Comparative Genomic Analysis of Three Salmonid Species Identifies Functional Candidate Genes Involved in Resistance to the Intracellular Bacterium *Piscirickettsia salmonis*

José M. Yáñez[1,2]*, Grazyella M. Yoshida[1], Ángel Parra[1,3,4,5], Katharina Correa[6], Agustín Barría[1,7], Liane N. Bassini[8], Kris A. Christensen[9], Maria E. López[1,10], Roberto Carvalheiro[11,12], Jean P. Lhorente[6] and Rodrigo Pulgar[3]*

[1] Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile, [2] Núcleo Milenio INVASAL, Concepción, Chile, [3] Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Santiago, Chile, [4] Doctorado en Acuicultura. Programa Cooperativo Universidad de Chile, Universidad Católica del Norte, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, [5] Facultad de Ciencias del Mar, Universidad Católica del Norte, Coquimbo, Chile, [6] Benchmark Genetics Chile, Puerto Montt, Chile, [7] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh Easter Bush, Midlothian, United Kingdom, [8] Escuela de Medicina Veterinaria, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile, [9] Fisheries and Oceans Canada, West Vancouver, BC, Canada, [10] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [11] School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), Jaboticabal, Brazil, [12] National Council for Scientific and Technological Development (CNPq), Brasília, Brazil

Piscirickettsia salmonis is the etiologic agent of salmon rickettsial syndrome (SRS) and is responsible for considerable economic losses in salmon aquaculture. The bacterium affects coho salmon (CS; *Oncorhynchus kisutch*), Atlantic salmon (AS; *Salmo salar*), and rainbow trout (RT; *Oncorhynchus mykiss*) in several countries, including Norway, Canada, Scotland, Ireland, and Chile. We used Bayesian genome-wide association study analyses to investigate the genetic architecture of resistance to *P. salmonis* in farmed populations of these species. Resistance to SRS was defined as the number of days to death and as binary survival (BS). A total of 828 CS, 2130 RT, and 2601 AS individuals were phenotyped and then genotyped using double-digest restriction site-associated DNA sequencing and 57K and 50K Affymetrix® Axiom® single nucleotide polymorphism (SNP) panels, respectively. Both traits of SRS resistance in CS and RT appeared to be under oligogenic control. In AS, there was evidence of polygenic control of SRS resistance. To identify candidate genes associated with resistance, we applied a comparative genomics approach in which we systematically explored the complete set of genes adjacent to SNPs, which explained more than 1% of the genetic variance of resistance in each salmonid species (533 genes in total). Thus, genes were classified based on the following criteria: i) shared function of their protein domains among species, ii) shared orthology among species, iii) proximity to the SNP explaining the highest proportion of the genetic variance, and iv) presence in more than one genomic region explaining more than 1% of the genetic variance within species. Our results allowed us to identify 120 candidate genes belonging to at least one of the

four criteria described above. Of these, 21 of them were part of at least two of the criteria defined above and are suggested to be strong functional candidates influencing *P. salmonis* resistance. These genes are related to diverse biological processes, such as kinase activity, GTP hydrolysis, helicase activity, lipid metabolism, cytoskeletal dynamics, inflammation, and innate immune response, which seem essential in the host response against *P. salmonis* infection. These results provide fundamental knowledge on the potential functional genes underpinning resistance against *P. salmonis* in three salmonid species.

INTRODUCTION

Infectious diseases are responsible for large economic losses in salmon farming. *Piscirickettsia salmonis*, the causal agent of salmon rickettsial syndrome (SRS), affects several salmon species and is considered one of the major pathogens affecting the salmon farming industry (Rozas and Enríquez, 2014). *P. salmonis* was identified in 1989 from farmed coho salmon (CS; *Oncorhynchus kisutch*) sampled in Chile (Cvitanich et al., 1991). Since then, *P. salmonis* has been confirmed as the causative agent for clinical and chronic SRS in CS, Atlantic salmon (AS; *Salmo salar*), and rainbow trout (RT; *Oncorhynchus mykiss*) in several countries, including Norway, Canada, Scotland, Ireland, and Chile (Fryer and Hedrick, 2003). Current control protocols and treatments are based on antibiotics and vaccines. The effectiveness of both strategies in field conditions is not optimal (Rozas and Enríquez, 2014). From the total mortalities ascribed to infectious diseases in Chile, SRS is responsible for 18.3%, 92.6%, and 67.9% in CS, RT, and AS, respectively (Sernapesca, 2018). These mortality rates, together with other factors such as antibiotic treatments and vaccinations, have generated economic losses up to USD $450 million per year (Camusetti et al., 2015).

A feasible and sustainable alternative to prevent disease outbreaks is genetic selection for disease resistance (Bishop and Woolliams, 2014). The estimated levels of heritability for resistance to *P. salmonis* in CS, AS, and RT range from 0.11 to 0.41 (Correa et al., 2015; Yáñez et al., 2016; Bangera et al., 2017; Barría et al., 2018a; Yoshida et al., 2018a; Bassini et al., 2019), demonstrating the feasibility of improving *P. salmonis* resistance through artificial selection in farmed salmon species.

Currently, the advancement of molecular technologies has allowed the generation of dense marker panels for salmonid species (Houston et al., 2014; Palti et al., 2015; Yáñez et al., 2016; Macqueen et al., 2017). The use of genotypes from dense panels of single nucleotide polymorphism (SNP) markers, together with phenotypes for the traits of interest, assessed in a large number of individuals could provide opportunities to discover the genetic architecture of complex traits. When genetic markers are linked to a major effect of quantitative trait loci (QTL), marker-assisted selection (MAS) could then be implemented into breeding programs. For instance, a QTL explaining ~80% of the genetic variance for resistance to infectious pancreatic necrosis virus (IPNV) has been identified in Scottish and Norwegian AS farmed populations (Houston et al., 2008; Moen et al., 2009). To date,

the number of IPNV outbreaks has been significantly reduced in Norwegian AS populations because of MAS for IPNV resistance (Hjeltnes et al., 2018). Interestingly, Moen et al. (2015) mapped the QTL to a region containing an epithelial cadherin (*cdh1*) gene encoding a protein that binds to IPNV, indicating that the protein is part of the machinery used by the virus for host internalization.

P. salmonis resistance has been suggested to be polygenic, with many loci explaining a small amount of the total genetic variance (Correa et al., 2015; Barría et al., 2018a), suggesting that the implementation of genomic selection (GS) is the most appropriate strategy to accelerate the genetic progress for this trait. Methods that can model all available SNPs simultaneously, including Bayesian regression methods (Fernando and Garrick, 2013), appear to be better for estimating marker effects than conventional methods of modeling each SNP individually and therefore are becoming increasingly more popular for genome-wide association study (GWAS; Goddard et al., 2009).

Due to the fact that *P. salmonis* affects farmed populations of three phylogenetically related salmonid species, including CS, AS, and RT, generating mortalities in a similar manner and that genetic variation for *P. salmonis* resistance has been already reported, we believe that exploring the genetic architecture of this trait simultaneously in the three species can provide further insights into the biology of the differential response against this intracellular bacteria among individuals. Thus, a comparative genomics approach aiming at evaluating and comparing genomic regions involved in *P. salmonis* resistance in CS, AS, and RT would help in narrowing down the list of potential candidate genes associated with the trait for further functional validation in salmonid species.

The aims of this study were i) to dissect the genetic architecture of resistance to *P. salmonis* in CS, AS, and RT using SNP and phenotype data modeled together using Bayesian GWAS approach, ii) to identify genomic regions involved in *P. salmonis* resistance among the three salmonid species, and iii) to identify candidate genes associated with *P. salmonis* resistance through a comparative genomics analysis.

MATERIALS AND METHODS

Challenge Tests

A total of 2,606, 2,601, and 2,416 fish belonging to 107, 118, and 105 full-sib families from CS, AS, and RT, respectively, were independently challenged with an isolate of *P. salmonis* (strain

LF-89; Mandakovic et al., 2016) as described in Barría et al. (2018a), Bassini et al. (2019), and Yáñez et al. (2013), Yáñez et al. (2014), Yáñez et al. (2016). Before the beginning of each experimental challenge, quantitative polymerase chain reaction (qPCR) was performed in a sub-sample of each population to confirm the absence of *Flavobacterium* spp., infectious salmon anemia virus, and IPNV. Subsequently, fish were intraperitoneally (IP) injected with 0.2 ml of an LD_{50} inoculum of *P. salmonis*. Although an IP challenge is not a natural form of infection, it is an effective method for presenting a naïve animal with a known and controlled amount of bacteria, making sure that the bacterial load and the time of infection are the same in every fish (Pulgar et al., 2015). After IP injection, infected fish were equally distributed by family into three different test tanks. Each challenge was maintained until mortalities returned to baseline levels. At the end of the challenges, all surviving fish were anesthetized and euthanized. A sample of caudal fin was taken from each survivor and dead fish from each of the experimental challenges for DNA extraction. Body weight was measured at the beginning of the challenge and at the time of death for each individual. The presence of *P. salmonis* was confirmed in a random sample of dead fish through qPCR and necropsy. Each experimental challenge was performed at Aquainnovo's Research Station, Xth Region, Chile.

Genotyping

A total of 828 CS, 2130 RT, and 2601 AS were genotyped using double-digest restriction site-associated DNA (ddRAD) and 57K and 50K Affymetrix® Axiom® SNP panels, respectively. Total DNA was extracted using commercial kits following the manufacturer's protocols. For CS, we used the Wizard SV Genomic DNA purification System (Promega), whereas DNeasy Blood & Tissue (Qiagen) was used for RT and AS.

For CS, 10 ddRAD libraries were prepared following the protocol proposed by Peterson et al. (2012) and sequenced on an Illumina Hiseq2500 (150 bp single-end). Raw sequences were analyzed using STACKS version 1.41 (Catchen et al., 2011; Catchen et al., 2013). rad-tags that passed the *process_radtags* quality control (QC) were aligned to the CS reference genome (GCF_002021735.1). Loci were built with *pstacks* setting a minimum depth coverage of three. After catalog construction, rad-tags were matched using *sstacks* followed by *populations* using default parameters. QC included the removal of SNPs below the following thresholds: Hardy-Weinberg equilibrium (HWE) $P < 1 \times 10^{-6}$, minor allele frequency (MAF) < 0.05, and genotyping call rate < 0.80. Individuals with a call rate below 0.70 were removed from the subsequent analysis. For a detailed protocol of library construction and SNP identification, see Barría et al. (2018a).

RT individuals were genotyped using the commercial 57K Affymetrix® Axiom® SNP array developed by the National Center of Cool and Cold Water Aquaculture at the U.S. Department of Agriculture (Palti et al., 2015). SNPs were filtered with the following QC parameters: HWE $P < 1 \times 10^{-6}$, MAF < 0.05, and SNP call rate < 0.95. Individuals with call rates lower than 0.95 were also removed.

The 50K Affymetrix® Axiom® SNP array used to genotype AS was developed by Universidad de Chile and Aquainnovo (Correa et al., 2015; Yáñez et al., 2016). These markers were selected from a 200K array, as described in detail by Correa et al. (2015). Genotypes were subjected to QC using the following criteria: HWE $P < 1 \times 10^{-6}$, MAF < 0.05, SNP, and samples were discarded when the genotype rate was < 0.95.

GWAS analysis

Resistance to SRS was defined as both the number of days to death (DD) after experimental challenge and the binary survival (BS; 0 for surviving individuals at the end of the experimental challenge and 1 for deceased fish). GWAS analyses were performed using the Bayes C method that assumes distributed mixture distribution for marker effects. All model parameters are defined in the following equation:

$$y = Xb + Zu + \sum_{i=1}^{n} g_i a_i \delta_i + e \qquad (EQ1)$$

where y is the vector of phenotypic records (DD or BS); X and Z are the incidence matrix of fixed effects and polygenic effect, respectively; b is the vector of fixed effects (tank and body weight); u is the 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; δ_i is an indicator variable (0, 1) sampled from a binomial distribution with parameters determined such that π value of 0.99; and e is a vector of residual effects.

The prior assumption is that SNP effects have independent and identical mixture distributions, where each SNP has a point mass at zero (with probability π) and a univariate Gaussian distribution (with probability $1 - \pi$) with a mean equal to zero and variance equal to σ_a^2 having in turn a scaled inverse $\mathcal{X}^2$ prior, with $\nu_a = 4$ and $\nu_e = 10$ degrees of freedom and scale parameter, respectively (Fernando and Garrick, 2013). These hyperparameter values were chosen based on previous studies (Peters et al., 2012; Santana et al., 2016; Wolc et al., 2016; Yoshida et al., 2017; Yoshida et al., 2018a).

The analyses were performed using the GS3 software (Legarra et al., 2013). A total of 200,000 iterations in Gibbs sampling were used, with a burn-in period of 20,000 cycles, and the results were saved every 50 cycles. Convergence was assessed by visual inspection of trace plots of the posterior density of genetic and residual variances.

The proportion of the genetic variance explained (GEV) by each significant SNP was calculated as

$$Vg_i = \left(\frac{2p_i q_i a_i^2}{\sigma_u^2} \right) \qquad (EQ2)$$

where p_i and q_i are the allele frequencies for the *i*th SNP, a_i is the estimated additive effect of the *i*th SNP on the phenotype, and σ_u^2 is the estimate of the polygenic variance (Lee et al., 2013).

The association between the SNPs and the phenotypes was assessed using the proportion of the GEV by each marker. To be inclusive regarding the genomic regions to be compared across the three species, we selected each of the regions explaining at least 1% of the genetic variance for the trait in each species.

The heritability values were calculated as

$$h^2 = \frac{V_A'}{V_A' + \sigma_e^2} \tag{EQ3}$$

where V_A' is the total additive genetic variance estimated as the sum of additive marker $\left(2\sigma_a^2 \pi \sum p_i q_i\right)$ and the polygenic pedigree based $\left(\sigma_g^2\right)$ additive genetic variance.

Comparative Genomic Analysis

Initially, sequence homologies between chromosomes containing regions with SNPs explaining more than 1% of the genetic variance were compared. Synteny among these chromosomes was identified using Symap (Soderlund et al., 2011). The relationship between the chromosomes from CS, RT, and AS and the association between SNPs and resistance to *P. salmonis* (Manhattan plot) was plotted using Circos (Krzywinski et al., 2009).

To identify caandidate genes asstociated with *P. salmonis* resistance, we used a comparative genomic analysis among CS, RT, and AS. For this, we mapped the location of each SNP that explained 1% or more of the genetic variance for the trait on the reference genome (NCBI_RefSeq) of each species: CS (GCF_002021735.1), RT (GCF_002163495.1; Pearse et al., 2018), and AS (GCF_000233375.1; Lien et al., 2016). Subsequently, we retrieved the sequences of all the genes (and their protein products) adjacent to each SNP within a window of 1 Mb (500 kb downstream and 500 kb upstream to
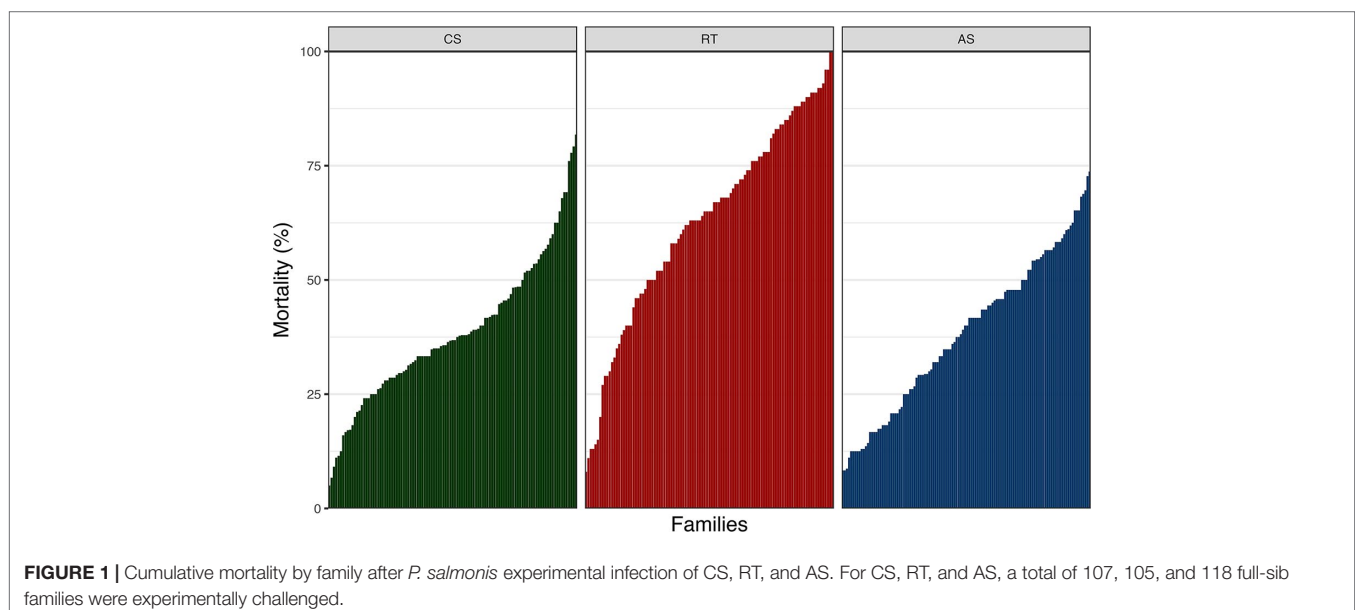
the associated SNP). We then used this information to apply the following criteria to classify and prioritize functional candidate genes by comparing the genomic regions involved in *P. salmonis* defined as DD and BS within and among the three species:

i) The complete set of genes was identified and classified into homologous superfamilies based on InterPro (Mitchell et al., 2019) protein domain signatures using Blast2GO software version 5.2.5 (Götz et al., 2008; referred to as Group A);

ii) Orthologous and paralogous genes among species were identified using the ProteinOrtho tool (Lechner et al., 2011). Multidirectional alignments were performed using the full-length sequences among complete sets of proteins encoded in each of the three species to obtain orthologous groups, with a 35% threshold for identity and similarity (Group B);

iii) The complete set of genes within 1 Mb windows adjacent to SNPs explaining the highest proportion of the genetic variation for each trait (leader SNP) was recovered and classified as high priority genes (Group C); and

iv) The complete set of genes located at the intersection of more than 1 Mb windows within a species was also identified and considered as high priority genes (Group D).

RESULTS

Challenge Test and Genetic Parameters

There was considerable phenotypic variation for *P. salmonis* resistance across fish species (**Figure 1**). The average cumulative mortality for different families ranged from 5% to 81%, 8% to 100%, and 8.3% to 73.7% for CS, RT, and AS, respectively. This

FIGURE 1 | Cumulative mortality by family after *P. salmonis* experimental infection of CS, RT, and AS. For CS, RT, and AS, a total of 107, 105, and 118 full-sib families were experimentally challenged.

TABLE 1 | Estimates of total additive genetic variance $\left(V_a'\right)$, residual variance $\left(\sigma_e^2\right)$, heritability (h²), and standard deviation (SD) for resistance against *P. salmonis* in three salmonid species.

Species	DD				Binary survival			
	V_a'	σ_e^2	h²	SD	V_a'	σ_e^2	h²	SD
CS	28.91	60.70	0.32	0.07	7.53	1.00	0.88	0.03
RT	30.42	32.71	0.48	0.04	1.87	1.00	0.64	0.05
AS	16.52	53.17	0.24	0.04	0.47	1.00	0.32	0.05

result suggests that the phenotypic variation for this trait could be related to the genetic background on each species. Estimated heritabilities for *P. salmonis* resistance were significant for the three species, indicating the feasibility to improve the trait by means of artificial selection (**Table 1**). The genomic heritability values for DD were 0.32 for CS, 0.48 for RT, and 0.24 for AS. When resistance was defined as BS, genomic heritability estimates increased to 0.88, 0.64, and 0.32 for CS, RT, and AS, respectively, representing moderate to high levels of genetic variation for *P. salmonis* resistance.
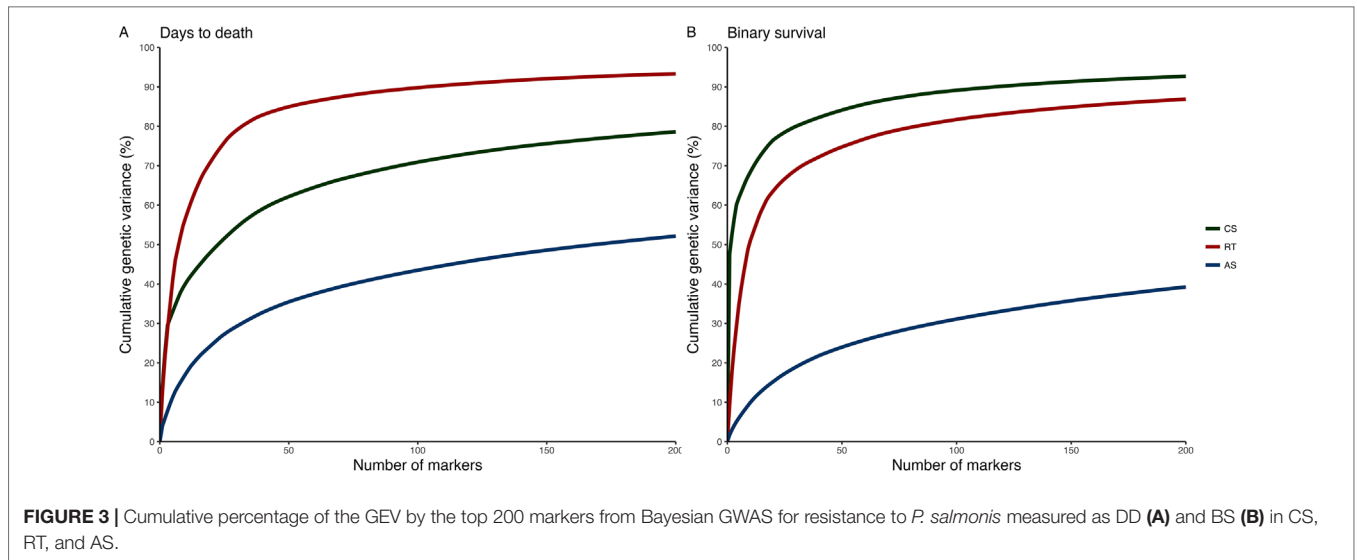
GWAS Analysis

A total of 580 CS (9,389 SNPs), 1,929 RT (24,916 SNPs), and 2,383 AS (42,624 SNPs) were retained after QC. For CS and RT, we found relatively few SNPs explaining a moderate to high percentage of genetic variance for *P. salmonis* resistance. In contrast, for AS, a large number of SNPs with small effect were found and the percentage of GEV by a single marker was not higher than 5% (**Figure 2**; **Supplementary Figure 1**). Although there were multiple shared syntenic regions with associated SNPs (4 for DD and 5 for BS) in two species, there were no shared syntenic regions where all three species had common

associated SNPs (**Figure 2**). **Figure 3** (and **Supplementary Figure 2**) highlights the different genetic architecture for resistance to *P. salmonis* among the three salmonid species studied. For CS, the top 200 SNPs explained about 70% and 90% of genetic variance for DD and BS, respectively, and just a marker located in chromosome 29 represented more than 50% of total genetic variance for BS. For RT, the top 200 SNPs explained 90% and 80% for DD and BS, respectively, whereas, in AS, they explained slightly more than 30% for both traits. These results suggested that CS and RT both appear to have oligogenic control with few markers having large effect loci, whereas the small effect of loci suggested the polygenic nature for resistance to *P. salmonis* in AS.

Comparative Genomic Analysis

We mapped the location of each SNP that explained 1% or more of the genetic variance for both DD and BS to the reference genome of CS, RT, and AS and searched for genes within 1 Mb windows flanking each SNP. This search allowed us to identify 533 unique genes that encoded 957 proteins. The complete list of genes and proteins can be found in **Supplementary Table S1: Sheets 1 to 6**.
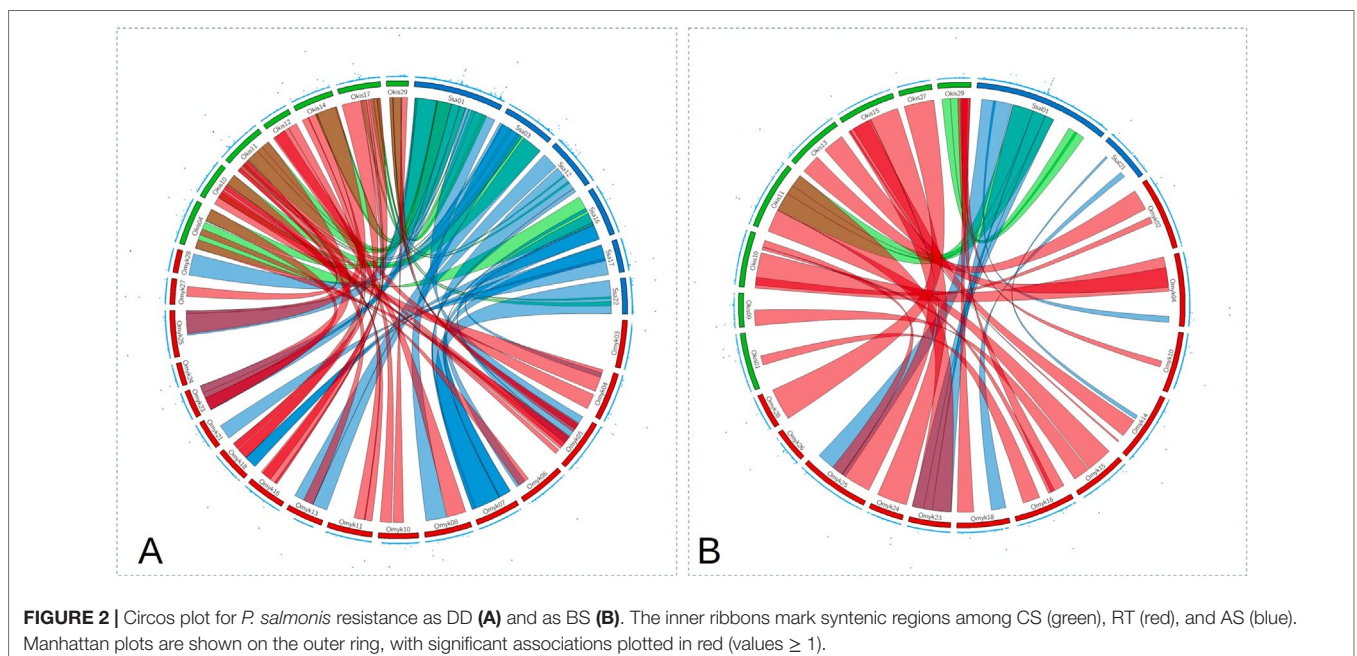
FIGURE 2 | Circos plot for *P. salmonis* resistance as DD **(A)** and as BS **(B)**. The inner ribbons mark syntenic regions among CS (green), RT (red), and AS (blue). Manhattan plots are shown on the outer ring, with significant associations plotted in red (values ≥ 1).

FIGURE 3 | Cumulative percentage of the GEV by the top 200 markers from Bayesian GWAS for resistance to *P. salmonis* measured as DD **(A)** and BS **(B)** in CS, RT, and AS.

To prioritize functional candidate genes, we annotated and classified the complete set of encoded proteins in homologous superfamilies for each trait and species based on InterPro protein domain signatures. We identified 194 and 129 homologous superfamilies for DD and BS, respectively, 103 of which were shared between traits (**Supplementary Table S1: homologous superfamilies**). The homologous superfamilies and the number of proteins present in at least two salmonid species are shown in **Figure 4**. Remarkably, about 30% of the proteins from genes present in regions associated with DD belong to five homologous

superfamilies [*P-loop containing nucleoside triphosphate hydrolase* (also known as P-loop_NTPase), *immunoglobulin-like fold*, *zinc finger C2H2 superfamily*, *zinc finger RING/FYVE/ PHD-type*, *and protein kinase-like domain superfamily*]. A total of 30% of proteins from genes present in regions associated with BS belong to only three homologous superfamilies (*P-loop_NTPase*, *immunoglobulin-like fold*, and *immunoglobulin-like domain superfamily*). Interestingly, the P-loop_NTPase superfamily contained the largest group of proteins for both traits, and at least one representative protein from each salmonid species belonged
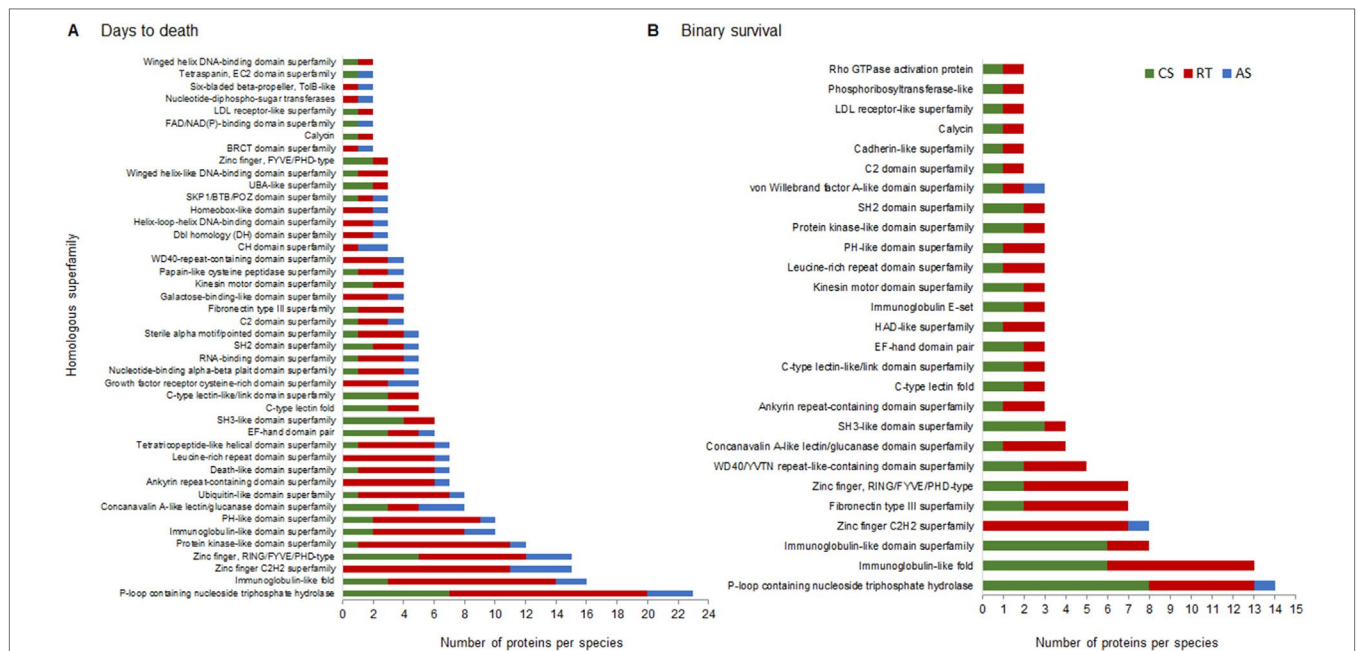
FIGURE 4 | Homologous superfamilies (InterPro) adjacent to the complete set of SNPs that explain more than 1% of the genetic variance of resistance to SRS measured as DD **(A)** and BS **(B)**. Bars represent the abundance of genes in each homologous superfamily present in at least two salmonid species. CS, RT, and AS.

to this superfamily. Thirty-one of the proteins identified in this study are part of this superfamily, including some GTPases, kinesin and myosin proteins, and ATP-dependent RNA helicases [**Supplementary Table S1, sheet: P-loop NTPases (Group_A)**].

To complement these analyses, we looked for orthologous proteins through multi-directional alignments using full-length sequences of the complete set of proteins for each species (Group B). Only five groups of orthologous genes were identified in at least two species, highlighting three non-receptor tyrosine-protein kinases (nr-TPK) with representative genes in the three species for DD and two species for BS. In addition, for DD, two ATP-dependent RNA helicases (DDX) and two Ras-related proteins (RAB) were identified in CS and RT, whereas two FYVE, RhoGEF/PH domain-containing proteins (FGD) were identified in RT and AS. For BS, two fatty acid-binding proteins (L-FABP) and two ankyrin repeat domain-containing proteins were identified in CS and RT [**Supplementary Table S1, sheet: Orthologous genes (Group_B)**]. The proteins nr-TPK, DDX, and L-FABP are also encoded by genes adjacent to SNPs that explained the highest proportion for the genetic variance (leader SNP) for both trait definitions (Group C).

Group C contained other genes (n=42) that encoded proteins such as myosin-IIIb (MYO3B), ATP-dependent RNA helicase (TDRD9), kinesin protein (KIF15), and kinesin protein (KIF2C) that are also included into the P-loop_NTPase superfamily as well as members of the orthologous groups such as FABP. Other genes encoding proteins classically associated with immune response such as tripartite motif-containing protein 35 (TRIM35) and lysozyme C II (LYZ2) are also part of this group. A complete list of these genes and proteins is in **Supplementary Table S1, sheet: Adjacent to leader SNP (Group_C)**.

Group D was composed of genes (n=58) located adjacent to more than one SNP simultaneously (within overlapped windows). Among them, we identified GTPase IMAP family member 4 (GIMAP4), GTPase IMAP family member 8 (GIMAP8), NLR family CARD domain-containing protein 3 (NLRC3), ADP-ribosylation factor protein 5B (ARL5B), voltage-dependent L-type calcium channel subunit beta-2 (CACNB2), and heparan sulfate glucosamine 3-O-sulfotransferase 3A1 (HS3ST3A1), all of which are also P-loop_NTPases. In addition, we identified histidine triad nucleotide-binding protein 1 (HINT1), which is also adjacent to the leader SNP for DD in AS, and other genes associated with immune response such as collectin-12 (COL12), macrophage mannose receptor 1 (MRC1), and tapasin-related protein (TAPBPR). A complete list of these genes and proteins can be found in **Supplementary Table S1, sheet: Genes overlapped windows (Group_D)**. Additionally, the gene that codes for NACHT, LRR, and PYD domains-containing protein 12 (NLRP12) was found in Groups A, C, and D.

We identified several candidate genes associated with *P. salmonis* resistance (n=120), which were present in at least one of the groups described previously. These genes are associated with the following biological processes: dependence on kinase activity, GTP hydrolysis, helicase activity, lipid metabolism, cytoskeletal dynamics, and inflammation. To rank the genes, we scored them based on the counting of each of them across following categories: i) species (CS, RT, and AS), ii) trait definitions (DD and BS), and iii) groups (A–D); thus, the maximum score for one particular gene was equal to 9. The prioritized functional candidate genes based on the score described above are shown in **Table 2** and the complete list of unique candidate genes (n=120) can be found in **Supplementary Table S1. sheet: Candidate genes.**

TABLE 2 | Summary of candidate genes associated with *P. salmonis* resistance for CS, RT, and AS ranked by score, which is simply based on the number of appearance of each gene across the following categories: i) species (CS, RT, and AS), ii) trait definitions (DD and BS), and iii) groups (A–D).

Gene symbol	Protein description	Species	Trait	Group	Score[a]
NRTPK	nr-TPK (cytosolic)	CS, RT, and AS	DD and BS	B–D	8
DDX	ATP-dependent RNA helicase DDX	CS and RT	DD	A–C	6
ARL5B	ADP-ribosylation factor protein 5B	CS	DD and BS	A and D	5
LFABP	Fatty acid-binding protein, liver	CS and RT	BS	B and C	5
GIMAP4	GTPase IMAP family member 4	RT	DD and BS	A and D	5
HS3ST3A1	Heparan sulfate glucosamine 3-O-sulfotransferase 3A1	AS	DD and BS	A and D	5
KIF2C	Kinesin protein KIF2C	RT	DD and BS	A and C	5
MYO3B	Myosin-IIIb	CS	DD and BS	A and C	5
NLRP12	NACHT, LRR, and PYD domains-containing protein 12	AS	DD	A, C, and D	5
RAB	Ras-related protein Rab	CS and RT	DD	B and C	5
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	CS	DD and BS	A and D	5
TDRD9	ATP-dependent RNA helicase TDRD9	CS	DD	A and C	4
FGD	FYVE, RhoGEF, and PH domain-containing protein	RT and AS	DD	B	4
GIMAP8	GTPase IMAP family member 8	RT	DD	A and D	4
HINT1	Histidine triad nucleotide-binding protein 1	AS	DD	C and D	4
KIF15	Kinesin protein KIF15	CS	DD	A and C	4
NLRC3	NACHT, LRR, and CARD domains-containing protein 3	RT	DD	A and D	4
COL12	Collectin-12	CS	BS	D	3
LYZ2	Lysozyme C II	AS	DD	C	3
MRC1	Macrophage mannose receptor 1	CS	BS	D	3
TAPBPR	Tapasin-related protein	RT	DD	D	3

[a] The maximum score possible for one particular gene is equal to 9.

DISCUSSION

The comparative genomic strategy used in this study allowed us to identify groups of homologous superfamilies and orthologous genes common to more than one species of salmonids among genes adjacent to SNPs that explain more than 1% of the genetic variance for *P. salmonis* resistance. To our knowledge, this is the first study that aims at identifying and prioritizing functional candidate genes involved in the differential response against bacterial infection by means of comparing results from GWAS mapping across different phylogenetically related salmonid species.

GENETIC ARCHITECTURE OF RESISTANCE TO *P. SALMONIS*

Heritability estimates are in agreement with previous studies aimed to estimate levels of genetic variation for resistance to bacterial diseases in salmonid species. For instance, Vallejo et al. (2016), Vallejo et al. (2017) presented heritabilities ranging from 0.26 to 0.54 and from 0.31 to 0.48 for resistance to bacterial cold water disease in a farmed RT population. The levels of genetic variation observed in the current study are consistent or somewhat higher than previous estimates of heritabilities for resistance to *P. salmonis* depending on the species and the trait definition. For instance, previous heritability values for *P. salmonis* resistance estimated based on pedigree information reached a maximum of 0.16, 0.44, and 0.41 for CS, RT, and AS, respectively (Yáñez et al., 2013; Yáñez et al., 2014; Yáñez et al., 2016; Bassini et al., 2019). When heritability for *P. salmonis* resistance was estimated based on genomic information, the maximum values reported previously were 0.39 and 0.62 for AS and RT, respectively (Bangera et al., 2017; Yoshida et al., 2018a).

Our results show evidence of alleles of medium to large effect involved in resistance to *P. salmonis* in CS and RT. In contrast, for AS, our results suggest that if alleles of large effect do exist, they are at such low frequency that they individually explain a small proportion of the variance for resistance to *P. salmonis*. The identification of genomic regions harboring associated SNPs was based on GWAS using the Bayes C approach, which is more suitable for oligogenic traits (Habier et al., 2011). In a few cases, the same SNP was significantly associated with both trait definitions (DD and BS). This could be the result of pleiotropy, closely linked genes [local linkage disequilibrium (LD)], or by a strong correlation between both traits. For example, we observed the same SNP associated with DD and BS in CS (58185_41 and 24601_47) and RT (AX-89926208 and AX-89966072) among the top 10 SNPs explaining most of the genetic variance for the trait.

Based on the LD of the AS population (measured as r^2), the number of SNPs used for AS (~43K) should be enough to cover the entire genome (Barría et al., 2018b). There is a lack of studies aimed at evaluating the LD and population structure of the current farmed RT population. Based on results from a different RT farmed population, at least 20K SNPs are necessary to cover the whole genome (Vallejo et al., 2018). If the LD levels of the present RT population are similar to those reported by Vallejo et al. (2018), the 23K SNPs used here will most likely cover the whole genome. However, this is not the case for

CS. Using a high-density SNP array, Rondeau et al. (In preparation) and Barría et al. (2019) suggested that at least 74K SNPs are necessary for whole-genome studies of the current CS population. The small number of SNPs assayed in this study for CS (9389) most likely affected the identification of markers with a moderate to high effect on resistance to *P. salmonis* in this species.

Candidate Proteins Associated With the Resistance to *P. salmonis*

Whereas the complete set of proteins predicted from reference genomes of CS, RT, and AS consisted of 57,592, 58,925, and 97,738, respectively, the proteins neighboring SNPs associated with resistance (range of 1 Mb) represent less than 1% of the different proteomes. The characterization of the complete set of proteins among species established that the most prevalent homologous superfamily was the P-loop_NTPase. However, as this superfamily contains proteins with at least 21 functions (Shalaeva et al., 2018), it is possible that the high frequency of proteins identified from this group was due to the overall high representation in salmonid genomes. For this reason, we retrieved the sequences of 100 randomly selected proteins from the genomes of CS, RT, and AS and classified them into subfamilies (**Supplementary Figure S3**). The results indicate that P-loop_NTPase is not the most prevalent in any of the salmonid species, which suggests that this homologous superfamily is actually enriched in the regions analyzed and is not a consequence of their high representation in CS, RT, and AS genomes.

When traits are polygenic in nature, the identification of genes underlying them is a challenging task and often depends on previous knowledge of the function of genes adjacent to the associated SNPs (Jiang et al., 2014; Bouwman et al., 2018; Robledo et al., 2019). Our strategy was based on identifying orthologous proteins between the salmonid species and families of homologous proteins in the complete set of proteins adjacent to all the SNPs that explained more than 1% of the genetic variance, without searching for a specific function. The identification of genes directly associated with the innate immune response, after applying all the classification criteria, such as LYZ2, MRC1, COL12, and TAPBPR, suggests that our strategy was successful in finding strong functional candidate genes involved in resistance to *P. salmonis*. Interestingly, about 100 genes not classically associated with the immune system were also identified; among which, 17 were part of at least two of the groups described previously and hence are considered strong candidates for being responsible on trait variation (**Table 2**).

Previously, lysozymes have primarily been described as having a bacteriolytic activity against Gram-positive bacteria; however, the expression of LYZ2 has been shown to be induced in a resistant RT line in response to *Flavobacterium psychrophilum* infection (Langevin et al., 2012) and in AS families in response to *P. salmonis* infection (Pulgar et al., 2015), indicating that the transcriptional regulation of this enzyme in salmonids responds to Gram-negative bacterial infection. MRC1 and COL12 are membrane receptors that display several functions associated with innate immunologic defense, particularly in the recognition of carbohydrate structures of pathogens and as phagocytic receptors of bacteria, yeasts, and

other pathogenic microorganisms (Harris et al., 1992; Ma et al., 2015). It has been reported that enhanced infection in human phagocytes with *Francisella tularensis*, a bacterium phylogenetically related to *P. salmonis*, is mediated by MRC1 (Schulert and Allen, 2006), whereas COL12 led to the activation of the alternative pathway of complement *via* association with properdin, a key positive regulator of the pathway by increment of the half-life of the C3 and C5 convertases (Ma et al., 2015). TAPBPR has been described as a second major histocompatibility complex class I-dedicated chaperone essential to providing specificity for T-cell responses against viruses and bacteria (Hermann et al., 2015) and the related protein tapasin has been shown to be induced in monocyte/macrophage in RT by chum salmon reovirus infection (Sever et al., 2014).

Another set of candidate proteins for SRS resistance in the three salmonid species studied are a cluster of cytosolic nr-TPKs. These proteins are a subgroup of the tyrosine kinase family, enzymes that phosphorylate tyrosine residues, and regulate many cellular functions, such as cell growth and survival, apoptosis, cell adhesion, cytoskeleton remodeling, and differentiation (Neet and Hunter, 1996). Although these proteins are not classically related to the response to pathogens, it has been described that the interaction of T- and B-cell antigen receptors with some nr-TPKs is critical to the activation of lymphocytes by an antigen (Sefton and Taddie, 1994). Moreover, some cellular signaling pathways are hijacked by intracellular pathogens, which can subvert protein phosphorylation to control host immune responses and facilitate invasion and dissemination (Haenssler and Isberg, 2011). It has been described that some bacterial effectors are injected into host cells through their secretion systems where they inhibit the Src kinase. In particular, the effector EspJ, an ADP-ribosyltransferase of the bacteria *Escherichia coli* and *Citrobacter rodentium*, regulates multiple host nr-TPKs *in vivo* by ADP-ribosylation, demonstrating that part of its target protein repertoire involves Src kinases such as YES1 and LYN as well as the adapter SYK (Young et al., 2014; Pollard et al., 2018), all of which were identified in this study in CS, RT, and AS. Remarkably, among the candidate genes, we also identified the small ARL5B, suggesting that an adequate regulation of the activity of nr-RTKs by ADP-ribosylation could be critical to combat *P. salmonis* infection.

Other orthologous candidate genes identified in this study encode for proteins RAB1 and RAB18, both members of the GTPase superfamily. GTPases are a large family of hydrolase enzymes that bind and hydrolyze GTP and play an important role in signal transduction, protein translation, control and cellular differentiation, intracellular transport of vesicles, and cytoskeletal reorganization, among other cellular processes (Bourne et al., 1991). Specifically, RAB GTPases constitute a subfamily of small GTPases known as master regulators of intracellular membrane traffic (Stenmark, 2009). As *P. salmonis* drives the formation of host membrane-derived organelles, the development of these *P. salmonis*-containing vacuoles is dependent on the bacterium's ability to usurp the intracellular membrane system of the fish. Furthermore, two orthologous of FGD were identified in RT and AS. These proteins activate CDC42, a GTPase involved in the organization of the actin cytoskeleton and with a role in early contractile events in phagocytes (Ching et al., 2007). As it has been described that the

infective process of *P. salmonis* depends on the exploitation of the actin monomers (Ramírez et al., 2015), the identification in this study of candidate genes that encode for cytoskeletal motor proteins (two kinesins and a myosin) highlights their relevance not only for the reorganization of the cytoskeleton but also for its motility and involvement in the development of the infection (Hoyt et al., 1997). Remarkably, two other candidate proteins associated with SRS resistance are also members of the GTPase superfamily, GIMAP4 and GIMAP8. This is a family of proteins abundantly expressed in lymphocytes and whose function is to contribute in the regulation of apoptosis and the maintenance of T-cell numbers in the organism (Yano et al., 2014).

Another group of orthologous genes code for ATP-dependent RNA helicases DDX24 in CS and DDX47 in RT for DD. The ATP-dependent RNA helicase DDX family, also known as DEAD-box helicases, is required for different cellular processes such as transcription, pre-mRNA processing, ribosome biogenesis, nuclear mRNA export, translation initiation, RNA turnover, and organelle function. The protein structure is very similar to viral RNA helicases and to DNA helicases, which suggests that the fundamental activities of these enzymes are similar (Rocak and Linder, 2004). Viruses also use RNA helicases at various stages of their life cycle. Many viruses carry their own helicases to assist with the synthesis of their genome, but others synthesize their genome within the cell nucleus, which tends to exploit cellular helicases and thus do not encode their own. We also identified the ATP-dependent RNA helicase TDRD9, which has not been directly implicated in infection but was differentially expressed in channel catfish in response to *Aeromonas hydrophila* infection (Li et al., 2013). Mechanistic studies of RNA helicases will allow the determination of the precise role of these helicases in the host-pathogen interaction.

The last group of orthologous genes identified code for two L-FABPs in CS and RT for BS. L-FABPs are abundant in hepatocytes and are known to be associated with lipid metabolism. In addition, these proteins are up-regulated in several types of cancer, but their role in infection remains unclear (Ku et al., 2016). Nevertheless, it has been recently reported that serum and urine L-FABP may be a new diagnostic marker for liver damage in patients with both acute and chronic hepatitis C infection (Cakir et al., 2017). Interestingly, in AS challenged with *P. salmonis*, L-FABP was up-regulated in resistant families and simultaneously down-regulated in susceptible families (Pulgar et al., 2015), suggesting a transcriptional regulation in response to *P. salmonis* infection and a putative expression marker of resistance to SRS.

Genes coding NLRP12, CACNB2, HS3ST3A1, and HINT1 were also selected as candidate genes for SRS resistance. NLRP12 and NLRC3 are two cytosolic proteins that share two functional domains (NACHT and LRR). NLRP12 was one of the best ranked genes, adjacent to the leader SNP and adjacent to more than one SNP simultaneously for DD in AS. This protein functions as an attenuating factor of inflammation in monocytes by negative regulation of the nuclear factor-κB (NF-κB) activation (Fata et al., 2013). In murine macrophages, a significant expression increase has been shown in cells infected with the intracellular parasite *Leishmania major* compared to non-infected macrophages (Fata et al., 2013). NLRC3 is also a negative regulator of the innate immune response mediated by the inhibition of Toll-like

receptor-dependent activation of the transcription factor NF-κB (Schneider et al., 2012). The presence of these genes suggests that the control of the inflammatory reaction in response to *P. salmonis* infection could be essential to combat SRS.

To the best of our knowledge, this is the first time that functional candidate genes underpinning resistance to *P. salmonis* are proposed based on a comparative genomics approach comparing GWAS results for the same trait in different fish genus/species. We hypothesize that variations in the sequences of these genes could play important roles in the host response to *P. salmonis* infection, which could be tested through new genetic approaches such as gene editing using CRISPR-Cas9 and used through GS or more traditional selection practices. All this information together can be used to generate better control and treatment measures for one of the most important bacterial diseases affecting salmon aquaculture.

CONCLUSIONS

Although *P. salmonis* resistance has previously been described as a polygenic trait, our comparative genomics approach based on GWAS results for the same trait in different salmonid species allowed us to identify about 100 candidate genes that may explain resistance to *P. salmonis*. Of these, 21 are suggested to be strong functional candidates influencing the trait. These genes are associated with multiple biological processes, including dependence on kinase activity, GTP hydrolysis, helicase activity, lipid metabolism, cytoskeletal dynamics, inflammation, and the innate immune response. We hypothesize that variations in the sequences of these genes could play an important role in the expression and/or activity of their encoded proteins and consequently in the resistance to *P. salmonis*. This information could be used to generate better control and treatment measures, based on selective breeding or new drug development, for one of the most important bacterial diseases affecting salmon aquaculture.

DATA AVAILABILITY

Genotype and phenotype data generated for this study are available as supplementary material. CS, RT, and AS data are found on Supplementary files 2 to 4, respectively.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All experimental challenges 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).

AUTHOR CONTRIBUTIONS

JY conceived of and designed the study and drafted the manuscript. GY assessed the GWAS analyses. AP and RP designed and assessed the comparative genomic analyses and contributed in the first draft of the manuscript and discussion. LB, ML, and KCo contributed in the RT and AS sampling, genotyping, and QC. AB performed DNA extraction from CS samples, contributed in the initial draft of the manuscript, and performed library construction. KCh performed ddRAD library construction and assessed the comparative sequences analyses between species. RC and JL contributed in the study design, analyses, and discussion. All authors have reviewed and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00665/full#supplementary-material

FIGURE S1 | Manhattan plots for resistance to *P. salmonis* measured as DD in CS, RT, and AS. Y-axis represents the percentage of the GEV by each marker.

FIGURE S2 | Manhattan plots for resistance to *P. salmonis* measured as BS in CS, RT, and AS. Y-axis represents the percentage of the GEV by each marker.

FIGURE S3 | Homologous superfamilies (InterPro) associated with 100 random selected proteins from CS, RT, and AS genomes.

REFERENCES

Antoine, T. E., Yakoub, A., Maus, E., Shukla, D., and Tiwari, V. (2014). Zebrafish 3-O-sulfotransferase-4 generated heparan sulfate mediates HSV-1 entry and spread. *PLoS One* 9 (2), e87302. doi: 10.1371/journal.pone.0087302

Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R., and Yáñez, J. M. (2017). Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genom.* 18, 121. doi: 10.1186/s12864-017-3487-y

Bassini, L. N., Lhorente, J. P., Oyarzún, M., Bangera, R., Yáñez, J. M., and Neira, R. (2019). Genetic parameters for *Piscirickettsia salmonis* resistance, sea lice (*Caligus rogercresseyi*) susceptibility and harvest weight in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture.* 510, 276–282. doi: 10.1016/j.aquaculture.2019.05.008

Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2018a). Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3-Genes Genomes Genet.* 8 (4), 1183–1194. doi: 10.1101/124099

Barría, A., Lopez, M. E., Yoshida, G., Carvalheiro, R., and Yáñez, J. M. (2018b). Population genomic structure and genome-wide linkage disequilibrium in farmed Atlantic salmon (*Salmo salar L.*) using dense SNP genotypes. *Front. Genet.* 9, 649. doi: 10.3389/fgene.2018.00649

Barría, A., Christensen, K. A., Yoshida, G., Jedlicki, A., Lhorente, J. P., Davidson, W. S., et al. (2019). Whole genome linkage disequilibrium and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding population using high density SNP array. *Front. Genet.* 10, 498. doi: 10.3389/fgene.2019.00498

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

Bourne, H., Sanders, D. A., and Frank McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* 349, 117–127. doi: 10.1038/349117a0

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., et al. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* 50, 362–367. doi: 10.1038/s41588-018-0056-5

Cakir, O. O., Toker, A., Ataseven, H., Demir, A., and Polat, H. (2017). The importance of liver-fatty acid binding protein in diagnosis of liver damage in patients with acute hepatitis. *J. Clin. Diagn. Res.* 11 (4), OC17–OC21. doi: 10.7860/JCDR/2017/24958.9621

Camusetti, M. A., Gallardo, A., Aguilar, D., and Larenas, J. (2015). Análisis de los costos por la utilización de quimioterápicos y vacunas en la salmonicultura. *Salmonexpert.*

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., and De Koning, D.-J. (2011). Stacks: building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genet.* 1, 171–182. doi: 10.1534/g3.111.000240

Ching, K. H., Kisailus, A. E., and Burbelo, P. D. (2007). Biochemical characterization of distinct regions of SPEC molecules and their role in phagocytosis. *Exp. Cell Res.* 313, 10–21. doi: 10.1016/j.yexcr.2006.09.011

Correa, K., Lhorente, J., Lopez, M., Bassini, L., Naswa, S., Deeb, N., et al. (2015). Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar L.*) chromosomes. *BMC Genom.* 16, 854. doi: 10.1186/s12864-015-2038-7

Cvitanich, J., Garate, O., and Smith, C. E. (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, 121–146. doi: 10.1111/j.1365-2761.1991.tb00584.x

Fata, A., Mahmoudian, M., Varasteh, A., and Sankian, M. (2013). Monarch-1 activation in murine macrophage cell line (J774 A.1) infected with Iranian strain of *Leishmania major. Iran. J. Parasitol.* 8 (2), 207–211.

Fernando, R. L., and Garrick, D. (2013). Bayesian methods applied to GWAS. *Genome-wide association studies and genomic prediction*, 1019, 237–274. doi: 10.1007/978-1-62703-447-0_10

Fryer, J. L., and Hedrick, R. P. (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

Goddard, M., and Hayes, B. (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., Wray, N. R., Verbyla, K., and Visscher, P. M. (2009). Estimating effects and making predictions from genome-wide marker data. Stat. Sci. 24, 517–529. doi: 10.1214/09-STS306.

Götz, S., García-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.* 2008 (10), 3420–3435. doi: 10.1093/nar/gkn176

Habier, D., Fernando, R. L., Kizilkaya, K., Garrick, D. J., Meuwissen, T., Hayes, B., et al. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* 12, 186. doi: 10.1186/1471-2105-12-186

Haenssler, E., and Isberg, R. R. (2011). Control of host cell phosphorylation by *Legionella pneumophila. Front. Microbiol.* 2, 64. doi: 10.3389/fmicb.2011.00064

Harris, N., Super, M., Rits, M., Chang, G., and Ezekowitz, R. A. (1992). Characterization of the murine macrophage mannose receptor: demonstration that the downregulation of receptor expression mediated by interferon-gamma occurs at the level of transcription. *Blood.* 80 (9), 2363–2373.

Hermann, C., Trowsdale, J., and Boyle, L. H. (2015). TAPBPR: a new player in the MHC class I presentation pathway. *Tissue Antigens* 85 (3), 155–166. doi: 10.1111/tan.12538

Hjeltnes, B., Bang-Jensen, B., Bornø, G., Haukaas, A., and Walde, C. S. (Ed.) (2018). The Health Situation in Norwegian Aquaculture 2017. Norwegian Veterinary Institute 108 p.

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

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 Genom.* 15, 90. doi: 10.1186/1471-2164-15-90

Hoyt, M., Hyman, A. A., and Bähler, M. (1997). Motor proteins of the eukaryotic cytoskeleton. *PNAS* 94 (24), 12747–12748. doi: 10.1073/pnas.94.24.12747

Jiang, L., Liu, X., Yang, J., Wang, H., Jiang, J., Liu, L., et al. (2014). Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genom.* 15 (1), 1105. doi: 10.1186/s12863-014-0125-4

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19 (9), 1639–1645. doi: 10.1101/gr.092759.109

Ku, C. Y., Liu, Y. H., Lin, H. Y., Lu, S. C., and Lin, J. Y. (2016). Liver fatty acid-binding protein (L-FABP) promotes cellular angiogenesis and migration in hepatocellular carcinoma. *Oncotarget* 7 (14), 18229–18246. doi: 10.18632/oncotarget.7571

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 (6), e39126. doi: 10.1371/journal.pone.0039126

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: detection of (co)-orthologs in large-scale analysis. *BMC Bioinform.* 28, 12:124. doi: 10.1186/1471-2105-12-124

Lee, S. H., Choi, B. H., Lim, D., Gondro, C., Cho, Y. M., Dang, C. G., et al. (2013). Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). *PLoS One* 8 (10), e74677. doi: 10.1371/journal.pone.0074677

Legarra, A., Ricard, A., and Filangi, O. (2013). GS3 software package and documentation. Accessed Jan. 2, 2017. http://snp.toulouse.inra.fr/~alegarra

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533 (7602), 200. doi: 10.1038/nature17164

Li, C., Wang, R., Su, B., Luo, Y., Terhune, J., Beck, B., et al. (2013). Evasion of mucosal defenses during *Aeromonas hydrophila* infection of channel catfish (*Ictalurus punctatus*) skin. *Dev. Comp. Immunol.*, 39(4), 447–455. doi: 10.1016/j.dci.2012.11.009

Ma, Y. J., Hein, E., Munthe-Fog, L., Skjoedt, M. O., Bayarri-Olmos, R., Romani, L., et al. (2015). Soluble collectin-12 (CL-12) is a pattern recognition molecule initiating complement activation *via* the alternative pathway. *J. Immunol.* 195 (7), 3365–3373. doi: 10.4049/jimmunol.1500493

Macqueen, D. J., Primmer, C. R., Houston, R. D., Nowak, B. F., Bernatchez, L., Bergseth, S., et al. (2017). Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genom.* 18, 1–9. doi: 10.1186/s12864-017-3862-8

Mandakovic, D., Glasner, B., Maldonado, J., Aravena, P., González, M., Cambiazo, V., et al. (2016). Genomic-based restriction enzyme selection for specific detection of *Piscirickettsia salmonis* by 16S rDNA PCR-RFLP. *Front. Microbiol.* 7, 643. doi: 10.3389/fmicb.2016.00643

Mitchell, A., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100

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 Genom.* 10, 368. doi: 10.1186/1471-2164-10-368

Moen, T., Torgersen, J., Santi, N., Davidson, W. S., Baranski, M., Ødegård, J., et al. (2015). Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* 115, 175406. doi: 10.1534/genetics.115.175406

Neet, K., and Hunter, T. (1996). Vertebrate non-receptor protein-tyrosine kinase families. *Genes Cells* 1 (2), 147–169. doi: 10.1046/j.1365-2443.1996.d01-234.x

Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., et al. (2015). The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol. Ecol. Resour.* 15, 662–672. doi: 10.1111/1755-0998.12337

Pearse, D., Barson, N., Nome, T., Gao, G., Campbell, M., Abadía-Cardoso, A., et al. (2018) Sex-dependent dominance maintains migration supergene in rainbow trout. *bioRxiv* 504621. doi: 10.1101/504621

Peters, S. O., Kizilkaya, K., Garrick, D. J., Fernando, R. L., Reecy, J. M., Weaber, R. L., et al. (2012). Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in *Brangus heifers*. *J. Anim. Sci.* 90, 3398–3409. doi: 10.2527/jas.2011-4507

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7 (5), e37135. doi: 10.1371/journal.pone.0037135

Pollard, D. J., Berger, C. N., So, E. C., Yu, L., Hadavizadeh, K., Jennings, P., et al. (2018). Broad-spectrum regulation of nonreceptor tyrosine kinases by the bacterial ADP-ribosyltransferase EspJ. *mBio* 9, 2) e00170–18. doi: 10.1128/mBio.00170-18

Pulgar, R., Hödar, C., Travisany, D., Zúñiga, A., Domínguez, C., Maass, A., et al. (2015). Transcriptional response of Atlantic salmon families to *Piscirickettsia salmonis* infection highlights the relevance of the iron-deprivation defence system. *BMC Genom.* 16, 495. doi: 10.1186/s12864-015-1716-9

Ramírez, R., Gómez, F. A., and Marshall, S. H. (2015). The infection process of *Piscirickettsia salmonis* in fish macrophages is dependent upon interaction with host-cell clathrin and actin. *FEMS Microbiol. Lett.* 362, 1–8. doi: 10.1093/femsle/fnu012

Robledo, D., Gutiérrez, A., Barría, A., Lhorente, J. P., Houston, R., and Yáñez, J. M. (2019). Discovery and functional annotation of quantitative trait loci affecting resistance to sea lice in Atlantic salmon. *Front. Genet.* 10, 56. doi: 10.3389/fgene.2019.00056

Rozas, M., and Enríquez, R. (2014). *Piscirickettsiosis* and *Piscirickettsia salmonis* in fish: a review. *J. Fish Dis.* 37, 163–188. doi: 10.1111/jfd.12211

Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell. Biol.* 5, 232–241. doi: 10.1038/nrm1335

Santana, M. H. A., Junior, G. A. O., Cesar, A. S. M., Freua, M. C., Gomes, R. C., Silva, S. L., et al. (2016). Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *J. Appl. Genet.* 57, 495–504. doi: 10.1007/s13353-016-0344-7

Schneider, M., Zimmermann, A. G., Roberts, R. A., Zhang, L., Swanson, K. V., Wen, H., et al. (2012). The innate immune sensor NLRC3 attenuates Toll-like receptor signaling *via* modification of the signaling adaptor TRAF6 and transcription factor NF-κB. *Nat. Immunol.* 13 (9), 823–831. doi: 10.1038/ni.2378

Schulert, G. S., and Allen, L. A. (2006). Differential infection of mononuclear phagocytes by *Francisella tularensis*: role of the macrophage mannose receptor. *J. Leukoc. Biol.* 80, 563–571. doi: 10.1189/jlb.0306219

Sefton, B. M., and Taddie, J. A. (1994). Role of tyrosine kinases in lymphocyte activation. *Curr. Opin. Immunol.* 6 (3), 372–379. doi: 10.1016/0952-7915(94)90115-5

Sernapesca (2018). Informe sanitario de salmonicultura en centros marinos. *Primer Semestre*, 2018.

Sever, L., Vo, N. T. K., Bols, N. C., and Dixon, B. (2014). Expression of tapasin in rainbow trout tissues and cell lines and up regulation in a monocyte/macrophage cell line (RTS11) by a viral mimic and viral infection. *Dev. Comp. Immunol.* 44, 86–93. doi: 10.1016/j.dci.2013.11.019

Shalaeva, D. N., Cherepanov, D. A., Galperin, M. Y., Golovin, A. V., and Mulkidjanian, A. Y. (2018). Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in relation to the basic catalytic mechanism. *Elife* 7, e37373. doi: 10.7554/eLife.37373

Soderlund, C., Bomhoff, M., and Nelson, W. (2011). SyMap v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* 39 (10), e68. doi: 10.1093/nar/gkr123

Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* 10, 513–525. doi: 10.1038/nrm2728

Vallejo, R. L., Leeds, T. D., Fragomeni, B. O., Gao, G., Hernandez, A. G., Misztal, I., 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. doi: 10.3389/fgene.2016.00096

Vallejo, R., Liu, S., Gao, G., Fragomeni, B. O., Hernandez, A. G., Leeds, T. D., et al. (2017). 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, 1–15. doi: 10.3389/fgene.2017.00156

Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Sixin, L., Parsons, J. E., et al. (2018). Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: evidence that long-range LD is a major contributing factor. *J. Anim. Breed. Genet.* 135, 263–274. doi: 10.1111/jbg.12335

Varona, L., García-Cortés, L., and Pérez-Enciso, M. (2001). Bayes factors for detection of quantitative trait loci. *Genet. Sel. Evol.* 33, 133. doi: 10.1186/1297-9686-33-2-133

Vidal, O., Noguera, J. L., Amills, M., Varona, L., Gil, M., Jiménez, N., et al. (2005). Identification of carcass and meat quality quantitative trait loci in a Landrace pig population selected for growth and leanness. *J. Anim. Sci.* 83, 293. doi: 10.2527/2005.832293x

Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with *P*-values. *Genet. Epidemiol.* 33, 79–86. doi: 10.1002/gepi.20359

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Dekkers, J. C. M., et al. (2016). Mixture models detect large effect QTL better than GBLUP and result in more accurate and persistent predictions. *J. Anim. Sci. Biotechnol.* 7, 7. doi: 10.1186/s40104-016-0066-z

Yano, K., Carter, C., Yoshida, N., Abe, T., Yamada, A., Nitta, , et al. (2014). Gimap3 and Gimap5 cooperate to maintain T-cell numbers in the mouse. *Eur. J. Immunol.* 44 (2), 561–572. doi: 10.1002/eji.201343750

Yáñez, J. M., Bangera, R., Lhorente, J. P., Barría, A., Oyarzún, M., Neira, R., et al. (2016). Negative genetic correlation between resistance against *Piscirickettsia salmonis* and harvest weight in coho salmon (*Oncorhynchus kisutch*). *Aquaculture* 459, 8–13. doi: 10.1016/j.aquaculture.2016.03.020

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. (2014). 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., Lopez, M. E., Bassini, L., Correa, K., Gilbey, J., et al. (2016). Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Mol. Ecol. Resour.* 16, 1002–1011. doi: 10.1111/1755-0998.12503

Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J. P., et al. (2018a). Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. *G3-Genes Genomes Genet.* 8, 719–726. doi: 10.1534/g3.117.300499

Yoshida, G. M., Carvalheiro, R., Lhorente, J. P., Correa, K., Barria, A., Figueroa, R. et al., (2018b). "Bayesian genome-wide association analyses reveal different genetic architecture of *Piscirickettsia salmonis* resistance in three salmonid species," in *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production*, (Auckland).

Yoshida, G. M., Lhorente, J. P., Carvalheiro, R., and Yáñez, J. M. (2017). Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (*Salmo salar L.*). *Anim. Genet.* 48, 698–703. doi: 10.1111/age.12621

Young, J. C., Clements, A., Lang, A. E., Garnett, J. A., Munera, D., Arbeloa, A., et al. (2014). The *Escherichia coli* effector EspJ blocks Src kinase activity *via* amidation and ADP ribosylation. *Nat. Commun.* 5, 5887. doi: 10.1038/ncomms6887

Conflict of Interest Statement: JL and KCo were employed by Benchmark Genetics Chile during the course of the study. The remaining 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.