray, P., and Vandeputte, M. (2010b). “Effect of culling on bias of BLUP estimates—case of fish species,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, Germany, CDROM paper 429.
- Dupont-Nivet, M., Karahan-Nomm, B., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., et al. (2010c). Genotype by environment interactions for growth in European seabass (*Dicentrarchus labrax*) are large when growth rate rather than weight is considered. *Aquaculture* 306, 365–368. doi: 10.1016/j.aquaculture.2010.05.025
- Dupont-Nivet, M., Vandeputte, M., and Chevassus, B. (2002). Optimization of factorial mating designs for inference on heritability in fish species. *Aquaculture* 204, 361–370. doi: 10.1016/S0044-8486(01)00839-0

- Dupont-Nivet, M., Vandeputte, M., Haffray, P., and Chevassus, B. (2006). Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture* 252, 161–170. doi: 10.1016/j.aquaculture.2005.07.005
- Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., et al. (2008). Heritabilities and G × E interactions for growth in the European sea bass (*Dicentrarchus labrax* L.) using a marker-based pedigree. *Aquaculture* 275, 81–87. doi: 10.1016/S0044-8486(01)00839-0
- Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P., and Guyomard, R. (1998). Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Can. J. Fish. Aquat. Sci.* 55, 715–725. doi: 10.1139/f97-268
- FAO. (2014). *The State of the World Fisheries and Aquaculture 2014*. FAO, Rome.
- Ferrari, S., Chatain, B., Cousin, X., Leguay, D., Vergnet, A., Vidal, M. O., et al. (2014). Early individual electronic identification of sea bass using RFID microtags: a first example of early phenotyping of sex-related growth. *Aquaculture* 426–427, 165–171. doi: 10.1016/j.aquaculture.2014.01.033
- Fessehaye, Y., El-bialy, Z., Rezk, M. A., Crooijmans, R., Bovenhuis, H., and Komen, H. (2006). Mating systems and male reproductive success in Nile tilapia (*Oreochromis niloticus*) in breeding hapas: a microsatellite analysis. *Aquaculture* 256, 148–158. doi: 10.1016/j.aquaculture.2006.02.024
- Galli, A., Greppi, G., Boscari, E., and Bonacina, C. (2011). Parental allocation in *Acipenser naccarii*: statistical aspects and software development. *J. Appl. Ichthyol.* 27, 651–654. doi: 10.1111/j.1439-0426.2011.01709.x
- Gheyas, A. A., Woolliams, J. A., Taggart, J. B., Sattar, M. A., Das, T. K., McAndrew, B. J., et al. (2009). Heritability estimation of silver carp (*Hypophthalmichthys molitrix*) harvest traits using microsatellite based parentage assignment. *Aquaculture* 294, 187–193. doi: 10.1016/j.aquaculture.2009.06.013
- Gjedrem, T. (2010). The first family-based breeding program in aquaculture. *Rev. Aquac.* 2, 2–15. doi: 10.1111/j.1753-5131.2010.01011.x
- Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture* 344–349, 12–22. doi: 10.1016/j.aquaculture.2012.03.003
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350–353, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Gjedrem, T., and Thodesen, J. (2005). “Selection,” in *Selection and Breeding Programs in Aquaculture*, ed. T. Gjedrem (Dordrecht: Springer), 89–111.
- Gjerde, B., Gjoen, H. M., and Villanueva, B. (1996). Optimum designs for fish breeding programmes with constrained inbreeding mass selection for a normally distributed trait. *Livest. Prod. Sci.* 47, 59–72. doi: 10.1016/S0301-6226(96)01000-7
- Glaubitz, J. C., Rhodes, O. E., and DeWoody, J. A. (2003). Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol. Ecol.* 12, 1039–1047. doi: 10.1046/j.1365-294X.2003.01790.x
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., et al. (2011). Current trends in microsatellite genotyping. *Mol. Ecol. Res.* 11, 591–611. doi: 10.1111/j.1755-0998.2011.03014.x
- Guy, D. R., Bishop, S. C., Brotherstone, S., Hamilton, A., Roberts, R. J., McAndrew, B. J., et al. (2006). Analysis of the incidence of infectious pancreatic necrosis mortality in pedigreed Atlantic salmon, *Salmo salar* L., populations. *J. Fish Dis.* 29, 637–647. doi: 10.1111/j.1365-2761.2006.00758.x
- Haffray, P., Bugeon, J., Rivard, Q., Quittet, B., Puyo, S., Allamelou, J. M., et al. (2013). Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 410–411, 236–244. doi: 10.1016/j.aquaculture.2013.06.016
- Haffray, P., Pincet, C., Chapuis, H., Mazeiraud, E., Petit, V., Rault, P., et al. (2012a). Negative genetic trade-offs between production traits and head or bony tissues in large all-female rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 368–369, 145–162. doi: 10.1016/j.aquaculture.2012.09.023
- Haffray, P., Vandeputte, M., Petit, V., Pincet, C., Chatain, B., Chapuis, H., et al. (2012b). Minimizing maternal effect in salmonid families mixed since eyed stages and a posteriori DNA-pedigree. *Livest. Sci.* 150, 170–178. doi: 10.1016/j.livsci.2012.08.017
- Hedgecock, D., Li, G., Hubert, S., Bucklin, K., and Ribes, V. (2004). Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *J. Shellfish Res.* 23, 379–385.
- Herbinger, C. M., Doyle, R. W., Pitman, E. R., Paquet, D., Mesa, K. A., Morris, D. B., et al. (1995). DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137, 245–256. doi: 10.1016/0044-8486(95)01109-9
- Herbinger, C. M., O'Reilly, P. T., Doyle, R. W., Wright, J. M., and O'Flynn, F. (1999). Early growth performance of Atlantic salmon full-sib families reared in single family tanks versus in mixed family tanks. *Aquaculture* 173, 105–116. doi: 10.1016/0044-8486(95)01109-9
- Herlin, M., Taggart, J. B., McAndrew, B. J., and Penman, D. J. (2007). Parentage allocation in a complex situation: a large commercial Atlantic cod (*Gadus morhua*) mass spawning tank. *Aquaculture* 272, S195–S203. doi: 10.1016/j.aquaculture.2007.08.018
- Jamieson, A. (1965). The genetics of transferrins in cattle. *Heredity* 20, 419–441. doi: 10.1038/hdy.1965.54
- Jerry, D. R., Preston, N. P., Crocos, P. J., Keys, S., Meadows, J. R. S., and Li, Y. (2006). Application of DNA parentage analyses for determining relative growth rates of *Penaeus japonicus* families reared in commercial ponds. *Aquaculture* 254, 171–181. doi: 10.1016/j.aquaculture.2005.10.035
- Jones, A. G., Small, C. M., Paczolt, K. A., and Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Mol. Ecol. Res.* 10, 6–30. doi: 10.1111/j.1755-0998.2009.02778.x
- Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x
- Kaspar, V., Vandeputte, M., Kohlmann, K., Hulak, M., Rodina, M., Gela, D., et al. (2008). A proposal and case study towards a conceptual approach of validating sperm competition in common carp (*Cyprinus carpio* L.), with practical implications for hatchery procedures. *J. Appl. Ichthyol.* 24, 406–409. doi: 10.1111/j.1439-0426.2008.01148.x
- Kocour, M., Mauger, S., Rodina, M., Gela, D., Linhart, O., and Vandeputte, M. (2007). Heritability estimates for processing and quality traits in common carp (*Cyprinus carpio* L.) using a molecular pedigree. *Aquaculture* 270, 43–50. doi: 10.1016/j.aquaculture.2007.03.001
- Kong, N., Li, Q., Yu, H., and Kong, L. F. (2013). Heritability estimates for growth-related traits in the Pacific oyster (*Crassostrea gigas*) using a molecular pedigree. *Aquac. Res.* doi: 10.1111/are.12205 [Epub ahead of print].
- Krishna, G., Gopikrishna, G., Gopal, C., Jahageerdar, S., Ravichandran, P., Kannappan, S., et al. (2011). Genetic parameters for growth and survival in *Penaeus monodon* cultured in India. *Aquaculture* 318, 74–78. doi: 10.1016/j.aquaculture.2011.04.028
- Kvingedal, R., Evans, B. S., Lind, C. E., Taylor, J. J. U., Dupont-Nivet, M., and Jerry, D. R. (2010). Population and family growth response to different rearing location, heritability estimates and genotype × environment interaction in the silver-lip pearl oyster (*Pinctada maxima*). *Aquaculture* 304, 1–6. doi: 10.1016/j.aquaculture.2010.02.035
- Lapègue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadiou, C., Gayral, P., et al. (2014). Development of SNP-genotyping arrays in two shellfish species. *Mol. Ecol. Res.* 14, 820–830. doi: 10.1111/1755-0998.12230
- Le Boucher, R., Vandeputte, M., Dupont-Nivet, M., Quillet, E., Ruelle, F., Vergnet, A., et al. (2013). Genotype by diet interactions in European sea bass (*Dicentrarchus labrax* L.): nutritional challenge with totally plant-based diets. *J. Anim. Sci.* 91, 44–56. doi: 10.2527/jas.2012-5311
- Li, X., Field, C., and Doyle, R. W. (2003). Estimation of additive genetic variance components in aquaculture populations selectively pedigreed by DNA fingerprinting. *Biomet. J.* 45, 61–72. doi: 10.1002/bimj.200290016
- Lillehammer, M., Meuwissen, T. H., and Sonesson, A. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45:39. doi: 10.1186/1297-9686-45-39
- Loughnan, S. R., Domingos, J. A., Smith-Keune, C., Forrester, J. P., Jerry, D. R., Beheregaray, L. B., et al. (2013). Broodstock contribution after mass spawning and size grading in barramundi (*Lates calcarifer*, Bloch). *Aquaculture* 404–405, 139–149. doi: 10.1016/j.aquaculture.2013.04.014
- Lucas, T., Macbeth, M., Degnan, S. M., Knibb, W., and Degnan, B. M. (2006). Heritability estimates for growth in the tropical abalone *Haliotis asinina*

- using microsatellites to assign parentage. *Aquaculture* 259, 146–152. doi: 10.1016/j.aquaculture.2006.05.039
- Matson, S. E., Camara, M. D., Eichert, W., and Banks, M. A. (2008). P-LOCI: a computer program for choosing the most efficient set of loci for parentage assignment. *Mol. Ecol. Res.* 8, 765–768. doi: 10.1111/j.1755-0998.2008.02128.x
- Miller, P. A., Elliott, N. G., Vaillancourt, R. E., Kube, P. D., and Koutoulis, A. (2014). Genetic diversity and pedigree assignment in tetraploid Pacific oysters (*Crassostrea gigas*). *Aquaculture* 433, 318–324. doi: 10.1016/j.aquaculture.2014.06.025
- Morton, R., and Howarth, J. M. (2005). Estimation of breeding values using selected pedigree records. *Biometrics* 61, 422–432. doi: 10.1111/j.1541-0420.2005.040305.x
- Navarro, A., Zamorano, M. J., Hildebrandt, S., Ginés, R., Aguilera, C., and Afonso, J. M. (2009). Estimates of heritabilities and genetic correlations for growth and carcass traits in gilthead seabream (*Sparus auratus* L.), under industrial conditions. *Aquaculture* 289, 225–230. doi: 10.1016/j.aquaculture.2008.12.024
- Nguyen, T. T. T., Hayes, B. J., and Ingram, B. A. (2014). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. *Aquaculture* 420–421, 295–301. doi: 10.1016/j.aquaculture.2013.11.021
- Nicholas, F. W. (1980). Size of population required for artificial selection. *Genet. Res.* 35, 85–105. doi: 10.1017/S0016672300013951
- Ninh, N. H., Ponzoni, R. W., Nguyen, N. H., Woolliams, J. A., Taggart, J. B., McAndrew, B. J., et al. (2011). A comparison of communal and separate rearing of families in selective breeding of common carp (*Cyprinus carpio*). *Aquaculture* 322–323, 39–46. doi: 10.1016/j.aquaculture.2011.09.031
- Norris, A. T., and Cunningham, E. P. (2004). Estimates of phenotypic and genetic parameters for flesh colour traits in farmed Atlantic salmon based on multiple trait animal model. *Livest. Prod. Sci.* 89, 209–222. doi: 10.1016/j.livprodsci.2004.02.010
- Perez-Enriquez, R., Takagi, M., and Taniguchi, N. (1999). Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. *Aquaculture* 173, 413–423. doi: 10.1016/S0044-8486(98)00469-4
- Pierce, L. R., Palti, Y., Silverstein, J. T., Barrows, F. T., Hallerman, E. M., and Parsons, J. E. (2008). Family growth response to fishmeal and plant-based diets shows genotype \times diet interaction in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 278, 37–42. doi: 10.1016/j.aquaculture.2008.03.017
- Rodzen, J. A., Famula, T. R., and May, B. (2004). Estimation of parentage and relatedness in the polyploid white sturgeon (*Acipenser transmontanus*) using a dominant marker approach for duplicated microsatellite loci. *Aquaculture* 232, 165–182. doi: 10.1016/S0044-8486(03)00450-2
- Sae-Lim, P., Kause, A., Mulder, H. A., Martin, K. E., Barfoot, A. J., Parsons, J. E., et al. (2013). Genotype-by-environment interaction of growth traits in rainbow trout (*Oncorhynchus mykiss*): a continental scale study. *J. Anim. Sci.* 91, 5572–5581. doi: 10.2527/jas.2012-5949
- Sae-Lim, P., Komen, H., and Kause, A. (2010). Bias and precision of estimates of genotype-by-environment interaction: a simulation study. *Aquaculture* 310, 66–73. doi: 10.1016/j.aquaculture.2010.10.020
- Saillant, E., Dupont-Nivet, M., Haffray, P., and Chatain, B. (2006). Estimates of heritability and genotype-environment interactions for body weight in sea bass (*Dicentrarchus labrax* L.) raised under communal rearing conditions. *Aquaculture* 254, 139–147. doi: 10.1016/j.aquaculture.2005.10.018
- Saillant, E., Dupont-Nivet, M., Sabourault, M., Haffray, P., Laureau, S., Vidal, M. O., et al. (2009). Genetic variation for carcass quality traits in cultured sea bass (*Dicentrarchus labrax*). *Aquat. Living Resour.* 22, 105–112. doi: 10.1051/alr/2009010
- Saillant, E., Fostier, A., Haffray, P., Menu, B., Thimonier, J., and Chatain, B. (2002). Temperature effects and genotype-temperature interactions on sex determination in the European sea bass (*Dicentrarchus labrax* L.). *J. Exp. Zool.* 292, 494–505. doi: 10.1002/jez.10071
- Sellars, M. J., Dierens, L., McWilliam, S., Little, B., Murphy, B., Coman, G. J., et al. (2014). Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquac. Res.* 45, 417–426. doi: 10.1111/j.1365-2109.2012.03243.x
- Skaarud, A., Woolliams, J. A., and Gjoen, H. M. (2011). Strategies for controlling inbreeding in fish breeding programs; an applied approach using optimum contribution (OC) procedures. *Aquaculture* 311, 110–114. doi: 10.1016/j.aquaculture.2010.11.023
- Sonesson, A. K. (2005). A combination of walk-back and optimum contribution selection in fish: a simulation study. *Genet. Sel. Evol.* 37, 587–599. doi: 10.1186/1297-9686-37-7-587
- Sonesson, A. K., Gjerde, B., and Robinson, N. (2011). A simple selection scheme to improve disease resistance and growth. *Aquaculture* 319, 337–341. doi: 10.1016/j.aquaculture.2011.07.009
- Sutton, J. T., Robertson, B. C., and Jamieson, I. G. (2011). Dye shift: a neglected source of genotyping error in molecular ecology. *Mol. Ecol. Res.* 11, 514–520. doi: 10.1111/j.1755-0998.2011.02981.x
- Taggart, J. B. (2007). FAP: an exclusion-based parental assignment program with enhanced predictive functions. *Mol. Ecol. Notes* 7, 412–415. doi: 10.1111/j.1471-8286.2006.01616.x
- Thodesen, J., Rye, M., Wang, Y., Yang, K. S., Bentsen, H. B., and Gjedrem, T. (2011). Genetic improvement of tilapia in China: genetic parameters and selection responses in growth of Nile tilapia (*Oreochromis niloticus*) after six generations of multi-trait selection for growth and fillet yield. *Aquaculture* 322, 51–64. doi: 10.1016/j.aquaculture.2011.10.010
- Trong, T. Q., van Bers, N., Crooijmans, R., Dibbitts, B., and Komen, H. (2013). A comparison of microsatellites and SNPs in parental assignment in the GIFT strain of Nile tilapia (*Oreochromis niloticus*): the power of exclusion. *Aquaculture* 388–391, 14–23. doi: 10.1016/j.aquaculture.2013.01.004
- Vandeputte, M. (2012). An accurate formula to calculate exclusion power of marker sets in parentage assignment. *Genet. Sel. Evol.* 44:36. doi: 10.1186/1297-9686-44-36
- Vandeputte, M., Dupont-Nivet, M., Chatain, B., and Chevassus, B. (2001). Setting up a strain testing design for the seabass, *Dicentrarchus labrax*: a simulation study. *Aquaculture* 202, 329–342. doi: 10.1016/S0044-8486(01)00782-7
- Vandeputte, M., Dupont-Nivet, M., Chavanne, H., and Chatain, B. (2007). A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics* 176, 1049–1057. doi: 10.1534/genetics.107.072140
- Vandeputte, M., Garouste, R., Dupont-Nivet, M., Haffray, P., Vergnet, A., Chavanne, H., et al. (2014). Multi-site evaluation of the rearing performances of 5 wild populations of European sea bass (*Dicentrarchus labrax*). *Aquaculture* 424–425, 239–248. doi: 10.1016/j.aquaculture.2014.01.005
- Vandeputte, M., Mauger, S., and Dupont-Nivet, M. (2006). An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Mol. Ecol. Notes* 6, 265–267. doi: 10.1111/j.1471-8286.2005.01167.x
- Vandeputte, M., Rossignol, M. N., and Pincet, C. (2011). From theory to practice: empirical evaluation of the assignment power of marker sets for pedigree analysis in fish breeding. *Aquaculture* 314, 80–86. doi: 10.1016/j.aquaculture.2011.01.043
- Vieira, V. L. A., Norris, A., and Johnston, I. A. (2007). Heritability of fibre number and size parameters and their genetic relationship to flesh quality traits in Atlantic salmon (*Salmo salar* L.). *Aquaculture* 272(Suppl. 1), S100–S109. doi: 10.1016/j.aquaculture.2007.08.028
- Villanueva, B., Verspoor, E., and Visscher, P. M. (2002). Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. *Anim. Genet.* 33, 33–41. doi: 10.1046/j.1365-2052.2002.00804.x
- Waldbieser, G. C., and Wolters, W. R. (1999). Application of polymorphic microsatellite loci in a channel catfish *Ictalurus punctatus* breeding program. *J. World Aquac. Soc.* 30, 256–262. doi: 10.1111/j.1749-7345.1999.tb00873.x
- Wang, C. M., Lo, L. C., Zhu, Z. Y., Lin, G., Feng, F., Li, J., et al. (2008). Estimating reproductive success of brooders and heritability of growth traits in Asian sea bass (*Lates calcarifer*) using microsatellites. *Aquac. Res.* 39, 1612–1619. doi: 10.1111/j.1365-2109.2008.02034.x
- Wang, J. (2007). Parentage and sibship exclusions: higher statistical power with more family members. *Heredity* 99, 205–217. doi: 10.1038/sj.hdy.6800984
- Wang, J., and Santure, A. W. (2009). Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* 181, 1579–1594. doi: 10.1534/genetics.108.100214

- Wang, J., and Scribner, K. T. (2014). Parentage and sibship inference from markers in polyploids. *Mol. Ecol. Res.* 14, 541–553. doi: 10.1111/1755-0998.12210
- Whatmore, P., Nguyen, N. H., Miller, A., Lamont, R., Powell, D., D'Antignana, T., et al. (2013). Genetic parameters for economically important traits in yellowtail kingfish *Seriola lalandi*. *Aquaculture* 400–401, 77–84. doi: 10.1016/j.aquaculture.2013.03.002
- Wilson, A. J., McDonald, G., Moghadam, H. K., Herbinger, C. M., and Ferguson, M. M. (2003). Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*. *Genet. Res. Camb.* 81, 145–156. doi:10.1017/S0016672302006055
- Yue, G. H., and Xia, J. H. (2014). Practical considerations of molecular parentage analysis in fish. *J. World Aquac. Soc.* 45, 89–103. doi: 10.1111/jwas.12107
- Zak, T., Deshev, R., Benet-Perlberg, A., Naor, A., Magen, I., Shapira, Y., et al. (2014). Genetic improvement of Israeli blue (Jordan) tilapia, *Oreochromis aureus* (Steindachner), through selective breeding for harvest weight. *Aquac. Res.* 45, 546–557. doi: 10.1111/are.12072
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 September 2014; paper pending published: 20 October 2014; accepted: 22 November 2014; published online: 12 December 2014.

Citation: Vandeputte M and Haffray P (2014) Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* 5:432. doi: 10.3389/fgene.2014.00432

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Vandeputte and Haffray. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic improvement of Pacific white shrimp [*Penaeus (Litopenaeus) vannamei*]: perspectives for genomic selection

Héctor Castillo-Juárez¹, Gabriel R. Campos-Montes², Alejandra Caballero-Zamora¹ and Hugo H. Montaldo^{3*}

¹ Departamento de Producción Agrícola y Animal, División de Ciencias Biológicas y de la Salud, Unidad Xochimilco, Universidad Autónoma Metropolitana, Mexico City, Mexico, ² Departamento El Hombre y su Ambiente, Unidad Xochimilco, Universidad Autónoma Metropolitana, Mexico City, Mexico, ³ Departamento de Genética y Bioestadística, Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México, Mexico City, Mexico

OPEN ACCESS

Edited by:

José Manuel Yáñez,
University of Chile, Chile

Reviewed by:

Juan Steibel,
Michigan State University, USA
Ross Houston,
University of Edinburgh, UK

*Correspondence:

Hugo H. Montaldo,
Departamento de Genética y
Bioestadística, Facultad de Medicina
Veterinaria y Zootecnia, Ciudad
Universitaria, Universidad Nacional
Autónoma de México, Circuito
Exterior s/n, 04510 Mexico City,
Mexico
montaldo@unam.mx

Specialty section:

This article was submitted to
Livestock Genomics, a section of the
journal *Frontiers in Genetics*

Received: 22 September 2014

Accepted: 20 February 2015

Published: 24 March 2015

Citation:

Castillo-Juárez H, Campos-Montes
GR, Caballero-Zamora A and
Montaldo HH (2015) Genetic
improvement of Pacific white shrimp
[*Penaeus (Litopenaeus) vannamei*]:
perspectives for genomic selection.
Front. Genet. 6:93.
doi: 10.3389/fgene.2015.00093

The uses of breeding programs for the Pacific white shrimp [*Penaeus (Litopenaeus) vannamei*] based on mixed linear models with pedigreed data are described. The application of these classic breeding methods yielded continuous progress of great value to increase the profitability of the shrimp industry in several countries. Recent advances in such areas as genomics in shrimp will allow for the development of new breeding programs in the near future that will increase genetic progress. In particular, these novel techniques may help increase disease resistance to specific emerging diseases, which is today a very important component of shrimp breeding programs. Thanks to increased selection accuracy, simulated genetic advance using genomic selection for survival to a disease challenge was up to 2.6 times that of phenotypic sib selection.

Keywords: *P. vannamei*, growth, survival rate, disease resistance, genomic selection

Introduction

Shrimp production is an important activity both in economic terms and from the perspective of its contribution to human nutrition. Total world shrimp production had a value of approximately \$13.6 billion USD in 2013. The most important shrimp production regions in the world are located in Asia, principally China, India, Vietnam, Indonesia, and Bangladesh, and in the Americas, primarily in Ecuador, Brazil, and Mexico (Food and Agriculture Organization of the United Nations [FAO], 2014).

Genetic improvement is an important option for increasing profitability in agriculture and aquaculture (Gjedrem et al., 2012). Several shrimp breeding programs have been implemented in a number of countries, some of which have been reviewed by Neira (2010) and Rye (2012).

The increasing importance of disease to shrimp farming worldwide has stimulated research for developing breeding programs for increasing disease resistance/tolerance to disease. Evidence for successful selection for disease resistance exists in shrimp for Taura virus and other diseases (Cock et al., 2009; Lightner et al., 2012). In a few cases, alternative programs which combine mass selection in communally raised animals with recovery of family identification using DNA markers have been used in order to perform mating that avoids excessive increases in inbreeding rates and to select animals in the presence of disease, allowing natural selection to

act toward increased genetic resistance (Rocha, 2012). Other programs have used experimental challenges combined with mass selection in successive generations and obtained increases in the genetic resistance to White Spot Syndrome Virus (Cuéllar-Anjel et al., 2012).

Shrimp Breeding Programs

Although many commercial shrimp breeding programs are not fully described, most are based on population structures typical for aquaculture breeding, using full- and half-sib family structures (Gjedrem et al., 2012). Most shrimp breeding programs focus on the improvement of growth traits and general survival rate. Some have concentrated on selection for specific diseases. These traits can be improved genetically by within-line selection.

The Maricultura del Pacífico hatchery commercial line population was started in 1998 in Mexico, from a heterogeneous population formed by a mix of domesticated lines and wild shrimp from several origins. The main efforts have been directed toward instrumenting a breeding program to develop a genetic line oriented towards improving profitability of biomass production under the production conditions predominating in the northwest part of Mexico. Since biomass production depends on harvesting body weight and survival rates, these two traits have been incorporated as the broodstock selection criteria. The relative importance of each of these traits in the selection index used is 5:1 for harvesting body weight and survival, respectively, which is based on economic studies under the main production system conditions found in Mexico.

Each year from 2003 to 2010, an average of 15,445 shrimp obtained from 130 females and 93 males were evaluated for body weight at harvesting. These evaluations were performed in 2–4 ponds where commercial like production conditions are reproduced. Starting in 2009, survival from 65 to 130 days of age also has been genetically evaluated. Mating is designed based on the use of shrimp from the selected families with the higher selection index values. Hence, broodstock animals come on average, from the upper 27% families in the case of males, and from the upper 53% families in the case of females. In addition to these selection procedures, within-family selection is performed at various growing stages based on approximations to the individual body weight. This is carried out in the genetic nucleus, under strict biosecurity conditions. Animals coming from the within-family selection procedures are the animals ultimately used as the next generation broodstock. Reproduction techniques are based on artificial insemination only, using two females per male. These procedures have been described by Castillo-Juárez et al. (2007) and by Campos-Montes et al. (2013), and some methodological implications have been discussed by Montaldo et al. (2013). Considering the time needed to evaluate animals and to obtain mature broodstock ready to use, the complete production cycle required to yield a new generation (generational interval) is 1 year.

Genetic Progress

Genetic progress has been evaluated in several shrimp selection programs. Hetzel et al. (2000) estimated the selection response per generation for 6-month weight in *Peneaus japonicus* at 8.3%. Andriantahina et al. (2012) estimated the genetic response per generation for 5-month weight in *Litopenaeus vannamei* at 10.7%. The genetic response for body weight at harvesting (130 days of age) in the Maricultura del Pacífico commercial line has been evaluated using linear models (BLUP-animal model). The estimated genetic gain as a linear trend from 2003 to 2010 represents an increment of 18.4% of the average body weight for the period. For survival, the estimated genetic gain as a linear trend from 2004 to 2010 was also positive (1.56%).

Potential of Genomic Selection for Disease Resistance

Diseases are major constraints for aquaculture production. As vaccination is not an option in shrimp and management contention measures are frequently unfeasible, genetic selection is considered a possible option in fighting many diseases in *P. vannamei* and other shrimp species. Genomic selection (GS) increases accuracy if compared to conventional selection, by taking advantage of both between and within-family variance, in situations where family testing for disease is used in sibs of the actual selection candidates, in order to avoid the introduction of the pathogen into the breeding nucleus population (Villanueva et al., 2011).

We used SelAction software (Rutten et al., 2002) and methods developed by Dekkers (2007) to deterministically simulate selection response in *P. vannamei* using dense genetic marker arrays (chips). This was done to assess the potential effects of GS in a breeding program oriented to the improvement of disease resistance. This simulation considered the context of a typical shrimp breeding program based on (sib) family selection. We used seven heritability values: 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, and 0.50 for survival under experimental infestation of shrimp produced in breeding shrimp populations of different size, and relatively high selection intensities, if compared to that from our actual breeding program (Campos-Montes et al., 2013). This allows us to evaluate the possible genetic response for survival to different pathogens, which may include a wide range of viral as well as bacterial infections. The considered proportion of common full-sib environmental effects was 0.15 for all cases.

Accuracy from GS was obtained with the formula developed by Daetwyler et al. (2010), assuming a genome size of 28 Morgans, and an effective population size of 50. This yielded a value for the number of independent chromosome segments close to 324, which may be considered conservative (Villanueva et al., 2011); therefore accuracy from GS was probably not overestimated. The proportion of the additive genetic variance explained by markers to reflect the preliminary stage of development of SNP chips in shrimp was also conservatively set at a value of 0.64 which is below the value of 0.80 used for the 50 K SNP chip for cattle (Daetwyler, 2009).

All breeding populations were derived from 30 males and 38 females, with an incomplete nested structure, similar to that described by Montaldo et al. (2013), to produce 150 families to be tested. Population sizes corresponding to 6, 50, and 100 shrimp per female were 900, 7,500, and 15,000 measured offspring, respectively.

Three breeding strategies were considered here for between family selection: (1) phenotypic information, (2) GS (based on using contemporary training and testing population of similar size to avoid an increase of generation interval), and (3) combined selection program which uses both genomic and phenotypic information.

Results (Figure 1) show that the potential of GS to develop lines with improved disease resistance is high. GS and combined selection programs had large increases in selection responses measured in phenotypic standard deviation units for survival, in populations of size 7,500 and 15,000, but less for the population of size 900. Ratio from combined and GS programs, with respect to phenotypic programs, increased with lower heritability values. These ratios for populations of size 7,500 were 2.6, 1.7, 1.6, 1.5, 1.4, and 1.4 for combined selection for heritability values of 0.01, 0.10, 0.20, 0.30, 0.40, and 0.50, respectively, and only slightly lower for GS. Analogous ratios for combined and GS relative to phenotypic selection for a population of size 15,000 were very similar. Results for combined selection for a population of size 900 showed an advantage over phenotypic selection with lower ratios (from 1.2 to 1.3), while ratios for GS were slightly below 1. These results indicate that with a training population of adequate size (probably of less than 7,500), GS could more than double the rate of selection response for survival to specific diseases, which is higher than previous estimates for GS programs in the context of aquaculture breeding for continuous traits (Nielsen et al., 2011). Interestingly, the relative advantage was greater for smaller heritability values. This is promising for improving survival to disease challenges in aquaculture species, which often have low heritability values (Yáñez and Martínez, 2010).

In this preliminary GS shrimp evaluation, we compared all the programs at the same selection intensity. This may give some unrealistic detriment to GS programs, because GS would allow selection of individual animals at higher selection intensities, when the individual identification of candidates is possible. Therefore, not all the potential advantages of GS and larger candidate populations to increase selection response were included in this study. Combined and GS responses were similar for populations of sizes 7,500 or 15,000, indicating that GS is capturing almost all the variation explained by the phenotypes, making GS more accurate. A more detailed comparison may include changes in inbreeding rates, and the optimization of the different factors affecting selection response. Advances in shrimp genomics (Yu et al., 2014; Zhang et al., 2014) may lead to the future development of SNP chips for *P. vannamei*, making the possibility of performing GS in this species more viable in the near future.

Within the framework of our preliminary calculations, GS for survival rates to disease challenges in shrimp may lead to large increases in selection responses across a wide range of heritability values.

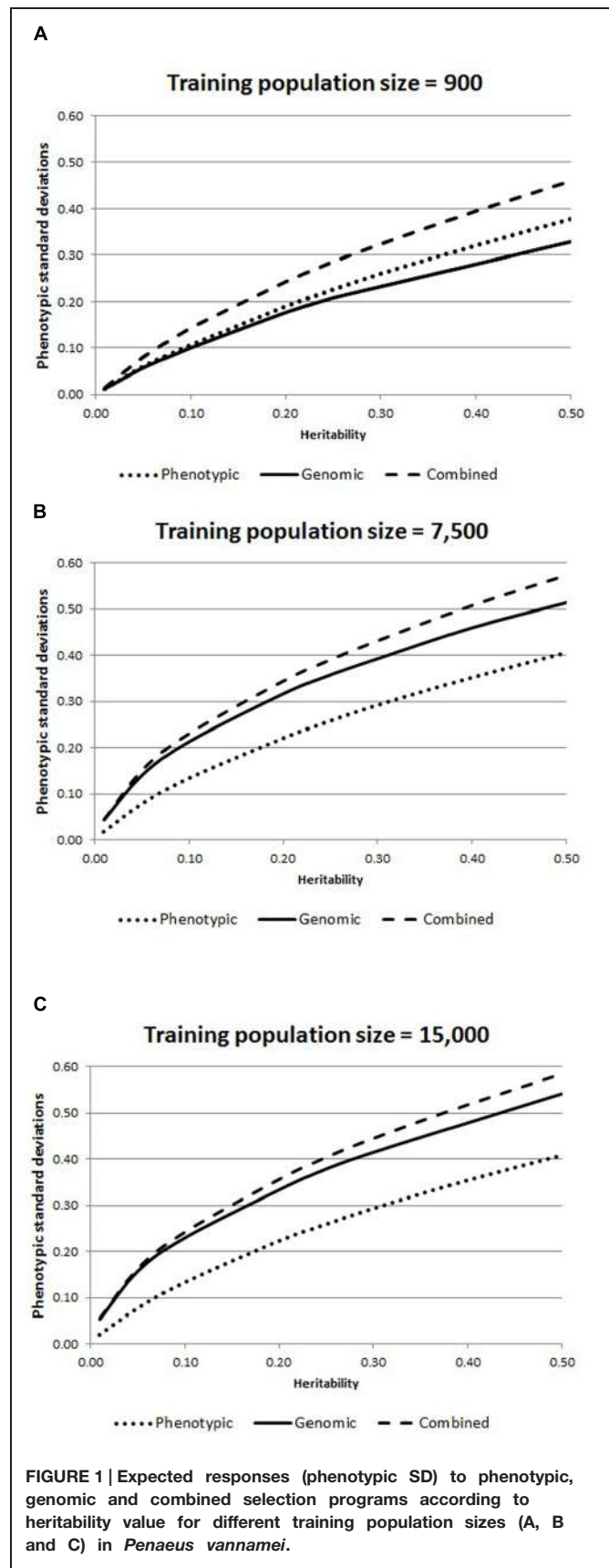


FIGURE 1 | Expected responses (phenotypic SD) to phenotypic, genomic and combined selection programs according to heritability value for different training population sizes (A, B and C) in *Penaeus vannamei*.

Conclusion

The use of the so-called “classic” or conventional methods of quantitative genetics to genetically improve the Pacific white shrimp has allowed for continuous progress of great value to increase the profitability of the shrimp industry in several countries and other aquaculture species. Recent advances in such areas as genomics will allow, in the near future, for the development of animal breeding methods, which may increase, and hence accelerate shrimp genetic progress. In particular, these novel techniques may help

increase disease resistance to specific emerging diseases, which is today a very important aspect in shrimp breeding programs.

Acknowledgments

The authors are thankful to CONACyT, Mexico, for providing funds for our breeding research in the hatchery since 2007. Special thanks to Maricultura del Pacífico workers, for their contribution to operating the breeding program.

References

- Andriantahina, F., Liu, X., Huang, H., and Xiang, J. (2012). Response to selection, heritability and genetic correlations between body weight and body size in Pacific white shrimp, *Litopenaeus vannamei*. *Chin. J. Oceanol. Limnol.* 30, 200–205. doi: 10.1007/s00343-012-1066-2
- Campos-Montes, G. R., Montaldo, H. H., Martínez-Ortega, A., Jiménez, A. M., and Castillo-Juárez, H. (2013). Genetic parameters for growth and survival traits in Pacific white shrimp *Penaeus (Litopenaeus) vannamei* from a nucleus population undergoing a two-stage selection program. *Aquaculture Int.* 21, 299–310. doi: 10.1007/s10499-012-9553-1
- Castillo-Juárez, H., Casares, J. C. Q., Campos-Montes, G., Villela, C. C., Ortega, A. M., and Montaldo, H. H. (2007). Heritability for body weight at harvest size in the Pacific white shrimp, *Penaeus (Litopenaeus) vannamei*, from a multi-environment experiment using univariate and multivariate animal models. *Aquaculture* 273, 42–49. doi: 10.1016/j.aquaculture.2007.09.023
- Cock, J., Gitterle, T., Salazar, M., and Morten, R. (2009). Breeding for disease resistance of Penaeid shrimps. *Aquaculture* 286, 1–11. doi: 10.1016/j.aquaculture.2008.09.011
- Cuéllar-Anjel, J., White-Noble, B., Schofield, P., Chamorro, R., and Lightner, D. V. (2012). Report of significant WSSV-resistance in the Pacific white shrimp, *Litopenaeus vannamei*, from a Panamanian breeding program. *Aquaculture* 368, 36–39. doi: 10.1016/j.aquaculture.2012.08.048
- Daetwyler, H. D. (2009). *Genome-Wide Evaluation of Populations*. Ph.D. thesis, Wageningen University, Wageningen.
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi: 10.1534/genetics.110.116855
- Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124, 331–341. doi: 10.1111/j.1439-0388.2007.00701.x
- Food and Agriculture Organization of the United Nations [FAO]. (2014). *The State of Food and Agriculture 2014*, Rome: Food and Agriculture Organization of the United Nations.
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Hetzel, D. J. S., Crocos, P. J., Davis, G. P., Moore, S. S., and Preston, N. C. (2000). Response to selection and heritability for growth in the Kuruma prawn, *Penaeus japonicus*. *Aquaculture* 181, 215–223. doi: 10.1016/S0044-8486(99)00237-9
- Lightner, D. V., Redman, R. M., Pantoja, C. R., Tang, K. F. J., Noble, B. L., Schofield, P., et al. (2012). Historic emergence, impact and current status of shrimp pathogens in the Americas. *J. Invertebr. Pathol.* 110, 174–183. doi: 10.1016/j.jip.2012.03.006
- Montaldo, H. H., Castillo-Juárez, H., Campos-Montes, G., and Pérez-Enciso, M. (2013). Effect of the data family structure, tank replication and the statistical model, on the estimation of genetic parameters for body weight at 28 days of age in the Pacific white shrimp (*Penaeus (Litopenaeus) vannamei* Boone, 1931). *Aquaculture Res.* 44, 1715–1723. doi: 10.1111/j.1365-2109.2012.03176.x
- Neira, R. (2010). “Breeding in aquaculture species: genetic improvement programs in developing countries,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Leipzig.
- Nielsen, H. M., Sonesson, A. K., and Meuwissen, T. H. E. (2011). Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *J. Anim. Sci.* 89, 630–638. doi: 10.2527/jas.2009-2731
- Rocha, J. L. (2012). “Alternatives to family programs in shrimp breeding: a progeny-testing scheme based on commercial performance,” in *Proceedings of the 9th Biennial Symposium of the Brazilian Society of Animal Breeding*, João Pessoa, Paraíba.
- Rutten, M. J. M., Bijma, P., Woolliams, J. A., and Van Arendonk, J. A. M. (2002). SelAction: Software to predict selection response and rate of inbreeding in livestock breeding programs. *J. Hered.* 93, 456–458. doi: 10.1093/jhered/93.6.456
- Rye, M. (2012). “Current status and prospects for the application of genetic improvement in aquaculture species,” in *Proceedings of the 9th Biennial Symposium of the Brazilian Society of Animal Breeding*, João Pessoa, Paraíba.
- Villanueva, B., Fernández, J., García-Cortés, L. A., Varona, L., Daetwyler, H. D., and Toro, M. A. (2011). Accuracy of genome-wide evaluation for disease resistance in aquaculture breeding programs. *J. Anim. Sci.* 89, 3433–3442. doi: 10.2527/jas.2010-3814
- Yáñez, J. M., and Martínez, V. (2010). Genetic factors involved in resistance to infectious diseases in salmonids and their application in breeding programmes. *Arch. Med. Vet.* 42, 1–13. doi: 10.4067/S0301-732X2010000200002
- Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., et al. (2014). SNP discovery in the transcriptome of White Pacific shrimp *Litopenaeus vannamei* by next generation sequencing. *PLoS ONE* 9:e87218. doi: 10.1371/journal.pone.0087218
- Zhang, X., Yu, Y., Li, F., Xiang, J., and Liu, D. (2014). “Construction of a high-density SNP genetic map for Pacific White shrimp *Litopenaeus vannamei*,” in *Proceedings of the Plant and Animal Genome XXII Conference*, Plant and Animal Genome, San Diego, CA.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Castillo-Juárez, Campos-Montes, Caballero-Zamora and Montaldo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

