

# Genomic Predictions and Genome-Wide Association Study of Resistance Against *Piscirickettsia salmonis* in Coho Salmon (*Oncorhynchus kisutch*) Using ddRAD Sequencing

Agustín Barría,<sup>\*,†</sup> Kris A. Christensen,<sup>\*,1</sup> Grazyella M. Yoshida,<sup>\*,§</sup> Katharina Correa,<sup>\*,\*\*</sup> Ana Jedlicki,<sup>\*</sup> Jean P. Lhorente,<sup>\*\*</sup> William S. Davidson,<sup>‡</sup> and José M. Yáñez<sup>\*,\*\*††,2</sup>

<sup>\*</sup>Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, La Pintana, Santiago 8820808, Chile, <sup>†</sup>Doctorado en Acuicultura, Programa Cooperativo Universidad de Chile, Universidad Católica del Norte, Pontificia Universidad Católica de Valparaíso, 8820808 Chile, <sup>‡</sup>Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada, <sup>§</sup>Animal Science Department, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências Agrárias e Veterinárias, Campus Jaboticabal, Jaboticabal 14884-900, Brazil, <sup>\*\*</sup>Aquainnovo S.A., Puerto Montt 5503032, Chile, and <sup>††</sup>Núcleo Milenio INVASAL, Concepción 4070386, Chile

ORCID IDs: 0000-0002-2813-4559 (A.B.); 0000-0003-0021-0326 (K.A.C.); 0000-0002-6788-7369 (G.M.Y.); 0000-0003-1386-8522 (K.C.); 0000-0003-0626-1247 (A.J.); 0000-0002-9157-4231 (J.P.L.); 0000-0002-3002-3225 (W.S.D.); 0000-0002-6612-4087 (J.M.Y.)

**ABSTRACT** *Piscirickettsia salmonis* is one of the main infectious diseases affecting coho salmon (*Oncorhynchus kisutch*) farming, and current treatments have been ineffective for the control of this disease. Genetic improvement for *P. salmonis* resistance has been proposed as a feasible alternative for the control of this infectious disease in farmed fish. Genotyping by sequencing (GBS) strategies allow genotyping of hundreds of individuals with thousands of single nucleotide polymorphisms (SNPs), which can be used to perform genome wide association studies (GWAS) and predict genetic values using genome-wide information. We used double-digest restriction-site associated DNA (ddRAD) sequencing to dissect the genetic architecture of resistance against *P. salmonis* in a farmed coho salmon population and to identify molecular markers associated with the trait. We also evaluated genomic selection (GS) models in order to determine the potential to accelerate the genetic improvement of this trait by means of using genome-wide molecular information. A total of 764 individuals from 33 full-sib families (17 highly resistant and 16 highly susceptible) were experimentally challenged against *P. salmonis* and their genotypes were assayed using ddRAD sequencing. A total of 9,389 SNPs markers were identified in the population. These markers were used to test genomic selection models and compare different GWAS methodologies for resistance measured as day of death (DD) and binary survival (BIN). Genomic selection models showed higher accuracies than the traditional pedigree-based best linear unbiased prediction (PBLUP) method, for both DD and BIN. The models showed an improvement of up to 95% and 155% respectively over PBLUP. One SNP related with B-cell development was identified as a potential functional candidate associated with resistance to *P. salmonis* defined as DD.

## KEYWORDS

selective breeding genotyping by sequencing *Oncorhynchus kisutch* disease resistance GWAS Genomic Selection GenPred Shared Data Resources

Chile is the largest producer of coho salmon (*Oncorhynchus kisutch*) globally, reaching about 160,000 tons in 2014, representing more than 90% of total production (FAO 2016). However, the success and sustainability of this industry is constantly threatened by infectious diseases, including Salmon Rickettsial Syndrome (SRS). This disease is caused by *Piscirickettsia salmonis*, a gram-negative and facultative intracellular bacteria, which was isolated for the first time in Chile in coho salmon (Cvitanich *et al.* 1991). Data from the Chilean National Fisheries and

Aquaculture Service (Sernapesca) indicates that during the first half of 2016, 53% of the mortalities ascribed to infectious diseases in coho salmon were associated with SRS (Sernapesca 2016). To date, control measures and treatments for SRS are based on antibiotics and vaccines. However, both strategies have not had the expected effectiveness under field conditions (Rozas and Enríquez 2014).

Because of this, it is necessary to develop alternative strategies for the control of this disease (Yáñez *et al.* 2014a). In this regard, breeding for

enhanced disease resistance is a feasible and sustainable option to improve animal health, welfare and productivity (Stear *et al.* 2001). A primary requisite for including disease resistance into a breeding program is the presence of significant additive genetic variation for the trait (Falconer and Mackay 1996). Commonly, data to evaluate resistance comes from experimental challenges carried out using siblings of the selection candidates (Ødegård *et al.* 2011; Yáñez *et al.* 2014a). Quantitative studies have estimated significant genetic variation for resistance against different pathogens in salmonid species (Ødegård *et al.* 2011; Yáñez *et al.* 2014a). For instance, low to moderate heritabilities for resistance against *P. salmonis* in Atlantic salmon (*Salmo salar*) ( $h^2 = 0.11$  to  $0.41$ ) (Yáñez *et al.* 2013; Yáñez *et al.* 2014b) and coho salmon ( $h^2 = 0.16$ ) (Yáñez *et al.* 2016a) have been estimated.

Marker assisted selection (MAS) can improve production traits in cases where the phenotypes are difficult to measure in the selected candidates (*e.g.*, disease resistance traits) and the total additive genetic variance explained by genetic markers is high (Hayes and Goddard 2010). This methodology has been successfully applied for the improvement of resistance against the Infectious Pancreatic Necrosis Virus (IPNV) in Atlantic salmon, which is controlled by a major quantitative trait locus (QTL) (Houston *et al.* 2012; Moen *et al.* 2015). In the case of polygenic traits, genomic selection (GS) (Meuwissen *et al.* 2001) can significantly improve selection accuracy of breeding values compared to traditional selection, and therefore enhance the response of selection for disease resistance in salmonid species (Tsai *et al.* 2016; Vallejo *et al.* 2016, 2017a; Bangera *et al.* 2017; Correa *et al.* 2017; Yoshida *et al.* 2017).

Genotyping by sequencing (GBS) is an alternative for genotyping in cases when SNP panels are not available. This approach reduces the complexity of the genome, and can be used to identify thousands of markers without prior marker discovery efforts or a reference genome. Currently, several approaches of GBS have been developed, significantly reducing the cost and labor (Baird *et al.* 2008; Elshire *et al.* 2011; Peterson *et al.* 2012). These methodologies have been widely used in salmonid species, to generate dense linkage maps (Briec *et al.* 2014; Gonen *et al.* 2014), perform association studies to identify genomic regions involved in the resistance against pathogens (Campbell *et al.* 2014; Liu *et al.* 2015; Palti *et al.* 2015b) and generate SNPs resources (Houston *et al.* 2012).

Double-digest restriction-site associated DNA (ddRAD) reduces DNA complexity by digesting DNA with two restriction enzymes (REs) simultaneously, without random shearing (Peterson *et al.* 2012). This approach has been widely used in genetic studies in aquaculture species (reviewed in Robledo *et al.* 2017).

In the present study, we used ddRAD sequencing to dissect the genetic architecture of resistance against *P. salmonis* in a farmed coho salmon population and identify molecular markers associated with the trait. Furthermore, GS models were used to evaluate the potential to accelerate the genetic improvement of resistance against *P. salmonis* in this coho salmon population by means of using genome-wide molecular information.

## MATERIALS AND METHODS

### Coho salmon breeding population

The coho salmon population used in the present study came from a unique 2012 year-class population. This population belongs to a genetic improvement program that was established in 1997 and is owned by Pesquera Antares and managed by Aquainnovo (Puerto Montt, Chile). Further details about this breeding population, in terms of reproductive management, rearing conditions, fish tagging and breeding objectives are described by Yáñez *et al.* (2014c; 2016a) and Dufflocq *et al.* (2016).

### Experimental challenge

The experimental challenge against *P. salmonis* was performed as it is described in detail by Correa *et al.* (2015a) and Yáñez *et al.* (2016a). Briefly, 2,606 individuals, belonging to 107 maternal full-sib families and 60 paternal half-sib families, were challenged against *P. salmonis*. Prior to the experiment, each fish was tagged with a passive integrated transponder (PIT-tag), placed in the abdominal cavity for genealogy traceability during the challenge test. The *P. salmonis* challenge was performed at Aquainnovo's research station, located in Lenca River, X Region, Chile.

For the lethal dose 50 (LD<sub>50</sub>) calculation, a random sample of 80 fish were selected from the population. Four different dilutions from the *P. salmonis* inoculum were evaluated (1/10, 1/100, 1/1000 and 1/10000). Twenty fish were challenged at each dilution. The dilutions were intraperitoneally (IP) injected with a volume of 0.2ml/fish. Daily mortality was recorded. This preliminary test spanned 26 days and a dilution of 1:680 was estimated as the LD<sub>50</sub>.

For the main challenge, fish were distributed into three tanks (7m<sup>3</sup>) with a salt water concentration of 31 ppt. An average of eight individuals (ranging from 1 to 18) from each of the 107 families were distributed into each tank. The experimental challenge was performed through an intraperitoneal (IP) injection with 0.2ml/fish of the LD<sub>50</sub> inoculum. The average weight of the fish at the inoculation was 279g (SD = 138g). To ensure that these fish were free from other pathogens, qRT-PCR was previously performed in order to control for the presence of Infectious Salmon Anemia Virus (ISAV), IPNV and *Flavobacterium spp.*

The *P. salmonis* challenge continued for up to 50 days post IP injection. Throughout the challenge, environmental parameters (pH, temperature, salinity and oxygen) were measured and controlled. Fish were removed from the tanks after death, and a sample of the anterior kidney was taken and stored at -80° in RNALater. A necropsy assay was performed in conjunction with qRT-PCR to confirm the cause of death and the presence of *P. salmonis*. This was also done to control for the presence of other pathogens, such as *Vibrio ordalii*, *Renibacterium salmoninarum* and IPNV.

### ddRAD library preparation and sequencing

Ten ddRAD libraries were produced by multiplexing 828 individuals following the protocol described by Peterson *et al.* (2012). For this, 64 parents (males and females) and 764 offspring representing the

Copyright © 2018 Barria *et al.*

doi: <https://doi.org/10.1534/g3.118.200053>

Manuscript received September 26, 2017; accepted for publication February 1, 2018; published Early Online February 22, 2018.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental Material is available online at [www.g3journal.org/lookup/suppl/doi:10.1534/g3.118.200053/-/DC1](http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.118.200053/-/DC1)

<sup>1</sup>Present address: Marine Ecosystems and Aquaculture Division, Science Branch Fisheries and Oceans Canada/Government of Canada, 4160 Marine Drive, West Vancouver, BC, Canada V7V 1N6.

<sup>2</sup>Corresponding author: Facultad de Ciencias Agronómicas, Universidad de Chile, Avenida Santa Rosa 11735, 8820808, La Pintana, Santiago, Chile. E-mail: [jmayanez@uchile.cl](mailto:jmayanez@uchile.cl)

17 most resistant and 16 most susceptible families were selected. An average of 23 (ranging from 11 to 43 individuals) offspring per family were chosen. Briefly, total DNA was extracted using the commercial kit Wizard SV Genomic DNA purification System (Promega) according to the manufacturer's protocol. Between 80 and 200 ng of DNA, from each individual was digested with two restriction enzymes (New England Biolabs, UK; NEB); 10 U of *SbfI* (specific for the CCTGCA|GG motif) and *MseI* (specific for the T|TAA motif) in a 12  $\mu$ l reaction volume, including 1  $\mu$ l of *SbfI* and *MseI* adapter (8.3 pM), for 90 min at 37°. The ligation reaction was carried out by adding 1  $\mu$ l of T4 ligase (NEB) diluted 1:100 in T4 buffer and incubating for 150 min at 37° and subsequently at 16° overnight.

Each ligation mix was diluted with 189  $\mu$ l of dilute TE buffer (1:10). Kodak DNA Polymerase (ABM), a high-fidelity polymerase, was used to amplify DNA fragments with the correct adapters. PCR reactions (20  $\mu$ l) were prepared containing 10  $\mu$ l of PCR mix 2x, 1  $\mu$ l of primer mix (10  $\mu$ M each), 6  $\mu$ l of diluted ligation mix and 3  $\mu$ l of nuclease-free water. Each sample was PCR amplified using the following conditions: 95° for 2 min, followed by 17 cycles of 95° for 20s, 66° for 30s and 68° for 40s. After PCR, amplicon quality was checked by loading 5  $\mu$ l on a 2% agarose gel. Subsequently, samples were pooled, so that the final concentration was similar among them within each library. Each library was concentrated through an evaporation step for 80 min in a Centrивap Mobile Console Centrifugal Evaporator (Labconco). This step was conducted until 300  $\mu$ l of the generated library was obtained. Final volume of each library was loaded on a 1% agarose gel. Size of the bands selected for sequencing ranged from 750 and 1,500 bp and between 1,800 and 2,500 bp. DNA was purified through the QIAquick gel extraction kit (Qiagen) following manufacturer's instructions. Finally, libraries were sequenced on an Illumina HiSeq2500 platform, using 150 base single-end.

### SNP identification

Raw sequence reads obtained from Illumina sequencing were analyzed using STACKS v. 1.41 (Catchen *et al.* 2011, 2013). This software was specifically developed to analyze short-read data generated through next generation sequencing (NGS) (Davey *et al.* 2013).

Sample reads were trimmed to 134 bp for all subsequent analyses, demultiplexed and filtered using process\_radtags. Rad-tags which passed the quality filter were aligned to the *Oncorhynchus kisutch* reference genome (GenBank: MPMKV00000000.1) using BWA v. 0.7.12 (Li and Durbin 2009). The reference genome was indexed (using the index function) and alignments were performed using the mem algorithm; all parameters were set as default. Loci were then built using pstacks with a minimum depth of coverage of three to build a locus (-m 3).

A catalog of loci was constructed using the cstacks program using only the parents' loci from pstacks. To build the catalog, the maximum number of mismatches allowed between sample tags was set to three (-n 3), and the matching was based on genomic location (g). After catalog construction, the sstacks program was used in order to match rad-tags against the catalog based again on genomic location (g), followed by the populations software, using defaults parameters. Loci were considered as valid if they were present in at least 75% of the individuals of the population. As a quality control step, the following parameters were used to filter low-confidence SNPs: Minor Allele Frequency (MAF)  $\leq$  0.05, Hardy-Weinberg Equilibrium (HWE)  $P < 1 \times 10^{-6}$  and genotyping call rate  $< 0.80$ . Individuals were removed from the analyses if their genotyping call rates were below 0.70.

### Trait definitions

Resistance against *P. salmonis* was defined as the day of death (DD) with values ranging from 1 to 50 depending on the time of death. Additionally, resistance was also evaluated as a binary (BIN) trait, either dead or alive at the end of the challenge. Values for this trait were 1 in cases where the fish died during the challenge, or 0 if the fish survived until the end of the challenge. Initial Body Weight (IW) for each fish, was measured prior to the IP injection.

### Pedigree-based BLUP

All challenged individuals (n = 2,606) were used for the pedigree-based approach, PBLUP, as a control for the performance evaluation of genomic predictions. A linear univariate animal model was used to estimate variance components and predict Estimated Breeding Values (EBVs) for DD, while for BIN a univariate threshold animal model was fitted (Table 1). The model used was as follows:

$$y = X\beta + Tp + e$$

In the previous equation,  $y$  is a vector of phenotypes (BIN or DD),  $\beta$  is a vector of fixed effects (sex and tank as factors, and initial weight as covariate),  $p$  is a vector of random additive polygenic genetic effects that follows a normal distribution  $\sim N(0, A\sigma_p^2)$ ,  $X$  and  $T$  are incidence matrices,  $A$  is the additive relationship matrix, and  $e$  is the random residual (Lynch and Walsh 1998). Both models were fitted using the BLUPF90 set of programs (Misztal *et al.* 2016) by means of the AIR-EMLF90 and THRGIBBS1F90 modules to analyze DD and BIN, respectively. The MCMC Gibbs sampling scheme set for running THRGIBBS1F90, included a total of 200,000 iterations. The first 20,000 were discarded as burn-in iterations, and then from the remaining 180,000 samples, one from every 50 samples was saved for analysis. This Gibbs sampling scheme collected 3,600 independent samples for analysis.

Heritabilities for PBLUP models were computed as follows:

$$h_i^2 = \frac{\sigma_{ai}^2}{\sigma_{ai}^2 + \sigma_{ei}^2}$$

where  $\sigma_{ai}^2$  and  $\sigma_{ei}^2$  are the additive genetic and residual variances for each trait. In the case of BIN, the residual variance was set to 1.

### Genomic BLUP

The SNP based variance components and GEBVs were estimated using genomic BLUP (GBLUP), similar to the PBLUP model described above, and implemented in the BLUPF90 software. The GBLUP is a modification of the PBLUP method, where  $g$  is a vector of random additive genetic polygenic effects with a distribution  $\sim N(0, G\sigma_g^2)$  and the numerator relationship matrix  $A$  is replaced by a genomic relationship matrix  $G$ , as described by (VanRaden 2008). Only genotyped animals, which passed all quality controls (n = 580) were analyzed.

### Single step genomic GBLUP

The single-step GBLUP (ssGBLUP) and weighted single-step GBLUP (wssGBLUP) models were similar to the PBLUP model except for a combined genomic and pedigree relationship. The kinship matrix used was  $H$  (Aguilar *et al.* 2010), in which genotype and pedigree data are combined. The inverse of the matrix  $H$  is:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

■ **Table 1** Estimated genetic parameters and accuracy of breeding values (EBV) estimation for resistance against *P. salmonis* using a pedigree-based model

Phenotype <sup>a</sup>	$\sigma_a^2$	$\sigma_e^2$	$h^2$ (SE)	Accuracy (R)
DD	12.55	77.60	0.14(0.034)	0.271
BIN	0.38	1.00	0.27 (0.043)	0.316

<sup>a</sup>The BLUP analysis included the phenotype of all the progeny of 107 families challenged against *Piscirickettsia salmonis* (n = 2,606)

where  $A^{-1}$  is the inverse numerator relationship matrix for all animals,  $A_{22}^{-1}$  is the inverse of a pedigree-based relationship matrix for genotyped animals only, and  $G^{-1}$  is the inverse genomic relationship matrix. SNPs were equally weighted and given an initial value of one in the ssGBLUP method. In the wssGBLUP method, the marker variances were used as weights. The marker variances were estimated from allele frequencies, and from marker effects that were calculated in the ssGBLUP method (Wang *et al.* 2014). The DD trait was analyzed as a linear trait using AIREMLF90 and BLUPF90, whereas, BIN was analyzed as a threshold trait with THRGIBBS1F90 in the BLUPF90 family of programs (Misztal *et al.* 2016). The MCMC Gibbs scheme for the estimation of the genetic parameters for BIN for ssGBLUP and wssGBLUP, were estimated identically as described above. The ssGBLUP and wssGBLUP models included all the genotyped animals which passed quality control (n = 580), and all the phenotyped fish (n = 2,606) from 107 families.

### Bayes C

The Bayes C (Habier *et al.* 2011) analyses were performed using GS3 software. A total of 200,000 iterations were used in the Gibbs sampling, with a burn-in period of 20,000 cycles. The results were saved every 50 cycles. The number of samples in this analysis totaled 4,000. Convergence and autocorrelation were assessed by visual inspection of trace plots of the posterior variance components. The adjusted model can be described, in matrix notation, as follows:

$$y = X\beta + Tp + \sum_{i=1}^n g_i a_i \delta_i + e$$

where  $y$  is the vector of phenotypic records (DD or BIN),  $X$  is an incidence matrix of fixed effects (sex and tank as factors and IW as covariate),  $\beta$  is the vector of fixed effects,  $T$  is an incidence matrix of polygenic effects,  $p$  is a random vector of polygenic effects of all individuals in the pedigree,  $g_i$  is the vector of the genotypes for the  $i^{th}$  SNP for each animal,  $a_i$  is the random allele substitution effect of the  $i^{th}$  SNP,  $\delta_i$  is an indicator variable (0, 1) sampled from a binomial distribution with parameters determined such that 1% of the markers were included in the model, and  $e$  is a vector of residual effects. The following prior distributions were assumed for the genetic random effects: Independent and identical mixture distributions for the SNP effects; each SNP has a point mass at zero having a probability  $\pi$  and a univariate normal distribution with a probability of  $1 - \pi$  with null mean and variance  $\sigma_a^2$ ; which in turn has a scaled inverse chi-squared prior with  $\nu_a = 4$  degrees of freedom and scale parameter  $s_a^2$  (or  $s_e^2$ ) (Fernando and Garrick 2013). The scale parameter was estimated as a function of the genetic variance population, based on the mean SNP allele frequency and number of markers assumed with nonzero effects (Fernando *et al.* 2007). Only genotyped animals, which passed quality control (n = 580) were used.

### Genomic prediction accuracy

The different models were compared using a fivefold cross validation scheme. To reduce stochastic effects of sampling, the cross-validation analysis was replicated ten times. Briefly, all challenged individuals (genotyped, phenotyped, or both), were randomly separated into five validation sets. For each set, predictions were made by masking the animals' phenotypes and using the remaining fish as a training set to estimate the marker effects. Thus, for each cross-validation run, the dataset was split into a training set (80%) and a validation set (20%). Accuracy was used to assess the performance of each model and was estimated as follows:

$$r_{EBV,BV} = \frac{r_{EBV,y}}{h}$$

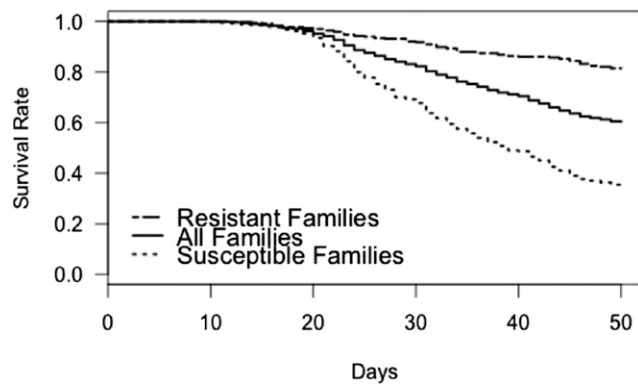
where  $r_{EBV,y}$  is the correlation between the EBV of a given model (predicted for the validation set using information from the training set) and the actual phenotype, while  $h$  is the square root of the pedigree-based estimate of heritability (Correa *et al.* 2017; Legarra 2008; Palaokostas *et al.* 2016; Tsai *et al.* 2016). Finally, accuracies were calculated for each model and compared to those obtained with the PBLUP model.

### Genome-wide association study

In order to identify associations between genetic markers and *P. salmonis* resistance, as DD or BIN, four genome-wide association methodologies were performed using the BLUPF90 set of programs (Misztal *et al.* 2016). GBLUP, ssGBLUP, wssGBLUP and w3ssGBLUP models were used to analyze the DD and BIN traits, using a linear and a threshold model respectively (as described above in model 1). For the GBLUP model, the pedigree-based relationship matrix ( $A$ ) was replaced by a genomic matrix ( $G$ ). For the ssGBLUP, wssGBLUP and w3ssGBLUP GWAS models the  $H$  matrix was used as described above. SNPs were weighted equally (and given a weight of one), for the ssGBLUP model. For the wssGBLUP and w3ssGBLUP methods, weights were determined based on individual marker variances that were estimated using both the marker effects, which were calculated previously for the ssGBLUP model, and marker allele frequencies (Wang *et al.* 2014). For the GBLUP GWAS model, only genotyped animals, which passed quality control (n = 580), were analyzed. For the ssGBLUP, wssGBLUP and w3ssGBLUP GWAS models, all the genotyped animals passing quality control (n = 580), and all the phenotyped fish (n = 2,606) were used. Due to practical reasons, the molecular markers anchored to the different scaffolds not placed in chromosomes, were assigned as chromosome 31. The parents of the challenged individuals were not included in the GWAS analysis because they did not have associated phenotype information, as they were not submitted to the challenge experiment.

### Data availability

Table S1 contains genotypic data (available at the public dryad digital repository <https://doi.org/10.5061/dryad.b273q6p>), Table S2 contains phenotypic data, and Table S3 contains the pedigree information. Table S4 contains the full list of genes located within the top ten 1-Mbp windows proximate to each SNP associated with *P. salmonis* resistance for DD and BIN identified through wssGBLUP. Table S5 contains information of the top ten markers which explain the highest percentage of the genetic variance for each method and trait. Table S6 contains results from the 10 replicated CV.



**Figure 1** Kaplan-Meier curves for *Piscirickettsia salmonis* experimental challenge in coho salmon. Average mortality curves for the 107 full-sib families, and the 17 best and 16 worst families.

## RESULTS

### Challenge test

Mortality began on the 10<sup>th</sup> day after the *P. salmonis* challenge, with evident symptoms of SRS and pathological lesions typical of SRS. These signs include swollen kidney, splenomegaly and yellowish liver tone and coloration (Rozas and Enríquez 2014). Challenged families showed considerable phenotypic variation for *P. salmonis* resistance. Average mortality of all the 107 families reached 38.53% during the 50-day challenge. The average cumulative mortality rate among the 17 best and 16 worst families, selected for genotyping, reached 19% and 63%, respectively (Figure 1).

### ddRAD sequencing

Prior to quality control (QC), per base quality (Phred score) was evaluated. The average quality score ranged from 36 to 38 among libraries, indicating high quality of data. Illumina sequencing, including parents, yielded an average of 156,058,078 ( $\pm$  16 million) of raw sequences per library. After initial QC, which included the removal of low quality sequences and reads with either missing or ambiguous barcodes, an average of 31,660,024 of the reads were removed. In parallel with QC, reads were trimmed to 134 bp. Thus, 79% of the raw reads were retained for subsequent analysis. To create a set of all possible alleles in the population, data sets of the parental samples were used to create a STACKS catalog. This catalog consisted of 106,309 unique ddRAD loci from which 20,068 markers from 757 individuals were identified. Quality filtering reduced this to 9,389 putative bi allelic SNPs (see Table S1) with an average sequencing depth of 38x ranging from 11x to 501x. These markers were identified segregating along the genome of 580 individuals (see Table S1).

### Genome-wide association analysis

Four genome-wide association methodologies were performed either for DD and BIN. These approximations include GBLUP, ssGBLUP, wssGBLUP and w3ssGBLUP. For both traits, all the models showed a similar association pattern. In the case of the ssGBLUP methodology, the GWAS plots become less noisy as the iterations progress, and the peaks associated with the traits become more distinct (Figure 2 and Figure 3). For DD, a marker potentially associated with *P. salmonis* resistance was located on chromosome 11 (Figure 2). This marker was identified in all of the four models and was within the top ten markers explaining most of the percentage of the genetic variance (Table 2).

The availability of a high quality coho salmon reference genome (GenBank accession number MPVK00000000.1), made it possible to identify genes near this marker. Within  $\sim$ 55 Kbp of this marker is the *phosphoinositide-3-kinase adaptor protein 1* (*pik3ap1*) gene; a gene related with innate host defense through B-cells development (Aiba *et al.* 2008; Herzog *et al.* 2009).

In the case of BIN, two molecular markers, explaining most of the genetic variance, were identified in all of the four models (Table 2). One of these markers is located on chromosome 29, while the second marker was identified only at a scaffold level (Scaffold04124) (Figure 3). For BIN, host immune response related genes were not found proximate to any of the suggestive molecular markers.

However, some genes within a 1-Mb window of these markers have been suggested to be involved with *P. salmonis* infection. The *14 KDa Phosphohistidine phosphatase-like* (PHPT1), *gelsolin-like* (GSN), and *glutamine synthase-like* (GS) genes are located on chromosome 29 and near associated markers. *Claudin-10* was found on scaffold04124. These genes have been previously identified as being up-regulated in *Salmo salar* individuals with low susceptibility to *P. salmonis* (Pulgar *et al.* 2015). Moreover, the *retinoic acid receptor RXR-alpha-A-like* (RXRA) gene, located on chromosome 29, has previously been identified as a molecular biomarker for *P. salmonis* infections, and has been found to be down-regulated in macrophages during infection (Rise *et al.* 2004).

A full list of genes that are located within a 1-Mbp window proximate to the suggestive markers associated with *P. salmonis* resistance for Okis11, Okis29 and at a scaffold level, identified through wssGBLUP model, is shown in Table S4. In the case of the marker located on the scaffold, the surrounding sequence was blasted against the *Salmo salar* reference genome (NC\_027300.1).

### Genetic parameters and predictions

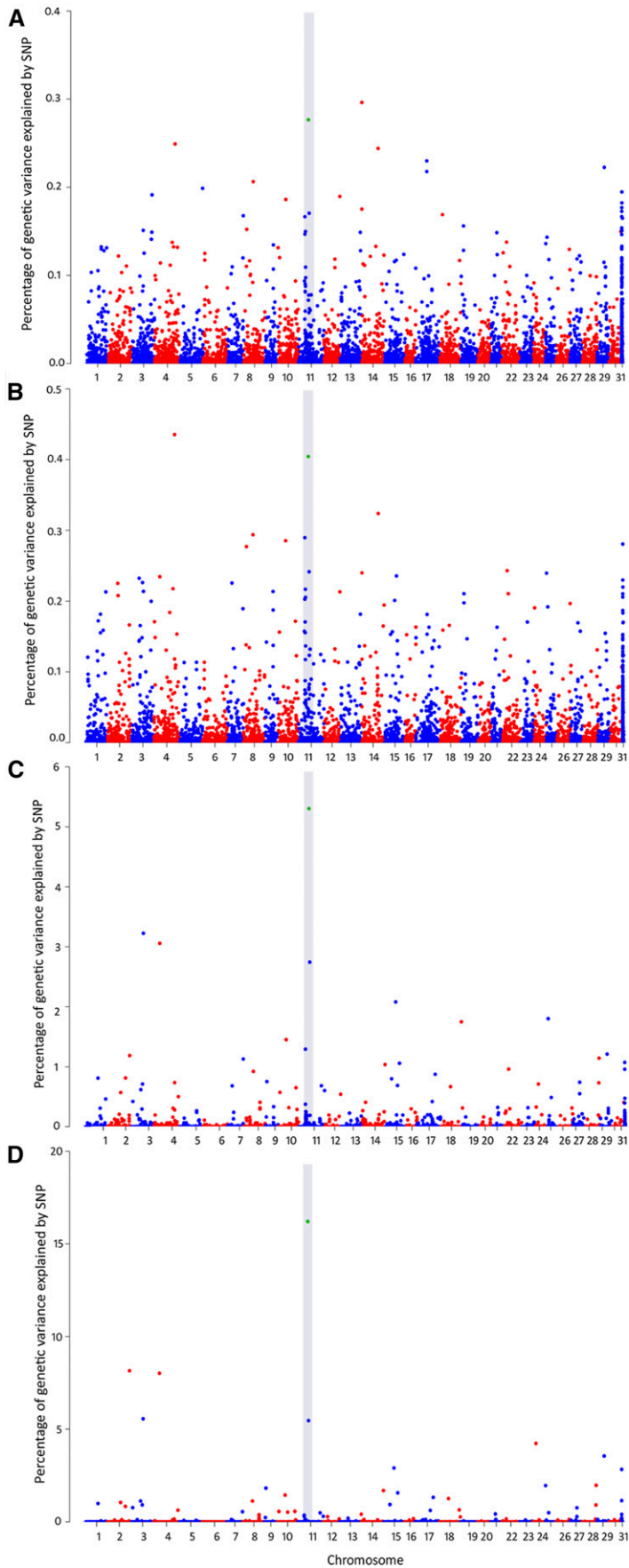
Significant additive genetic variation was estimated for both DD and BIN when using all the data from challenged individuals from the 107 maternal, full-sib families (Table 1). Using the pedigree-based model without genomic data, estimates of the narrow sense heritability for DD and BIN were equal to 0.14 ( $\pm$  0.034) and 0.27 ( $\pm$  0.043), respectively.

Based on a fivefold cross validation, the accuracy of the PBLUP model was slightly lower for DD (0.271) than for BIN (0.316) (Figure 4). When genomic data were included, accuracies for DD and BIN were higher than those achieved using only phenotypic data. However, there is considerable variation between models and trait definitions. The accuracies for the different models ranged from 0.299 (ssGBLUP) to 0.529 (GBLUP) for the DD trait, and from 0.314 (ssGBLUP) to 0.807 (GBLUP) for the BIN. For DD, all the models with genomic data outperformed the pedigree-based model. The relative increase in accuracy ranged from 10 (ssGBLUP) to 95% (GBLUP) (Figure 4). In the case of BIN, the relative increase in accuracy ranged from 20 (wssGBLUP) to 155% (GBLUP). However, one of the genomic models (ssGBLUP) had a similar accuracy to the PBLUP model, with a relative accuracy 1% lower than the pedigree-based model.

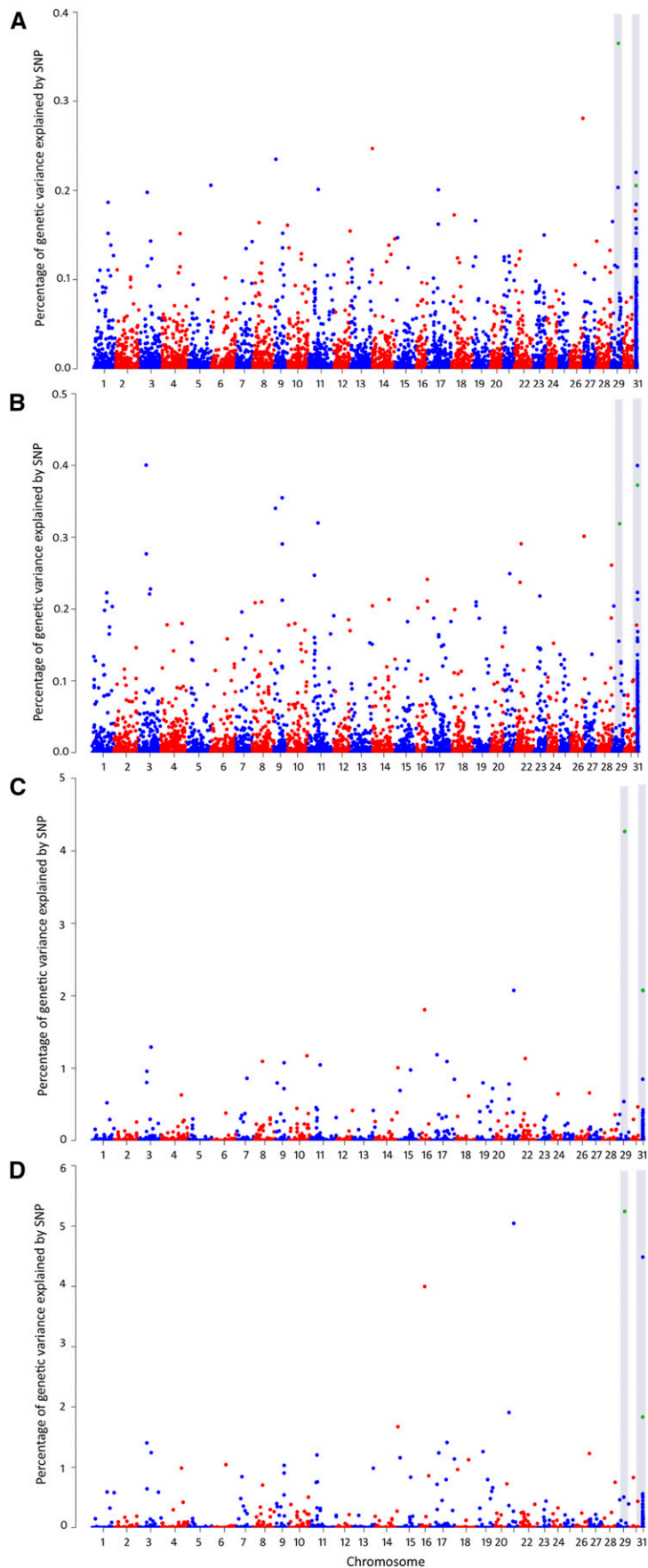
Interestingly, the accuracies obtained with GBLUP were higher than ssGBLUP and wssGBLUP for both traits. These accuracy values reached 0.529, 0.299 and 0.417 for DD and 0.807, 0.314 and 0.3797 for BIN. For both traits, the models with better performance were, GBLUP > Bayes C > wssGBLUP > ssGBLUP.

## DISCUSSION

Significant genetic variation for *P. salmonis* resistance was detected in the present study. Moderate heritabilities were estimated using different



**Figure 2** Genomic association analyses for resistance against *Piscirickettsia salmonis* in a coho salmon population (defined as day of death) for four different models; GBLUP (A), ssGBLUP (B), wssGBLUP (C) and w3ssGBLUP (D). The gray area highlights the SNPs (in green), which are among the top ten markers explaining a high percentage of the genetic variance in the four models.



**Figure 3** Genomic association analyses for resistance against *Piscirickettsia salmonis* in a coho salmon population (for survival as a binary trait) for four different models; GBLUP (A), ssGBLUP (B), wssGBLUP (C) and w3ssGBLUP (D). The gray area highlights SNPs (in green) that were among the top ten markers explaining a high percentage of the genetic variance in the four models.

■ Table 2 Top ten markers associated with *Piscirickettsia salmonis* resistance defined as DD and BIN in coho salmon, using wssGBLUP method

Ranking	Name	Chr <sup>c</sup>	Pos (BP) <sup>d</sup>	PEV <sup>e</sup>	Genes <sup>f</sup>
<b>DD</b>					
1	<b>24987_127</b>	<b>11<sup>a</sup></b>	<b>30525921</b>	<b>5.302</b>	PIK3AP1, TIAL1, PCBD1
2	6135_83	3	37136738	3.225	NOXA1, UBAC1, KLHL20
3	7914_47	4	18281619	3.056	LRP5, NTRK3, KLHL25
4	25096_120	11	33047788	2.744	VA, INPP5A, CSAD
5	34697_43	15	32261231	2.081	CCDC153, KMT2A, LXN
6	52922_94	25	656773	1.802	NHLRC2, NRG3, LRRC4
7	41979_18	18	61818285	1.750	ROBO2, KCNJ1, LCE
8	22393_114	10	24996755	1.451	SLC34A2, SH3RF1, FYB
9	24553_70	11	19605173	1.294	VOS41, CHMP5, FAM49B
10	58185_41	29	22363292	1.211	TSC1, GFI1B, STOM
<b>BIN</b>					
1	<b>58185_41</b>	<b>29<sup>a</sup></b>	<b>22363292</b>	<b>4.270</b>	PHPT1, GSN, GS
2	<b>66451_65</b>	<b>31<sup>ab</sup></b>	<b>12406</b>	<b>2.076</b>	PSMD14, CLDN10, CTSM
3	68326_79	31 <sup>b</sup>	211676	2.073	ROBO1, PIK3CB, KCNJ1
4	45949_127	21	17289066	2.073	FLVCR1, VTA1, HIVEP2
5	36367_15	16	15005555	1.807	SEC24D, KACNIP4, MYOZ2
6	6135_83	3	37136738	1.291	NOXA1, UBAC1, KLHL20
7	37641_86	17	19982735	1.184	NR0B2, HIVEP3, EDN2
8	23665_61	10	55479897	1.170	HDAC5, CADM1, ICAM1
9	47149_112	22	17215562	1.133	PPARA, CDKN1B, KCNQ1
10	18750_95	8	26751149	1.094	GRID2, SMARCAD1, TSPAN3,

<sup>a</sup> Markers in common within the top ten along the four models.

<sup>b</sup> *Salmo salar* used as reference specie.

<sup>c</sup> Chromosome.

<sup>d</sup> Position in coho salmon reference genome.

<sup>e</sup> Percentage of Phenotypic variance.

<sup>f</sup> Summary of the genes located within 1-Mb window are in supporting information Table S4.

trait definitions, either for DD or BIN. Estimated heritabilities were higher for resistance as a binary trait when compared to DD.

Previously, Yáñez *et al.* (2016a), estimated a heritability of 0.16 for resistance against *P. salmonis*, defined as day of death (from the same coho salmon population) through a bivariate linear model. Our study estimated a similar heritability value (0.14) in the same population. Differences in the estimations are likely due to the univariate model we used instead of a bivariate model. We also estimated heritability for *P. salmonis* resistance using a threshold model for the binary trait, which was higher (0.27) than the value reported for DD. These results are consistent with previous findings for resistance against *P. salmonis* in Atlantic salmon using pedigree information (Yáñez *et al.* 2013).

When resistance, defined as day of death, was analyzed using a linear model, heritability was estimated as 0.18 (0.03). When using a threshold model to analyze resistance as a binary trait, a heritability of 0.24 (0.04) was calculated (Yáñez *et al.* 2013, 2014b). The genetic variation and heritability values for *P. salmonis* resistance are in accord with different studies that have also found significant genetic resistance to other bacterial diseases in salmonid species (Gjøen *et al.* 1997; Ødegård *et al.* 2006; Vallejo *et al.* 2016).

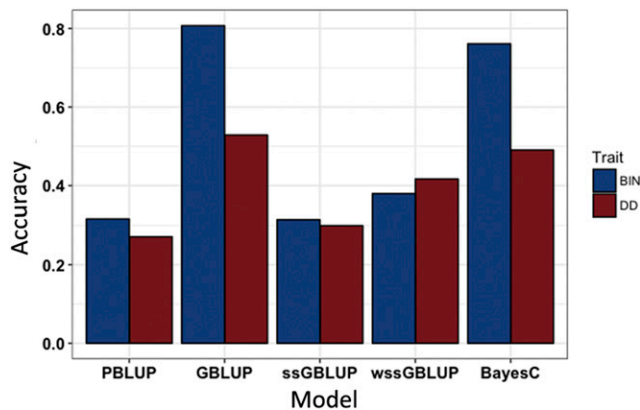
Bacterial disease resistance has been suggested to be a polygenic trait in aquaculture species. For example, Paliokostas *et al.* (2016) suggested that resistance against *Photobacterium damsela subsp.* has a polygenic architecture in Gilthead Sea Bream (*Sparus aurata*). Using a 50K SNP genotyping array, it was possible to elucidate a moderately polygenic architecture of *P. salmonis* resistance in Atlantic salmon (Correa *et al.* 2015b). In the current study, and using 9K SNPs, a similar genetic architecture for resistance against *P. salmonis* in coho salmon population was found. However, it is likely that the moderate number of

individuals could limit the power to detect QTL of larger effect controlling *P. salmonis* resistance in coho salmon.

Among the top ten genetic markers for DD and BIN, one and two markers were identified among all four models, respectively. The availability of an annotated, coho salmon genome made it possible to identify the *phosphoinositide-3-kinase adaptor protein 1* (*pik3ap1*) (also known as the *adaptor protein B-cell PI3K adaptor* (BCAP)), a gene that is related with B-cell development (Herzog *et al.* 2009), proximate to the genetic marker found to be associated with resistance. The role of B-cells, through the humoral response, has been widely investigated and elucidated (reviewed in Janeway *et al.* 2001). B-cells may also have phagocytic activity in both rainbow trout (Li *et al.* 2006) and Atlantic salmon (Øverland *et al.* 2010). These cells are capable of ingesting large particles and bacteria; killing them through phagolysosome fusion (Li *et al.* 2006). Studies *in vitro*, showed that from the total phagocytic leukocytes isolated from Atlantic salmon head kidney, 37% were B-cells. Additionally, 77% were B-cells when leukocytes were isolated from peripheral blood (PB). The phagocytic ability of B-cells was three times higher than those observed in neutrophils in head kidney, while in PB no differences were observed (Øverland *et al.* 2010).

We hypothesize that B-cell development could help in the immune response against *P. salmonis* in coho salmon population through its phagocytic activity. Further studies are needed in order to have a better understanding of the role of these cells in the resistance against *P. salmonis*, and its ability to digest and kill bacteria. For resistance defined as the BIN trait, no genes related with immune response were identified near genetic markers associated with this trait. A possible explanation of this observation, is that the region near the genetic marker acts as a regulatory sequence. Also, it could be due to the relative small sample





**Figure 4** Comparison of predicted accuracies (R) for *Piscirickettsia salmonis* resistance in a coho salmon population comparing between PBLUP and models with genomic data for DD (red bars) and BIN (blue bars).

size for the GWAS, which ideally should be over 1000 animals (limiting the resolution of the GWAS).

The underlying genetic basis for this trait may have an important impact on the accuracy of genomic predictions. Thus, performance comparisons of different algorithms is required for either GS and GWAS when a complex trait is studied for the first time within a population (Vallejo *et al.* 2017b). The current study is the first to evaluate the genetic architecture of resistance against *Piscirickettsia salmonis* in coho salmon through different algorithms (to our knowledge). We used a genomic model that assumes that genetic variances are controlled by an infinite number of markers with minimum effects on the trait (*i.e.*, GBLUP). The GBLUP method calculates genomic relationship (G matrix) using all the genotyped markers for this reason. An extension of this method, which combines genomic (G) and pedigree-based (A) relationship information into the H relationship matrix (Aguilar *et al.* 2010; Legarra *et al.* 2014) was also evaluated. This method, called single-step GBLUP (ssGBLUP), mimics a Bayesian selection model in that it only fits SNPs that explain moderate to large genetic variance of the trait (Wang *et al.* 2012). Results of these different genomic-wide association analyses suggest that resistance against *Piscirickettsia salmonis* has a polygenic architecture, with no major QTL (*i.e.*, explaining  $\geq 10\%$  of the genetic variation) segregating in the current population (Figure 2 and Figure 3).

When resistance was defined as DD, the accuracy for the PBLUP model was slightly lower than resistance measured as BIN (0.271 and 0.316, respectively). These values are consistent with the results obtained for resistance against the bacterial disease pasteurellosis in *S. aurata* defined as day of death, authors reached an accuracy up to 0.30 through a pedigree-based model (Palaïokostas *et al.* 2016). However, both accuracy values are slightly lower when compared with values obtained for resistance against sea lice in Atlantic salmon. In this regard, Correa *et al.* (2017) reached an accuracy of 0.41 for resistance against *Caligus rogercresseyi*, while Tsai *et al.* (2016) reached a prediction accuracy  $\sim 0.5$  for resistance against *Lepeophtheirus salmonis* using PBLUP.

The results from our genomic predictions are in agreement with previous studies that showed higher estimated accuracies using GS than with PBLUP in the same half/full-sib family structure in salmon breeding programs (Nielsen *et al.* 2009; Lillehammer *et al.* 2013; Bangera *et al.* 2017).

The GS model that showed the best performance in terms of accuracy of predictions for DD was GBLUP. This was followed by Bayes C,

wssGBLUP and ssGBLUP. The addition of genomic information allowed an improvement up to 95% in accuracy for this trait. The current study showed a relative accuracy prediction improvement of 54% when comparing PBLUP to wssGBLUP for this trait. This improvement is different to that obtained by Vallejo *et al.* (2016) for resistance against Bacterial Cold Water Disease (BCWD), defined as day of death, in rainbow trout (*Oncorhynchus mykiss*). The authors reported a reduction in the predictive ability (PA) of 20 and 26% using wssGBLUP, either with a chip array (40K) or through RAD sequencing (10K) respectively. The authors attributed this pronounced reduction to stochastic fluctuations due to a small training group of their study ( $n = 583$ ). However, when the number of individuals was increased, this model outperformed the pedigree-based model, reaching a relative increase in accuracy up to 108%, for the same trait definition (Vallejo *et al.* 2017a).

In the case of BIN, the use of the ssGBLUP model, did not show an improvement in the accuracy prediction (with a reduction of 1% in comparison to the accuracy obtained through the pedigree-based model). However, there was an estimated relative increase in accuracy ranging from 20 to 155% when comparing PBLUP to the others genomic models. In this regard, Bayes C and GBLUP showed an increase in accuracy of 140% and 155%, respectively. These high values are similar to the improvements seen with resistance against BCWD defined as a binary trait, which reached a relative increase up to 97% (Vallejo *et al.* 2017a).

The relative improvements ranged in accuracy from -1 to 155% for BIN and from 10 to 95% for DD are greater to those obtained for other diseases resistance studies in Atlantic salmon; even with lower marker numbers. Using an identical random selection design as in the current study, sea lice resistance showed a relative improvement in accuracy of 22% relative to PBLUP when using 37K SNPs (Correa *et al.* 2017). For reliability, an improvement up to 52% with 220K SNPs was reached (Ødegård *et al.* 2014). Tsai *et al.* (2016) reported that when a non-full siblings design was used, an improvement in accuracy of 250% and 500% was reached in two different populations compared to PBLUP. However, when the methodology was changed to a random selection scheme this improvement only reached up to 27% using 33K SNPs.

In case of *Piscirickettsia salmonis* resistance, and using the same cross-validation scheme in our study, the relative reliability was increased by 25% and 30% for resistance, defined as day of death or as a binary trait, respectively with 50K SNPs (Bangera *et al.* 2017).

In the current study we evaluated a wide range of different models of GS for their potential implementation in aquaculture. The performance of each implemented model varied according to the underlying genetic architecture of the trait (Meuwissen *et al.* 2001; Daetwyler *et al.* 2010). Thus, it is valuable to perform these comparisons to identify the best performing method using real data. Similar accuracies among PBLUP and ssGBLUP models could be due to the predicted GEBV as both models are overrepresented by polygenic EBV (Bangera *et al.* 2017). GBLUP estimated genetic relationships using genotype and pedigree data rather than just average relationship as PBLUP (Habier *et al.* 2007). This allows a more accurate genetic relationship matrix and provides an increase in performance, as seen in the current data, due to the close family relationship. Moreover, GBLUP had significantly better performance when resistance was defined as a linear or binary trait compared to the other evaluated genomic models. We suggest that this could be an effect of only genotyping families from the opposite sides of the mortality distribution (*i.e.*, most resistant and most susceptible), and not all the challenged individuals. Bangera *et al.* (2017), reported similar accuracies among GBLUP and Bayesian methods between GBLUP and Bayes C using 10K SNPs in Atlantic salmon. Similar results that have been seen in dairy cattle for most traits (de Roos *et al.* 2009; Goddard 2009)

The greater improvement in accuracy for the binary trait BIN, compared with the linear trait DD, could be due a better fit of the threshold model for BIN than the fit of the linear model for DD. It could also be due to the higher estimated heritability.

We hypothesize that the large improvement values seen in the current study are likely due to an increased level of linkage disequilibrium (LD) found within this farmed coho salmon population. Additional studies are needed to elucidate the minimum number of markers necessary for GS.

Our results, using ddRAD sequencing, are in agreement with other genomic studies, which utilize GBS techniques with aquaculture species. Using RAD sequencing, some authors have previously performed genomic-wide association studies in rainbow trout looking for associations with disease resistance. From a total of 4K identified SNPs, 31 markers were significantly associated with either BCWD or Infectious Hematopoietic Necrosis Virus (IHNV) resistance as a binary trait (Campbell *et al.* 2014). These authors also showed the potential of these RAD markers to predict an animal's phenotype. In the case of BCWD resistance, defined as a linear and binary trait, Palti *et al.* (2015b) identified suggestive and significant SNPs in two different families and candidate genes associated with this trait using ~5K markers per family. Similar numbers of SNPs were used by Liu *et al.* (2015) to significantly associate 18 SNPs.

Genomic selection predictions are in accordance with studies aimed to evaluate GS using others GBS techniques, in both, relative increase in accuracy and number of discovered SNPs. Dou *et al.* (2016) predicted higher accuracies for shell height and shell width using GBLUP and Bayes methods in Yesso scallop (*Patinopecten yessoensis*) using 2K SNPs identified by 2b-RAD. Identical methodology allowed Palaiokostas *et al.* (2016) to reach a relative increase in accuracy up to 53% with 12K SNPs using Bayesian methods compared to PBLUP. A study in rainbow trout using RAD sequencing identified 10K SNPs. Even then, the accuracies were similar with GS models compared to PBLUP, and the authors predicted that increasing the number of individuals could lead to a relative increase in accuracy up to 69% (Vallejo *et al.* 2016).

The genotyping strategy was aimed at the i) evaluation of genomic selection methods; and ii) allowing the identification of molecular markers associated with the trait by means of GWAS. The aim was maximizing the phenotypic variance within the sample while keeping a balanced representation of fish per family. Thus, genotyping strategy was not totally random, but specifically focused on the most extreme families; 17 resistant and 16 susceptible families. We aimed at genotyping all the fish belonging to each selected family. Thus, each family was represented within the sample with an average of 23 (ranging from 11 to 43) fish/family.

The availability of dense SNP arrays for coho salmon, as it is already the case for Atlantic salmon (Houston *et al.* 2014; Yáñez *et al.* 2016b) and rainbow trout (Palti *et al.* 2015a), may increase the accuracy for predicting genomic breeding values and the power for the determination of the genetic factors involved in economically-important traits. It is also expected that in the near future, further functional studies for a better understanding of *P. salmonis* resistance and other complex traits in salmonids will be facilitated by the international initiative on the Functional Annotation of All Salmonid Genomes, FAASG (Macqueen *et al.* 2017).

We have evaluated different GS models, and demonstrated that the use of genomic prediction is a feasible strategy for the improvement of breeding value prediction. This information could be used for the implementation of genomic information in genetic programs for *Piscirickettsia salmonis* resistance in farmed coho salmon populations.

## Conclusions

Moderate significant genetic variation was estimated for resistance against *Piscirickettsia salmonis* in coho salmon, using either pedigree or genomic information. These results highlight the feasibility of including this trait into genetic improvement programs. Our study shows that genomic prediction methods, using ddRAD genotypes (including 9K SNPs), has a substantial advantage in terms of accuracy when compared to pedigree-based model for either DD or BIN. The improvement was up to 95 and 155% respectively in the current population. The association analyses were used to identify a gene related with B-cell development, which could also be involved in resistance against *P. salmonis*. To our knowledge, this is the first study aimed at dissecting the genetic architecture of *P. salmonis* resistance in a coho salmon population.

## ACKNOWLEDGMENTS

AB and KC acknowledge the National Commission of Scientific and Technologic Research (CONICYT) for the funding through the National PhD funding program. AB acknowledges the Government of Canada for the funding through the Canada-Chile Leadership Exchange Scholarship. This project was funded by the U-Inicia grant, from the Vicerrectoria de Investigación y Desarrollo, Universidad de Chile. This work has been conceived on the frame of the grant FONDEF NEWTON-PICARTE (IT14I10100), funded by CONICYT (Government of Chile) and the Newton Fund - The British Council (Government of United Kingdom). This work has been partially supported by Núcleo Milenio INVASAL from Iniciativa Científica Milenio (Ministerio de Economía, Fomento y Turismo, Gobierno de Chile). This research was carried out in conjunction with EPIC4 (Enhanced Production in Coho: Culture, Community, Catch), a project supported by the government of Canada through Genome Canada, Genome British Columbia, and Genome Quebec.

Authors' contributions: AB performed DNA extraction, library construction, ddRAD analysis, GWAS analysis, and wrote the initial version of the manuscript. KrC performed library construction and contributed to the data analysis. AB, GY and KC performed genomic prediction analysis. AJ performed DNA extraction. JPL contributed with study design. WD contributed with analysis and discussion. JMY conceived and designed the study, supervised work of AB and contributed to the analysis, discussion and writing. All authors have reviewed and approved the manuscript.

Animal ethics approval: Challenge and sampling procedures were approved by the Comité de Bioética Animal from the Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile (Certificate N08-2015).

## LITERATURE CITED

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta *et al.*, 2010 Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93(2): 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Aiba, Y., M. Kameyama, T. Yamazaki, T. F. Tedder, and T. Kurosaki, 2008 Regulation of B-cell development by BCAP and CD19 through their binding to phosphoinositide 3-kinase. *Blood* 111(3): 1497–1503. <https://doi.org/10.1182/blood-2007-08-109769>
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10): e3376. <https://doi.org/10.1371/journal.pone.0003376>

- Bangera, R., K. Correa, J. P. Lhorente, R. Figueroa, and J. M. Yáñez, 2017 Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genomics* 18(1): 121. <https://doi.org/10.1186/s12864-017-3487-y>
- Brieuc, M. S. O., C. D. Waters, and J. E. Seeb, and K. A. Naish, 2014 A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3 Genes Genomes Genet.* 4: 447–460. <https://doi.org/10.1534/g3.113.009316>
- Campbell, N. R., S. E. LaPatra, K. Overturf, R. Towner, and S. R. Narum, 2014 Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. *G3 Genes Genomes Genet.* 4: 2473–81. <https://doi.org/10.1534/g3.114.014621>
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, 2011 *Stacks*: building and genotyping loci *de novo* from short-read sequences. *G3 Genes Genomes Genet.* 1: 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 *Stacks*: An analysis tool set for population genomics. *Mol. Ecol.* 22(11): 3124–3140. <https://doi.org/10.1111/mec.12354>
- Correa, K., R. Bangera, R. Figueroa, J. P. Lhorente, and J. M. Yáñez, 2017 The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genet. Sel. Evol.* 49(1): 15. <https://doi.org/10.1186/s12711-017-0291-8>
- Correa, K., M. Filp, D. Cisterna, M. E. Cabrejos, C. Gallardo-Escárate *et al.*, 2015a Effect of triploidy in the expression of immune-related genes in coho salmon *Oncorhynchus kisutch* (Walbaum) infected with *Piscirickettsia salmonis*. *Aquacult. Res.* 46: 59–63. <https://doi.org/10.1111/are.12584>
- Correa, K., J. Lhorente, M. Lopez, L. Bassini, S. Naswa *et al.*, 2015b Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* 16(1): 854. <https://doi.org/10.1186/s12864-015-2038-7>
- Cvitanich, J., O. Garate, and C. E. Smith, 1991 The isolation of a rickettsia-like organism causing disease and mortality in Chilean salmonids and its confirmation by Koch's postulate. *J. Fish Dis.* 14(2): 121–145. <https://doi.org/10.1111/j.1365-2761.1991.tb00584.x>
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams, 2010 The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185(3): 1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi *et al.*, 2013 Special features of RAD Sequencing data: Implications for genotyping. *Mol. Ecol.* 22(11): 3151–3164. <https://doi.org/10.1111/mec.12084>
- Dou, J., X. Li, Q. Fu, W. Jiao, Y. Li *et al.*, 2016 Evaluation of the 2b-RAD method for genomic selection in scallop breeding. *Sci. Rep.* 6(1): 19244. <https://doi.org/10.1038/srep19244>
- Dufflocq, P., J. P. Lhorente, R. Bangera, R. Neira, S. Newman *et al.*, 2016 Correlated response of flesh color to selection for harvest weight in coho salmon (*Oncorhynchus kisutch*). *Aquaculture* 472: 6–11.
- Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5): e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Falconer, D.S., and Mackay T.F.C., 1996. *Introduction to Quantitative Genetics*. fourth ed. Longman Group Limited, Harlow, Essex, U.K.
- FAO, 2016. Fisheries and Aquaculture Department [Online]. (Rome. Updated 31 January 2016. <http://www.fao.org/fishery/statistics/global-aquaculture-production/query/en>
- Fernando, R.L., and D. Garrick, 2013 Bayesian Methods Applied to GWAS. In: Gondro C., van der Werf J., Hayes B. (eds) *Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology (Methods and Protocols)*, vol 1019. Humana Press, Totowa, NJ.
- Fernando, R. L., D. Habier, C. Stricker, J. C. M. Dekkers, and L. R. Totir, 2007 Genomic selection. *Acta Agric. Scand. Sect. Anim. Sci.* 57: 192–195. <https://doi.org/10.1080/09064700801959395>
- Gjøen, H. M., T. Refstie, O. Ulla, and B. Gjerde, 1997 Genetic correlations between survival of Atlantic salmon in challenge and field tests. *Aquaculture* 158(3-4): 277–288. [https://doi.org/10.1016/S0044-8486\(97\)00203-2](https://doi.org/10.1016/S0044-8486(97)00203-2)
- Goddard, M., 2009 Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136(2): 245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- Gonen, S., N. R. Lowe, T. Cezard, K. Gharbi, S. C. Bishop *et al.*, 2014 Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* 15(1): 166. <https://doi.org/10.1186/1471-2164-15-166>
- Habier, D., R. L. Fernando, and J. C. M. Dekkers, 2007 The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177: 2389–2397. <https://doi.org/10.1534/genetics.107.081190>
- Habier, D., R. L. Fernando, K. Kizilkaya, D. J. Garrick, T. Meuwissen *et al.*, 2011 Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12(1): 186. <https://doi.org/10.1186/1471-2105-12-186>
- Hayes, B., and M. Goddard, 2010 Genome-wide association and genomic selection in animal breeding. *Genome* 53(11): 876–883. <https://doi.org/10.1139/G10-076>
- Herzog, S., M. Reth, and H. Jumaa, 2009 Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. *Nat. Rev. Immunol.* 9(3): 195–205. <https://doi.org/10.1038/nri2491>
- Houston, R. D., J. W. Davey, S. C. Bishop, N. R. Lowe, J. C. Mota-Velasco *et al.*, 2012 Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* 13(1): 244. <https://doi.org/10.1186/1471-2164-13-244>
- Houston, R. D., J. B. Taggart, T. Cézard, M. Bekaert, N. R. Lowe *et al.*, 2014 Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15(1): 90. <https://doi.org/10.1186/1471-2164-15-90>
- Janeway, C., P. Travers, M. Walport, and M. Shlomchik, 2001 B-cell activation by armed helper T cells, pp. 343–360 in *Immunobiology 5, The Immune System in Health and Disease*, Ed. 5. Churchill Livingstone, New York.
- Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal, 2014 Single Step, a general approach for genomic selection. *Livest. Sci.* 166: 54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>
- Legarra, A., C. Robert-Granié, E. Manfredi, and J.-M. Elsen, 2008 Performance of genomic selection in mice. *Genetics* 180: 611–618. <https://doi.org/10.1534/genetics.108.088575>
- Li, J., D. R. Barreda, Y.-A. Zhang, H. Boshra, A. E. Gelman *et al.*, 2006 B lymphocytes from early vertebrates have potent phagocytic and microbicidal abilities. *Nat. Immunol.* 7(10): 1116–1124. <https://doi.org/10.1038/ni1389>
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson, 2013 A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45(1): 39. <https://doi.org/10.1186/1297-9686-45-39>
- Liu, S., R. L. Vallejo, Y. Palti, G. Gao, D. P. Marancik *et al.*, 2015 Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front. Genet.* 6: 1–10. <https://doi.org/10.3389/fgene.2015.00298>
- Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*, Sinauer Associates, Sunderland.
- Macqueen, D. J., C. R. Primmer, R. D. Houston, B. F. Nowak, L. Bernatchez *et al.*, 2017 Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* 18: 484. <https://doi.org/10.1186/s12864-017-3862-8>

- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Misztal, I., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar *et al.*, 2016 Manual for BLUPF90 Family of Programs. University of Georgia, Athens, GA.
- Moen, T., J. Torgersen, N. Santi, W. S. Davidson, M. Baranski *et al.*, 2015 Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* 200: 1313–1326. <https://doi.org/10.1534/genetics.115.175406>
- Nielsen, H. M., A. K. Sonesson, H. Yazdi, and T. H. E. Meuwissen, 2009 Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture* 289(3-4): 259–264. <https://doi.org/10.1016/j.aquaculture.2009.01.027>
- Ødegård, J., M. Baranski, B. Gjerde, and T. Gjedrem, 2011 Methodology for genetic evaluation of disease resistance in aquaculture species: Challenges and future prospects. *Aquacult. Res.* 42: 103–114. <https://doi.org/10.1111/j.1365-2109.2010.02669.x>
- Ødegård, J., T. Moen, N. Santi, S. A. Korsvoll, S. Kjøglum *et al.*, 2014 Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5: 1–8. <https://doi.org/10.3389/fgene.2014.00402>
- Ødegård, J., I. Olesen, B. Gjerde, and G. Klemetsdal, 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(1-4): 116–123. <https://doi.org/10.1016/j.aquaculture.2006.05.034>
- Øverland, H. S., E. F. Pettersen, A. Rønneseth, and H. I. Wergeland, 2010 Phagocytosis by B-cells and neutrophils in Atlantic salmon (*Salmo salar* L.) and Atlantic cod (*Gadus morhua* L.). *Fish Shellfish Immunol.* 28 (1): 193–204. <https://doi.org/10.1016/j.fsi.2009.10.021>
- Palaïokostas, C., S. Ferrarero, R. Franch, R. D. Houston, and L. Bargelloni, 2016 Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3 Genes Genomes Genet.* Available at: <https://doi.org/10.1534/g3.116.035220>
- Palti, Y., G. Gao, S. Liu, M. P. Kent, S. Lien *et al.*, 2015a The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol. Ecol. Resour.* 15(3): 662–672. <https://doi.org/10.1111/1755-0998.12337>
- Palti, Y., R. L. Vallejo, G. Gao, S. Liu, A. G. Hernandez *et al.*, 2015b Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated dna sequencing. *PLoS One* 10(9): e0138435. <https://doi.org/10.1371/journal.pone.0138435>
- Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra, 2012 Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7: e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Pulgar, R., C. Hödar, D. Travisany, A. Zuñiga, C. Domínguez *et al.*, 2015 Transcriptional response of Atlantic salmon families to *Piscirickettsia salmonis* infection highlights the relevance of the iron-deprivation defence system. *BMC Genomics* 16(1): 495. <https://doi.org/10.1186/s12864-015-1716-9>
- Rise, M. L., S. R. M. Jones, G. D. Brown, K. R. von Schalburg, W. S. Davidson *et al.*, 2004 Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiol. Genomics* 20(1): 21–35. <https://doi.org/10.1152/physiolgenomics.00036.2004>
- Robledo, D., C. Palaïokostas, L. Bargelloni, P. Martínez, and R. Houston, 2017 Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* 1–13. <https://doi.org/10.1111/raq.12193>
- de Roos, A. P. W., B. J. Hayes, and M. E. Goddard, 2009 Reliability of genomic predictions across multiple populations. *Genetics* 183(4): 1545–1553. <https://doi.org/10.1534/genetics.109.104935>
- Rozas, M., and R. Enríquez, 2014 *Piscirickettsiosis* and *Piscirickettsia salmonis* in fish: a review. *J. Fish Dis.* 37(3): 163–188. <https://doi.org/10.1111/jfd.12211>
- Sernapesca, 2016 Informe Sanitario de Salmonicultura en Centros Marinos 2016. ([Online]. Valparaíso, Chile. Updated 14 August 2017. [http://www.sernapesca.cl/index.php?option=com\\_remository&Itemid=246&func=fileinfo&id=27560](http://www.sernapesca.cl/index.php?option=com_remository&Itemid=246&func=fileinfo&id=27560)).
- Stear, M. J., S. C. Bishop, B. A. Mallard, and H. Raadsma, 2001 The sustainability, feasibility and desirability of breeding livestock for disease resistance. *Res. Vet. Sci.* 71(1): 1–7. <https://doi.org/10.1053/rvsc.2001.0496>
- Tsai, H.-Y., A. Hamilton, A. E. Tinch, D. R. Guy, J. E. Bron *et al.*, 2016 Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48(1): 47. <https://doi.org/10.1186/s12711-016-0226-9>
- Vallejo, R. L., T. D. Leeds, B. O. Fragoneni, G. Gao, A. G. Hernandez *et al.*, 2016 Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. *Front. Genet.* 7: 1–13. <https://doi.org/10.3389/fgene.2016.00096>
- Vallejo, R. L., T. D. Leeds, G. Gao, J. E. Parsons, K. E. Martin *et al.*, 2017a Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49(1): 17. <https://doi.org/10.1186/s12711-017-0293-6>
- Vallejo, R. L., S. Liu, G. Gao, B. O. Fragoneni, A. G. Hernandez *et al.*, 2017b Similar genetic architecture with shared and unique quantitative trait loci for bacterial cold water disease resistance in two rainbow trout breeding populations. *Front. Genet.* 8: 156. <https://doi.org/10.3389/fgene.2017.00156>
- VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11): 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando *et al.*, 2014 Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5: 1–10. <https://doi.org/10.3389/fgene.2014.00134>
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir, 2012 Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* 94(02): 73–83. <https://doi.org/10.1017/S0016672312000274>
- Yáñez, J. M., R. Bangera, J. P. Lhorente, A. Barria, M. Oyarzun *et al.*, 2016a Negative genetic correlation between resistance against *Piscirickettsia salmonis* and harvest weight in coho salmon (*Oncorhynchus kisutch*). *Aquaculture* 459: 8–13. <https://doi.org/10.1016/j.aquaculture.2016.03.020>
- Yáñez, J. M., S. Naswa, M. E. Lopez, L. Bassini, K. Correa *et al.*, 2016b Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Mol. Ecol. Resour.* 16(4): 1002–1011. <https://doi.org/10.1111/1755-0998.12503>
- Yáñez, J. M., R. Bangera, J. P. Lhorente, M. Oyarzún, and R. Neira, 2013 Quantitative genetic variation of resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *Aquaculture* 414–415: 155–159. <https://doi.org/10.1016/j.aquaculture.2013.08.009>
- Yáñez, J. M., L. N. Bassini, M. Filp, J. P. Lhorente, R. W. Ponzoni *et al.*, 2014a Inbreeding and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding nucleus in Chile. *Aquaculture* 420–421: S15–S19. <https://doi.org/10.1016/j.aquaculture.2013.05.028>
- Yáñez, J. M., R. D. Houston, and S. Newman, 2014b Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5: 1–13. <https://doi.org/10.3389/fgene.2014.00415>
- Yáñez, J. M., J. P. Lhorente, L. N. Bassini, M. Oyarzún, R. Neira *et al.*, 2014c Genetic co-variation between resistance against both *Caligus rogercresseyi* and *Piscirickettsia salmonis*, and body weight in Atlantic salmon (*Salmo salar*). *Aquaculture* 433: 295–298. <https://doi.org/10.1016/j.aquaculture.2014.06.026>
- Yoshida, G.M., R. Bangera, R. Carvalheiro, K. Correa, R. Figueroa, J. P. Lhorente, and J. M. Yáñez, 2017 Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. *G3 Genes Genomes Genet.* 8:719–726. <https://doi.org/10.1534/g3.117.300499>

Communicating editor: R. Houston