Comparative genomic analysis of three salmonid species identifies functional candidate genes involved in resistance to the intracellular bacteria *Piscirickettsia salmonis*

Running title: Comparative analysis for resistance to *Piscirickettsia salmonis*

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

[1] Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Av. Santa Rosa 11735, La Pintana, Santiago 8820808, Chile.

[2] Núcleo Milenio INVASAL, Concepción, Chile.

[3] Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, El Líbano 5524, Macul, Santiago 7830490 Chile.

[4] Benchmark Genetics Chile, Ruta 7 Carretera Austral Km 35, Puerto Montt, Chile.

[5] Escuela de Medicina Veterinaria, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile.

[6] Fisheries and Oceans Canada 4160 Marine Drive, West Vancouver, BC, Canada V7V 1N6.

[7] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

[8] School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), Jaboticabal, Via de Acesso Prof. Paulo Donato Castellane, 14884-900, Jaboticabal, Brazil.

[9] National Council for Scientific and Technological Development (CNPq), Brasília, DF, 71605-001, Brazil.

27 ***Correspondence**: jmayanez@uchile.cl, rpulgar@inta.uchile.cl.

28 **Word count:**

29 **Number of figures: 4**

30 **Keywords: Coho salmon, rainbow trout, Atlantic salmon, *Piscirickettsia salmonis*,**

31 **Genome-wide Association study, comparative genomics, piscirickettsiosis.**

Abstract

Piscirickettsia salmonis is the etiological agent of Salmon Rickettsial Syndrome (SRS), and is responsible for considerable economic losses in salmon aquaculture. The bacteria affect 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 (GWAS) 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 (DD) and as binary survival (BS). A total of 828 CS, 2,130 RT and 2,601 AS individuals were phenotyped and then genotyped using ddRAD sequencing, 57K SNP Affymetrix® Axiom® and 50K Affymetrix® Axiom® SNP panels, respectively. Both trait 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*

57 infection. These results provide fundamental knowledge on the potential functional genes
58 underpinning resistance against *P. salmonis* in three salmonid species.
59
60 **Introduction**
61 Infectious diseases are responsible for large economic losses in salmon farming.
62 *Piscericketssia salmonis*, the causal agent of Salmon Rickettsial Syndrome (SRS), affects
63 several salmon species and is considered one of the major pathogens affecting the salmon
64 farming industry (Rozas and Enríquez, 2014). *P. salmonis* was identified in 1989 from farmed
65 coho salmon (*Oncorhynchus kisutch*) sampled in Chile (Cvitanich et al., 1991). Since then, *P.*
66 *salmonis* has been confirmed as the causative agent for clinical and chronic SRS in coho
67 salmon, Atlantic salmon (*Salmo salar*) and rainbow trout (*Onchorhyncus mykiss*) in several
68 countries, including: Norway, Canada, Scotland, Ireland and Chile (Fryer and Hedrick, 2003).
69 Current control protocols and treatments are based on antibiotics and vaccines. The
70 effectiveness of both strategies in field conditions is not optimal (Rozas and Enríquez, 2014).
71 From the total mortalities ascribed to infectious diseases in Chile, SRS is responsible for 18.3%,
72 92.6% and 67.9% in coho salmon, rainbow trout and Atlantic salmon, respectively (Sernapesca,
73 2018). These mortality rates, together with other factors such as antibiotic treatments and
74 vaccinations, have generated economic losses up to USD $ 450 million per year (Camusetti et
75 al., 2015).
76 A feasible and sustainable alternative to prevent disease outbreaks is genetic selection
77 for disease resistance (Bishop and Woolliams, 2014). The estimated levels heritability for
78 resistance to *P. salmonis* in coho salmon, Atlantic salmon and rainbow trout, range from 0.11
79 to 0.41 (Bangera et al., 2017; Barría et al., 2018a; Bassini et al., submitted; Correa et al., 2015;
80 Yáñez et al., 2016; Yoshida et al., 2018a); demonstrating the feasibility of improving *P.*
81 *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; Macqueen et al., 2017; Palti et al., 2015; Yañez et al., 2016). The use of genotypes from dense panels of 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 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 Atlantic salmon farmed populations (Houston et al., 2008; Moen et al., 2009). To date, the number of IPN outbreaks has been significantly reduced in Norwegian Atlantic salmon populations because of MAS for IPNV resistance (Hjeltnes, 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 (Barría et al., 2018a; Correa et al., 2015), suggesting that the implementation of genomic selection (GS) is the most appropriate strategy to accelerate the genetic progress for this trait. Methods which can model all available SNPs simultaneously, including Bayesian regression methods (Fernando & 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 GWAS (Goddard et al. 2009).

Regarding that *P. salmonis* affects farmed populations of three phylogenetically related salmonid species, including coho salmon, Atlantic salmon and rainbow trout, 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 coho salmon, Atlantic salmon and rainbow trout 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 coho salmon, Atlantic salmon and rainbow trout 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**.**

Material 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 coho salmon (CS), Atlantic salmon (AS) and rainbow trout (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. (submitted) and Yáñez et al. (2013, 2014, 2016). Prior to the beginning of each experimental challenge, qPCR was performed in a sub-sample of each population to confirm the absence of *Flavobacterium* spp, Infectious Salmon Anemia Virus (ISAV) and IPNV. Subsequently, fish were intraperitoneally (IP) injected with 0.2 ml of a LD_{50} inoculum of *P. salmonis*. Post 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 qRT-PCR and necropsy. Each experimental challenge was performed at Aquainnovo's Research Station, Xth Region, Chile.

Genotyping

A total of 828 CS, 2,130 RT and 2,601 AS were genotyped using ddRAD, 57K SNP Affymetrix® Axiom® 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), while DNeasy Blood & Tissue (Qiagen) was used for RT and AS.

For CS, ten double digest Rad-seq (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 v. 1.41 (Catchen et al., 2011, 2013). Rad-tags which passed the *process_radtags* quality control were aligned to the coho salmon 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*, and followed by *populations* using default parameters. Quality control (QC) included the removal of SNPs below the following thresholds: Hardy-Weinberg Equilbrium (HWE) $p<1x10-6$, 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 USDA (Palti et al., 2015). SNPs were filtered with the following QC parameters: HWE p < 1x10-6, MAF < 0.05, and SNP call rate < 0.95. Individuals with call rates lower than 0.95 were also removed.

The 50K SNP Affymetrix® Axiom® array used to genotype AS, was developed by Universidad de Chile and Aquainnovo (Correa et al., 2015; Yañez et al., 2016). These markers were selected from a 200K array, as described in detail in Correa et al. (2015). Genotypes were quality-controlled using the following criteria: HWE p < 1x10-6, MAF < 0.05, SNP and samples were discarded when the genotype rate was < 0.95.

Genome-wide association analysis

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

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

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- π) with a mean equal to zero and variance equal to σ_a^2, having in turn a scaled inverse χ^2 prior, with $v_a = 4$ and $v_e = 10$ degrees of freedom (df) 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, 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 by each significant SNP was calculated as:

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

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 genetic variance explained 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:

$$h^2 = \frac{V_A'}{V_A' + \sigma_e^2} \qquad \text{(EQ3)}$$

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

Comparative genomic analysis

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

To identify candidate genes associated with *P. salmonis* resistance, we used a comparative genomic analysis between coho salmon CS, AS and RT. 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), AS (GCF_000233375.1, Lien et al., 2016) and RT (GCF_002163495.1, Pearse et al., bioRxiv). 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 in order 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 were 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, 2005) (referred to as: Group A);

ii) Orthologous and paralogous genes among species were identified using the ProteinOrtho tool (Lechner et al., 2011). Multi-directional 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) were recovered and classified as high priority genes (Group C); and

iv) The complete set of genes located at the intersection of more than one 1 Mb windows within a species were 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%, 8.3% to 73.7% and 8% to 100% for CS, AS and RT, respectively. This result suggests that the phenotypic variation for this trait could be related with 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, 0.24 for AS and 0.48 for RT. When resistance was defined as BS, genomic heritability estimates increased to 0.88, 0.32 and 0.64 for CS, AS and RT, respectively, representing moderate to high levels of genetic variation for *P. salmonis* resistance.

Genome-wide association analysis

A total of 580 CS (9,389 SNPs), 2,383 AS (42,624 SNPs) and 1,929 RT (24,916 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 genetic variance explained by a single marker was not higher than 5% (**Figure 2 and Supplementary Figure 1**). While 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 and RT, both traits appear to have oligogenic control with few moderate to large effect loci, and a large-unknown number of loci each having a small effect on the traits. For BS, the top 200 SNPs explained about 80% and 90% of the phenotypic variance in CS and RT, respectively, while in AS they explained slightly more than 30%. For DD, the top ten SNPs explained a 40%, 57% and 17% of the total genetic variance for *P. salmonis* resistance in CS, RT and AS, respectively.

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, AS and RT, and searched for genes within 1Mb 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 the **Supplementary Table S1: Sheets 1 to 6.**

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 salmonids species are shown in **Figure 4**. Remarkably, around 30% of the proteins from genes present in regions associated with DD belong to five homologous superfamilies (*P-loop containing nucleoside triphosphate hydrolase, 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 containing nucleoside triphosphate hydrolase, immunoglobulin-like fold and immunoglobulin-like domain superfamily*). Interestingly, the *P-loop containing nucleoside triphosphate hydrolase* superfamily (also known as P-loop_NTPase) contained the largest group of proteins for both traits, and at least one representative protein from each salmonid species belonged 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 in 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, while 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_NTPases superfamily, as well as members of the orthologous groups such as fatty acid-binding proteins (FABP). Other genes encoding proteins classically associated to immune response such as tripartite motif-containing protein 35 (TRIM35) and lysozyme C II (LYZ) 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), that 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 (TAPBPL). 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: dependent on kinase activity, GTP

hydrolysis, helicase activity, lipid metabolism, cytoskeletal dynamics and inflammation. In order 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, B, C and 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 the **Supplementary Table S1. sheet: Candidate Genes.**

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 which aims at identifying and prioritizing functional candidate genes involved in the differential response against bacterial infection by means of comparing results from genome-wide association mapping across different phylogenetically related salmonid species.

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; 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 rainbow trout 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, 0.41, and 0.44, for CS, AS and RT, respectively (Bassini et al., submitted; Yáñez et. al., 2013, 2014; 2016). When

352 heritability for *P. salmonis* resistance was estimated based on genomic information, the
353 maximum values previously reported were 0.39, and 0.62, for AS and RT, respectively
354 (Bangera et al., 2017; Yoshida et al., 2018a).

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

366 Based on the linkage disequilibrium (LD) of the Atlantic salmon population, (measured
367 as r^2), the number of SNPs used for AS (~ 43K) should be enough to cover the entire genome
368 (Barría et al., 2018b). There is a lack of studies aimed at evaluating the LD and population
369 structure of the current farmed rainbow trout population. Based on results from a different
370 rainbow trout farmed population, at least 20K SNPs are necessary to cover the whole genome
371 (Vallejo et al. 2018). If LD levels of the present rainbow trout population are similar to those
372 reported by Vallejo et al. (2018), the 23K SNPs used here will most likely cover the whole
373 genome. However, this is not the case for CS. Using a high density SNP array Rondeau et al.
374 (2018, in prep.) and Barría et al. (2018c), suggested that at least 74K SNPs are necessary for
375 whole-genome studies of the current coho salmon population. The small number of SNPs

376 assayed in this study for CS (9,389), most likely affected the identification of markers with a
377 moderate to high effect on resistance to *P. salmonis* in this species.

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

391 When traits are polygenic in nature, the identification of genes underlying them is a
392 challenging task and often depends on previous knowledge of the function of genes adjacent
393 to the associated SNPs (Jiang *et al*., 2014; Bouwman *et al*., 2018; Robledo et al., 2019). Our
394 strategy was based on identifying orthologous proteins between the salmonid species and
395 families of homologous proteins in the complete set of proteins adjacent to all the SNPs that
396 explained more than 1% of the genetic variance, without searching for a specific function. The
397 identification of genes directly associated with the innate immune response, after applying all
398 the classification criteria, such as lysozyme C II, macrophage mannose receptor 1, collectin-12
399 and tapasin-related protein, suggests that our strategy was successful in finding strong
400 functional candidate genes involved in resistance to *P. salmonis*. Interestingly, around one

401 hundred genes not classically associated with the immune system were also identified; from
402 which seventeen of them were part of at least two of the groups described previously and hence
403 are considered strong candidates for being responsible on trait variation **(Table 2).**

404 Previously, lysozymes have primarily been described as having a bacteriolytic activity
405 against gram-positive bacteria; however, the expression of lysozyme C II has been shown to
406 be induced in a resistant rainbow trout line in response to *Flavobacterium psychrophilum*
407 infection (Langevin et al., 2012) and in Atlantic salmon families in response to *Piscirickettsia*
408 *salmonis* infection (Pulgar et al., 2015), indicating that the transcriptional regulation of this
409 enzyme in salmonids responds to gram-negative bacterial infection. Macrophage mannose
410 receptor 1 and collectin-12 are membrane receptors which display several functions associated
411 with innate immunologic defense, particularly in the recognition of carbohydrate structures of
412 pathogens and as phagocytic receptors of bacteria, yeasts and other pathogenic microorganisms
413 (Harris *et al*., 1992; Ma *et al*., 2015). It has been reported that enhanced infection in human
414 phagocytes with *Francisella tularensis*, a bacterium phylogenetically related to *P. salmonis*, is
415 mediated by the macrophage mannose receptor (Schulert and Allen, 2006), while collectin-12
416 led to the activation of the alternative pathway of complement via association with properdin,
417 a key positive regulator of the pathway by increment of the half-life of the C3 and C5
418 convertases (Ma *et al*., 2015). Tapasin-related protein has been described as a second MHC
419 class I-dedicated chaperone, essential to providing specificity for T cell responses against
420 viruses and bacteria (Hermann et al., 2015) and the related protein tapasin has been shown to
421 be induced in monocyte/macrophage in rainbow trout by chum salmon reovirus infection
422 (Sever et al., 2014).

423 Another set of candidate genes for SRS resistance in the three salmonid species studied
424 are a cluster of cytosolic non-receptor tyrosine-protein kinases (nRTKs). These proteins are a
425 subgroup of the tyrosine kinase family, enzymes that phosphorylate tyrosine residues of

proteins, and regulate many cellular functions, such as: cell growth and survival, apoptosis, cell adhesion, cytoskeleton remodeling and differentiation (Neet and Hunter, 1996). Although these genes 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 non-receptor tyrosine protein kinases is critical to the activation of lymphocytes by an antigen (Sefton and Taddie, 1994). Moreover, some cellular signaling pathways are hijacked by intracellular pathogens (bacteria and viruses), thus pathogens can subvert protein phosphorylation to control host immune responses and facilitate invasion and dissemination. It has been described that some bacterial effectors are injected into host cells through their secretion systems where they inhibit the Src kinase (a subfamily of nRTKs). In particular, the effector EspJ, an ADP-ribosyltransferase of the bacteria *Escherichia coli* and *Citrobacter rodentium*, regulates multiple non-receptor tyrosine kinases *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 GTPase ADP-ribosylation factor protein 5B (ARL5B), suggesting that an adequate regulation of the activity of nRTKs 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. The 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, the RAB GTPases constitute a subfamily of small GTPases known as master regulators of intracellular membrane traffic (Stenmark, 2009). Since *P. salmonis* drives the formation of host membrane-derived organelles, the development of these *P. salmonis*-

451 containing vacuoles (PCVs) are dependent on the bacterium's ability to usurp the intracellular
452 membrane system of the fish. Interestingly, *Legionella pneumophila*, the closest phylogenetical
453 bacterium to *P. salmonis*, disturbs the intracellular vesicular trafficking of infected human cells
454 by recruiting the RAB1 to the cytosolic face of the Legionella-containing vacuole (LCV)
455 through the activity of its effector protein DRRA (Müller et al., 2010), suggesting that *P.*
456 *salmonis* could use similar mechanisms for the formation and maintenance of its replicative
457 vacuole. Furthermore, two orthologous of FYVE, RhoGEF and PH domain-containing proteins
458 were identified in RT and AS. These proteins activate CDC42, a GTPase involved in the
459 organization of the actin cytoskeleton and with a role in early contractile events in phagocytes
460 (Ching et al., 2007). Since it has been described that the infective process of *P. salmonis*
461 depends on the exploitation of the actin monomers (Ramírez et al., 2015), the identification in
462 this study of candidate genes that encode for cytoskeletal motor proteins (two kinesins and a
463 myosin), highlights their relevance not only for the reorganization of the cytoskeleton, but also
464 for its motility and involvement in the development of the infection (Hoyt et al., 1997).
465 Remarkably, two other candidate proteins associated with SRS resistance are also members of
466 the GTPase superfamily, the GTPases of the immunity-associated proteins (GIMAPs) 4 and 8.
467 This is a family of proteins abundantly expressed in lymphocytes and whose function is to
468 contribute in the regulation of apoptosis and the maintenance of T-cell numbers in the organism
469 (Yano et al., 2014).

470 Another group of orthologous genes code for ATP-dependent RNA helicases DDX24
471 in CS and DDX47 in RT for DD. The ATP-dependent RNA helicase DDX family, also known
472 as DEAD-box helicases, is required for different cellular processes such as transcription, pre-
473 mRNA processing, ribosome biogenesis, nuclear mRNA export, translation initiation, RNA
474 turnover and organelle function. The protein structure is very similar to viral RNA helicases
475 and to DNA helicases, which suggests that the fundamental activities of these enzymes are

476 similar (Rocak and Linder, 2004). Viruses also utilize RNA helicases at various stages of their
477 life cycle. Many viruses carry their own helicases to assist with the synthesis of their genome,
478 but others synthesize their genome within the cell nucleus which tends to exploit cellular
479 helicases and thus do not encode their own. We also identified the ATP-dependent RNA
480 helicase TDRD9, which has not been directly implicated in infection, but was differentially
481 expressed in channel catfish (*Ictalurus punctatus*) in response to *Aeromonas hydrophila*
482 infection (Li et al., 2012). Mechanistic studies of RNA helicases will allow the determination
483 of the precise role of these helicases in the host/pathogen interaction.

484 The last group of orthologous genes identified code for two liver fatty acid binding
485 proteins (L-FABP) in CS and RT for BS. Liver fatty acid-binding proteins are abundant in
486 hepatocytes and are known to be associated with lipid metabolism. In addition, these proteins
487 are up-regulated in several types of cancer but their role in infection remains unclear (Ku et al.,
488 2016). Nevertheless, it has been recently reported that serum and urine L-FABP may be a new
489 diagnostic marker for liver damage in patients with both acute and chronic hepatitis C infection
490 (Cakir et al., 2017). Interestingly, in Atlantic salmon challenged with *P. salmonis*, L-FABP
491 was up-regulated in resistant families and simultaneously down-regulated in susceptible
492 families (Pulgar et al., 2015), suggesting a transcriptional regulation in response to *P. salmonis*
493 infection and a putative expression marker of resistance to SRS.

494 Genes coding NACHT, LRR and PYD domains-containing protein 12 (NLRP12);
495 NACHT, LRR and CARD domains-containing protein 3 (NLRC3); voltage-dependent L-type
496 calcium channel subunit beta-2 (CACNB2); heparan sulfate glucosamine 3-O-sulfotransferase
497 3A1 (HS3ST3A1) and histidine triad nucleotide-binding protein 1 (HINT1) were also selected
498 as candidate genes for SRS resistance. NLRP12 and NLRC3 are two cytosolic proteins that
499 share two functional domains (NACHT and LRR). NLRP12 was one of the best ranked genes,
500 adjacent to the leader SNP and adjacent to more than one SNP simultaneously for DD in AS.

501 This protein functions as an attenuating factor of inflammation in monocytes, by negative
502 regulation of the NF-κB activation (Fata et al., 2013). In murine macrophages, a significant
503 expression increase has been shown in cells infected with the intracellular parasite *Leishmania*
504 *major* compared to non-infected macrophages (Fata et al., 2013). NLRC3 is also a negative
505 regulator of the innate immune response mediated by the inhibition of Toll-like receptor (TLR)-
506 dependent activation of the transcription factor NF-κB (Schneider et al., 2012). The presence
507 of these genes suggests that the control of the inflammatory reaction in response to *P. salmonis*
508 infection could be essential to combat SRS.

509 Finally, some members of heparan sulfate glucosamine 3-O-sulfotransferase (like
510 HS3ST3A1) have been shown to mediate the herpes simplex virus type-1 (HSV-1) entry and
511 spread in zebrafish (Antoine et al., 2014). Also, some members of CACNB2 changed their
512 expression levels in response to *P. salmonis* in multiples tissues of Atlantic salmon (Tacchi et
513 al., 2011), while HINT1 responds transcriptionally to *Salmonella typhimurium* and infectious
514 pancreatic necrosis virus (IPNV) infection in zebrafish and Atlantic salmon, respectively
515 (Stockhammer et al., 2010; Robledo et al., 2016).

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

525

Conclusions

Although *P. salmonis* resistance has previously been described as polygenic trait, our comparative genomics approach based on GWAS results for the same trait in different salmonid species allowed us to identify around one hundred 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: dependent 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 disease affecting salmon aquaculture.

Data availability

Genotype and phenotype data for Atlantic salmon and rainbow trout are available on the online digital repository figshare figshare.com/s/e70a3a84c05ea60e1074 and figshare.com/s/221a39319b236d46f9fc respectively. Coho salmon genotype and phenotype are at available doi.org/10.5061/dryad.b273q6p). and at 10.6084/m9.figshare.7886183.

Acknowledgements

Aguas Claras, Pesquera Antares and Salmones Chaicas provided the CS, RT and AT datasets, respectively. Thanks to FAPESP (2014/20626-4; 2015/25232-7) and CNPq (308636/2014-7) for financial support. AB and KC acknowledge the National Commission of Scientific and Technologic Research (CONICYT) for the funding through the National PhD funding

551 program. RP acknowledge the National Commission of Scientific and Technologic Research
552 (CONICYT) for the funding through the Fondecyt program (11161083). AB acknowledges the
553 Government of Canada for the funding through the Canada-Chile Leadership Exchange
554 Scholarship. We also want to thank the World Congress on Genetics Applied to Livestock
555 Production (WCGALP) as this work has been partially presented on this conference (Yoshida
556 et al. 2018b)

557

558 **Ethics approval and consent to participate**

559 All experimental challenges and sampling procedures were approved by the Comité de Bioética
560 Animal from the Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile
561 (Certificate N08-2015).

562

563 **Funding**

564 This project was funded by the U-Inicia grant, from the Vicerrectoria de Investigación y
565 Desarrollo, Universidad de Chile. This work was conceived of under the framework of the
566 grant FONDEF NEWTON-PICARTE (IT14I10100), funded by CONICYT (Government of
567 Chile). This work has been partially supported by Núcleo Milenio INVASAL from Iniciativa
568 Científica Milenio (Ministerio de Economía, Fomento y Turismo, Gobierno de Chile). This
569 research was carried out in conjunction with EPIC4 (Enhanced Production in Coho: Culture,
570 Community, Catch), a project supported by the government of Canada through Genome
571 Canada, Genome British Columbia, and Genome Quebec.

572

573 **Authors' contributions**

574 JMY conceived of and designed the study and drafted the manuscript. GrY assessed the GWAS
575 analyses. AP and RP assessed the comparative genomic analyses and contributed with the first

576 draft of the manuscript and discussion. LB, MEL and KC contributed with the RT and AS
577 sampling, genotyping and quality controls. AB performed DNA extraction from CS samples,
578 contributed with initial draft of the manuscript and performed library construction. KrC
579 performed ddRAD library construction and assessed the comparative sequences analyses
580 between species. RC and JPL contributed with study design, analyses and discussion. All
581 authors have reviewed and approved the manuscript.

582

583 **Conflict of Interest**

584 JPL was hired by Aquainnovo at the moment of the study. All other authors have no conflicts
585 of interest to declare.

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600 **Table 1**. Estimates of total additive genetic variance (V'_a), residual variance (σ_e^2), heritability
601 (h^2) and standard deviation (SD) for resistance against *Piscirickettsia salmonis* in three
602 salmonids species.

Species	**Days to death**				**Binary survival**			
	V'_a	σ_e^2	h^2	**SD**	V'_a	σ_e^2	h^2	**SD**
Coho salmon	28.91	60.70	0.32	0.07	7.53	1.00	0.88	0.03
Rainbow trout	30.42	32.71	0.48	0.04	1.87	1.00	0.64	0.05
Atlantic salmon	16.52	53.17	0.24	0.04	0.47	1.00	0.32	0.05

603

604

605

606 **Table 2.** Summary of candidate genes associated with *P. salmonis* resistance for coho salmon
607 (CS), rainbow trout (RT) and Atlantic salmon (AS) ranked by score, which is simply based on
608 the number of appearance of each gene across the following categories: i) species (CS, RT and
609 AS); ii) trait definitions (DD and BS); and iii) groups (A, B, C and D).

Gene Symbol	Protein description	Species	Trait	Group	Score[1]
NRTPK	non-receptor tyrosine-protein kinase (cytosolic)	CS, RT and AS	DD and BS	B, C and D	8
DDX	ATP-dependent RNA helicase DDX	CS and RT	DD	A, B and C	6
ARL5B	ADP-ribosylation factor protein 5B	CS	DD and BS	A and D	5
L-FABP	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
GIMAP4	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
COLEC12	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

610
611 [1] The maximum score possible for one particular gene was equal to 9.
612
613
614
615
616
617
618
619
620
621
622
623
624

625 **Bibliography**

626 Antoine, T. E., Yakoub, A., Maus, E., Shukla, D., & Tiwari, V. (2014). Zebrafish 3-O-
627 sulfotransferase-4 generated heparan sulfate mediates HSV-1 entry and spread. PloS one,
628 9(2), e87302. doi:10.1371/journal.pone.0087302

629 Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R., and Yáñez, J. M. (2017). Genomic
630 predictions can accelerate selection for resistance against Piscirickettsia salmonis in
631 Atlantic salmon (Salmo salar). *BMC Genomics* 18, 121.

632 Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., *et al.*
633 (2018a). Genomic predictions and genome-wide association study of resistance against
634 piscirickettsia salmonis in coho salmon (Oncorhynchus kisutch) using ddRAD
635 sequencing. *G3 Genes Genomes Genet.* 4231, g3.200053.2018.

636 Barría, A., Lopez, M. E., Yoshida, G., Carvalheiro, R., and Yáñez, J. M. (2018b). Population
637 genomic structure and genome-wide linkage disequilibrium in farmed Atlantic salmon
638 (Salmo salar L.) using dense SNP genotypes. *Front. Genet.* doi: 10.33.

639 Barría, A., Christensen, K. A., Yoshida, G., Jedlicki, A., Lhorente, J. P., Davidson, W. S., et
640 al. (2018c). Whole genome linkage disequilibrium and effective population size in a
641 coho salmon (Oncorhynchus kisutch) breeding population. bioRxiv [Preprint]. doi:
642 10.1101/335018.

643 Bishop, S. C., and Woolliams, J. A. (2014). Genomics and disease resistance studies in
644 livestock. *Livest. Sci.* 166, 190–198.

645 Bourne H, David A. Sanders & Frank McCormick. (1991). The GTPase superfamily:
646 conserved structure and molecular mechanism. Nature 349, pages117–127.

647 Bouwman, A. C. et al. (2018). Meta-analysis of genome-wide association studies for cattle
648 stature identifies common genes that regulate body size in mammals. Nature Genetics
649 (2018). doi:10.1038/s41588-018-0056-5

650 Cakir OO, Toker A, Ataseven H, Demir A, Polat H.(2017). The Importance of Liver-Fatty
651 Acid
652 Binding Protein in Diagnosis of Liver Damage in Patients with Acute Hepatitis. J Clin
653 Diagn Res. 2017 Apr;11(4):OC17-OC21. doi: 10.7860/JCDR/2017/24958.9621.
654 Camusetti, M. A., Gallardo, A., Aguilar, D., and Larenas, J. (2015). Análisis de los costos por
655 la utilización de quimioterápicos y vacunas en la salmonicultura. *Salmonexpert*.
656 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks:
657 An analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140.
658 Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., and De Koning,
659 D.-J. (2011). Stacks: Building and genotyping loci de novo from short-read sequences.
660 *G3 Genes Genomes Genet.* 1, 171–182.
661 Ching K.H., Kisailus A.E., Burbelo P.D. (2007). Biochemical characterization of distinct
662 regions of SPEC molecules and their role in phagocytosis. Exp. Cell Res. 313:10-21.
663 Correa, K., Lhorente, J., Lopez, M., Bassini, L., Naswa, S., Deeb, N., et al. (2015). Genome-
664 wide association analysis reveals loci associated with resistance against Piscirickettsia
665 salmonis in two Atlantic salmon (Salmo salar L.) chromosomes. *BMC Genomics* 16,
666 854.
667 Cvitanich, J., Garate, O., and Smith, C. E. (1991). The isolation of a rickettsia-like organism
668 causing disease and mortality in Chilean salmonids and its confirmation by Koch's
669 postulate. *J. Fish Dis.* 14, 121–146.
670 Fata, A., Mahmoudian, M., Varasteh, A., & Sankian, M. (2013). Monarch-1 Activation in
671 Murine Macrophage Cell Line (J774 A.1) Infected with Iranian Strain of Leishmania
672 major. Iranian journal of parasitology, 8(2), 207-11.
673 Fernando, R. L., and Garrick, D. (2013). "Bayesian Methods Applied to GWAS," in, 237–
674 274. doi:10.1007/978-1-62703-447-0_10.

675 Fryer, J. L., and Hedrick, R. P. (2003). Piscirickettsia salmonis: a Gram-negative intracellular
676 bacterial pathogen of fish. *J. Fish Dis.* 26, 251–262.
677 *Goddard, M., and Hayes, B. (2009). Mapping genes for complex traits in domestic animals*
678 *and their use in breeding programmes. Nat. Rev. Genet. 10, 381–391. doi:*
679 *10.1038/nrg2575*
680 Götz S., García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón
681 M, Dopazo J, Conesa A (2008). High-throughput functional annotation and data mining
682 with the Blast2GO suite. Nucleic Acids Res. 2008 (10):3420-35. doi:
683 10.1093/nar/gkn176. Epub 2008 Apr 29.
684 Habier, D., Fernando, R. L., Kizilkaya, K., Garrick, D. J., Meuwissen, T., Hayes, B., et al.
685 (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*
686 12, 186. doi:10.1186/1471-2105-12-186.
687 Harris N, Super M, Rits M, Chang G, Ezekowitz RA. (1992). Characterization of the murine
688 macrophage mannose receptor: demonstration that the downregulation of receptor
689 expression mediated by interferon-gamma occurs at the level of transcription. Blood.
690 1;80(9):2363-73.
691 Hermann C, Trowsdale J, Boyle LH (2015). TAPBPR: a new player in the MHC class I
692 presentation pathway. Tissue Antigens ;85(3):155-66. doi: 10.1111/tan.12538.
693 Hjeltnes, B. (Editor) (2018). Fish Health Report 2018. *Nor. Vet. Institute, Oslo.*
694 Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al.
695 (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in
696 Atlantic salmon (Salmo salar). *Genetics* 178, 1109–1115.
697 Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., Downing, A., et al.
698 (2014). Development and validation of a high density SNP genotyping array for Atlantic
699 salmon (Salmo salar). *BMC Genomics* 15, 90.

Hoyt M., Anthony A. Hyman, and Martin Bähler. (1997). Motor proteins of the eukaryotic cytoskeleton. PNAS, 94 (24) 12747-12748.

Jiang, L., Liu, X., Yang, J., Wang, H., Jiang, J., Liu, L., He, S., Ding, X., Liu J., and Zhang, Q. (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genomics 15:1* **15,** 1105.

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., Jones S.J., Marra M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Research 19(9): 1639-45.

Ku, C. Y., Liu, Y. H., Lin, H. Y., Lu, S. C., & Lin, J. Y. (2016). Liver fatty acid-binding protein (L-FABP) promotes cellular angiogenesis and migration in hepatocellular carcinoma. Oncotarget, 7(14), 18229-46.

Langevin C, Blanco M, Martin SA, Jouneau L, Bernardet JF, Houel A, Lunazzi A, Duchaud E, Michel C, Quillet E, Boudinot P. (2012). Transcriptional responses of resistant and susceptible fish clones to the bacterial pathogen Flavobacterium psychrophilum. PLoS One. 7(6):e39126.

Lechner M., Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011 Apr 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). 8. doi:10.1371/journal.pone.0074677.

Legarra, A., Ricard, A., and Filangi, O. (2013). GS3 software package and documentation.

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon

725 genome provides insights into rediploidization. Nature(2016) 533:200.
726 doi:10.1038/nature17164
727 Li C, Wang R, Su B, Luo Y, Terhune J, Beck B, Peatman E. (2012). Evasion of mucosal
728 defenses during Aeromonas hydrophila infection of channel catfish (Ictalurus punctatus)
729 skin. Dev Comp Immunol. 2013 Apr;39(4):447-55. doi: 10.1016/j.dci.2012.11.009.
730 Ma YJ, Hein E, Munthe-Fog L, Skjoedt MO, Bayarri-Olmos R, Romani L, Garred P (2015).
731 Soluble Collectin-12 (CL-12) Is a Pattern Recognition Molecule Initiating Complement
732 Activation via the Alternative Pathway. J Immunol. 1;195(7):3365-73. doi:
733 10.4049/jimmunol.1500493.
734 Macqueen, D. J., Primmer, C. R., Houston, R. D., Nowak, B. F., Bernatchez, L., Bergseth, S.,
735 et al. (2017). Functional Annotation of All Salmonid Genomes (FAASG): an
736 international initiative supporting future salmonid research, conservation and
737 aquaculture. *BMC Genomics* 18, 1–9.
738 Mandakovic D, Glasner B, Maldonado J, Aravena P, González M, Cambiazo V, Pulgar R.
739 (2016). Genomic-Based Restriction Enzyme Selection for Specific Detection of
740 Piscirickettsia salmonis by 16S rDNA PCR-RFLP. Frontiers in Microbiology 7 (643).
741 Mitchell A., Teresa K Attwood, Patricia C Babbitt, Matthias Blum, Peer Bork, Alan Bridge,
742 Shoshana D Brown, Hsin-Yu Chang, Sara El-Gebali, Matthew I Fraser, Julian Gough,
743 David R Haft, Hongzhan Huang, Ivica Letunic, Rodrigo Lopez, Aurélien Luciani, Fabio
744 Madeira, Aron Marchler-Bauer, Huaiyu Mi, Darren A Natale, Marco Necci, Gift Nuka,
745 Christine Orengo, Arun P Pandurangan, Typhaine Paysan-Lafosse, Sebastien Pesseat,
746 Simon C Potter, Matloob A Qureshi, Neil D Rawlings, Nicole Redaschi, Lorna J
747 Richardson, Catherine Rivoire, Gustavo A Salazar, Amaia Sangrador-Vegas, Christian J
748 A Sigrist, Ian Sillitoe, Granger G Sutton, Narmada Thanki, Paul D Thomas, Silvio C E
749 Tosatto, Siew-Yit Yong and Robert D Finn (2019). InterPro in 2019: improving

coverage, classification and access to protein sequence annotations. *Nucleic Acids Research, Jan 2019; 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 Genomics* 10, 368.

Moen T, Torgersen J, Santi N, Davidson WS, Baranski M, Ødegård J, Kjøglum S, Velle B, Kent M5, Lubieniecki KP, Isdal E, Lien S (2015). Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. Genetics.115.175406.

Müller MP, Peters H, Blümer J, Blankenfeldt W, Goody RS, Itzen A. (2010). The Legionella effector protein DrrA AMPylates the membrane traffic regulator Rab1b. Science;329(5994):946-9. doi: 10.1126/science.1192276. Epub 2010 Jul 22.

Neet K, Hunter T. (1996). Vertebrate non-receptor protein-tyrosine kinase families. Genes Cells. Feb;1(2):147-69.

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.

Pearse D,, Barson N., Nome T., Gao G., Campbell M., Abadía-Cardoso A., *et al.* (2018) bioRxiv 504621; doi: https://doi.org/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

775 digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in
776 model and non-model species. *PLoS One* 7. doi:10.1371/journal.pone.0037135.
777 Pollard D., Cedric N. Berger, Ernest C. So, Lu Yu, Kate Hadavizadeh, Patricia Jennings,
778 Edward W. Tate, Jyoti S. Choudhary, Gad Frankel (2018). Broad-Spectrum Regulation
779 of Nonreceptor Tyrosine Kinases by the Bacterial ADP-Ribosyltransferase EspJ. mBio
780 Apr 2018, 9 (2) e00170-18; DOI: 10.1128/mBio.00170-18.
781 Pulgar R., Hödar C., Travisany D., Zúñiga A., Domínguez C., Maass A., González M.,
782 Cambiazo V. (2015). Transcriptional response of Atlantic salmon families to
783 Piscirickettsia salmonis infection highlights the relevance of the iron-deprivation defence
784 system. BMC Genomics 16:495. DOI: 10.1186/s12864-015-1716-9.
785 Ramírez, R., Ǵomez, F. A., and Marshall, S. H. (2015). The infection process of
786 Piscirickettsia salmonis in fish macrophages is dependent upon interaction with host-cell
787 clathrin and actin. *FEMS Microbiol. Lett.* 362, 1–8. doi:10.1093/femsle/fnu012.
788 Robledo, D., Gutiérrez, A., Barría, A., Lhorente, JP., Houston, R., and Yáñez, JM (2019).
789 Discovery and Functional Annotation of Quantitative Trait Loci Affecting Resistance to
790 Sea Lice in Atlantic Salmon. Frontiers in Genetics 10, 56.
791 Robledo, D., Taggart, J. B., Ireland, J. H., McAndrew, B. J., Starkey, W. G., Haley, C. S.,
792 Hamilton, A., Guy, D. R., Mota-Velasco, J. C., Gheyas, A. A., Tinch, A. E., Verner-
793 Jeffreys, D. W., Paley, R. K., Rimmer, G. S., Tew, I. J., Bishop, S. C., Bron, J. E., …
794 Houston, R. D. (2016). Gene expression comparison of resistant and susceptible Atlantic
795 salmon fry challenged with Infectious Pancreatic Necrosis virus reveals a marked contrast
796 in immune response. BMC genomics, 17, 279. doi:10.1186/s12864-016-2600-y.
797 Rozas, M., and Enríquez, R. (2014). Piscirickettsiosis and Piscirickettsia salmonis in fish: a
798 review. *J. Fish Dis.* 37, 163–188.
799 Rocak S., and Patrick Linder. (2004). DEAD-box proteins: the driving forces behind RNA

800 metabolism. Nature Reviews Molecular Cell Biology volume 5, pages 232–241.

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

805 Schneider, M., Zimmermann, A. G., Roberts, R. A., Zhang, L., Swanson, K. V., Wen, H.,
806 Davis, B. K., Allen, I. C., Holl, E. K., Ye, Z., Rahman, A. H., Conti, B. J., Eitas, T. K.,
807 Koller, B. H., … Ting, J. P. (2012). The innate immune sensor NLRC3 attenuates Toll-
808 like receptor signaling via modification of the signaling adaptor TRAF6 and transcription
809 factor NF-κB. Nature immunology, 13(9), 823-31.

810 Schulert, G. S., and Allen, L. A. (2006). Differential infection of mononuclear phagocytes by
811 Francisella tularensis: role of the macrophage mannose receptor. J. Leukoc. Biol. 80,
812 563–571. doi: 10.1189/jlb.0306219.

813 Sefton BM, Taddie JA (1994). Role of tyrosine kinases in lymphocyte activation. Curr Opin
814 Immunol. 6(3):372-9.

815 Sernapesca (2018). Informe sanitario de salmonicultura en centros marinos. Primer Semestre
816 2018.

817 Sever, L., Vo, N. T. K., Bols, N. C. & Dixon, B (2014). Expression of tapasin in rainbow trout
818 tissues and cell lines and up regulation in a monocyte/macrophage cell line (RTS11) by a
819 viral mimic and viral infection. Developmental & Comparative Immunology 44, 86–93.

820 Shalaeva DN, Cherepanov DA, Galperin MY, Golovin AV, Mulkidjanian AY (2018).
821 Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in
822 relation to the basic catalytic mechanism. Elife.7: e37373. doi:10.7554/eLife.37373

823 Soderlund C., Bomhoff M., Nelson W., (2011). SyMap v3.4: a turnkey synteny system with
824 application to plant genomes. Nucleic Acid Research 39(10)

825 Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. Nat. Rev. Mol. Cell
826 Biol. 10, 513-525. doi:10.1038/nrm2728
827 Stockhammer OW, Rauwerda H, Wittink FR, Breit TM, Meijer AH, Spaink HP. (2010).
828 Transcriptome analysis of Traf6 function in the innate immune response of zebrafish
829 embryos. Mol Immunol. Nov-Dec;48(1-3):179-90. doi: 10.1016/j.molimm.2010.08.011.
830 Tacchi L, Bron JE, Taggart JB, Secombes CJ, Bickerdike R, Adler MA, et al.(2011). Multiple
831 tissue transcriptomic responses to Piscirickettsia salmonis in Atlantic salmon (Salmo
832 salar). Physiol Genomics.43(21):1241–54.
833 Vallejo, R. L., Leeds, T. D., Fragomeni, B. O., Gao, G., Hernandez, A. G., Misztal, I., et al.
834 (2016). Evaluation of genome-enabled selection for bacterial cold water disease
835 resistance using progeny performance data in rainbow trout: Insights on genotyping
836 methods and genomic prediction models. *Front. Genet.* 7, 1–13.
837 Vallejo, R., Liu, S., Gao, G., Fragomeni, B. O., Hernandez, A. G., Leeds, T. D., et al. (2017).
838 Similar genetic architecture with shared and unique quantitative trait loci for bacterial
839 cold water disease resistance in two rainbow trout breeding populations. *Front. Genet.* 8,
840 1–15.
841 Varona, L., García-Cortés, L., and Pérez-Enciso, M. (2001). Bayes factors for detection of
842 Quantitative Trait Loci. *Genet. Sel. Evol.* 33, 133. doi:10.1186/1297-9686-33-2-133.
843 Vidal, O., Noguera, J. L., Amills, M., Varona, L., Gil, M., Jiménez, N., et al. (2005).
844 Identification of carcass and meat quality quantitative trait loci in a Landrace pig
845 population selected for growth and leanness. *J. Anim. Sci.* 83, 293.
846 doi:10.2527/2005.832293x.
847 Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with *P*
848 -values. *Genet. Epidemiol.* 33, 79–86. doi:10.1002/gepi.20359.
849 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 T, Ishimaru N, Takada K, Butcher GW, Takahama Y (2014) Gimap3 and Gimap5 cooperate to maintain T-cell numbers in the mouse. Eur J Immunol.;44(2):561-72. doi: 10.1002/eji.201343750. Epub 2013 Oct 23.

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.

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.

Yañez, 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.

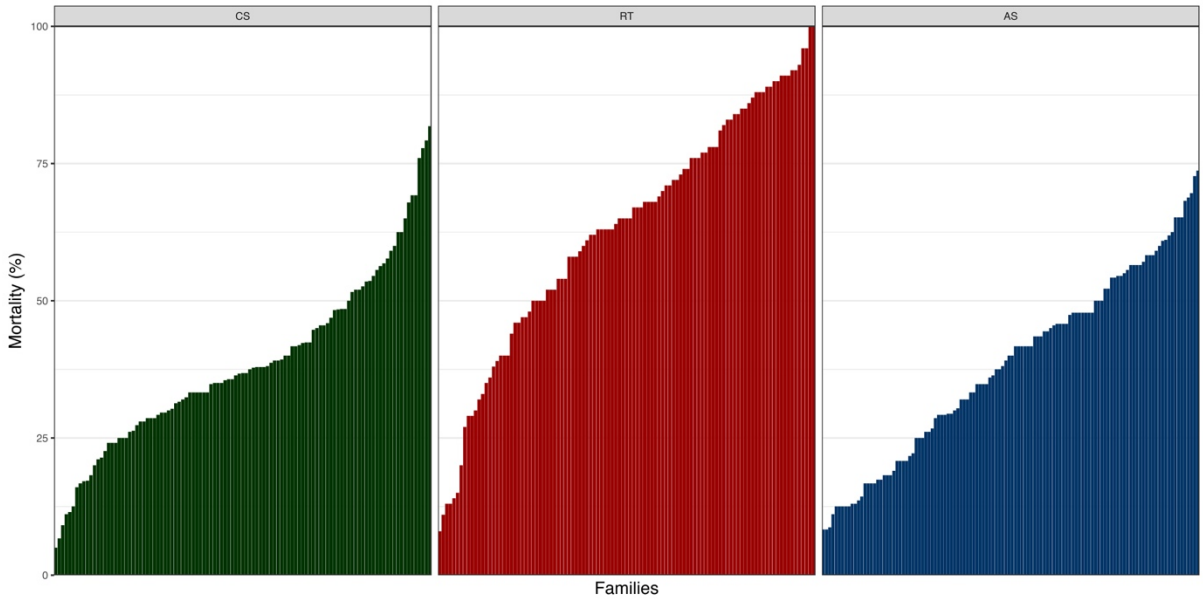
Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J. P., et al. (2018a). Genomic prediction accuracy for resistance against Piscirickettsia salmonis in farmed rainbow trout. *G3 Genes Genomes Genet.* 8, 719–726. doi:10.1534/g3.117.300499.

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.

Young J., Abigail Clements, Alexander E. Lang, James A. Garnett, Diana Munera, Ana Arbeloa, Jaclyn Pearson, Elizabeth L. Hartland, Stephen J. Matthews, Aurelie Mousnier, David J. Barry, Michael Way, Andreas Schlosser, Klaus Aktories & Gad Frankel (2014). The Escherichia coli effector EspJ blocks Src kinase activity via amidation and ADP ribosylation. Nature Communications volume 5, Article number: 5887.



895

896 **Figure 1**. Cumulative mortality by family after *Piscirickettsia salmonis* experimental infection

897 of coho salmon (CS), rainbow trout (RT) and Atlantic salmon (AS). For CS, RT and AS a total

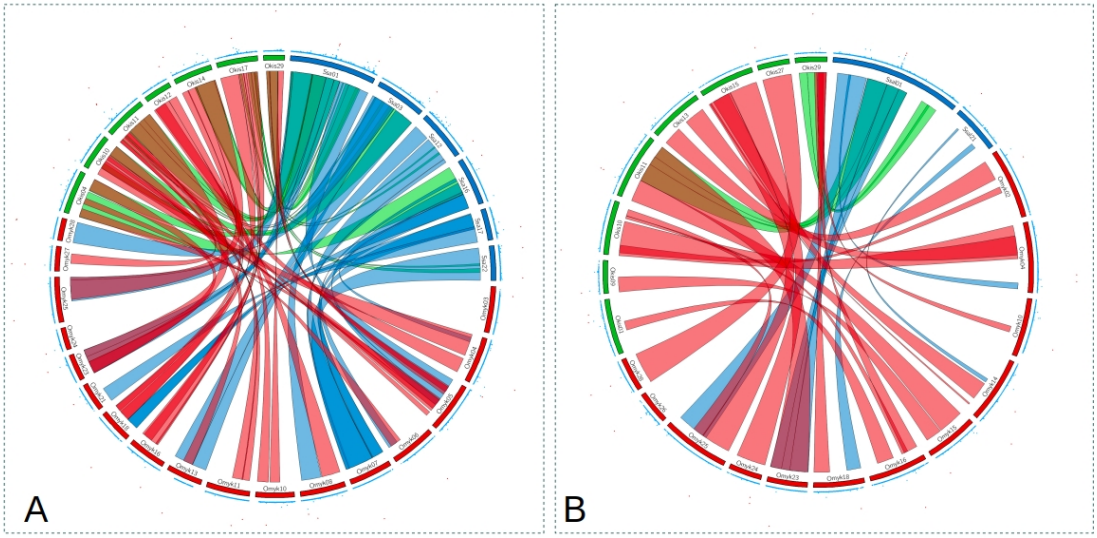
898 of 107, 105 and 118 full-sib families were experimentally challenged.

899

900

901

902



903

904 **Figure 2.** Circos plot for *P. salmonis* resistance as day of death (A) and as binary survival (B).
905 The inner ribbons mark syntenic regions between coho salmon (green) rainbow trout (red) and
906 Atlantic salmon (blue). Manhattan plots are showed on the outer ring, with significant
907 associations plotted in red (values >= 1).

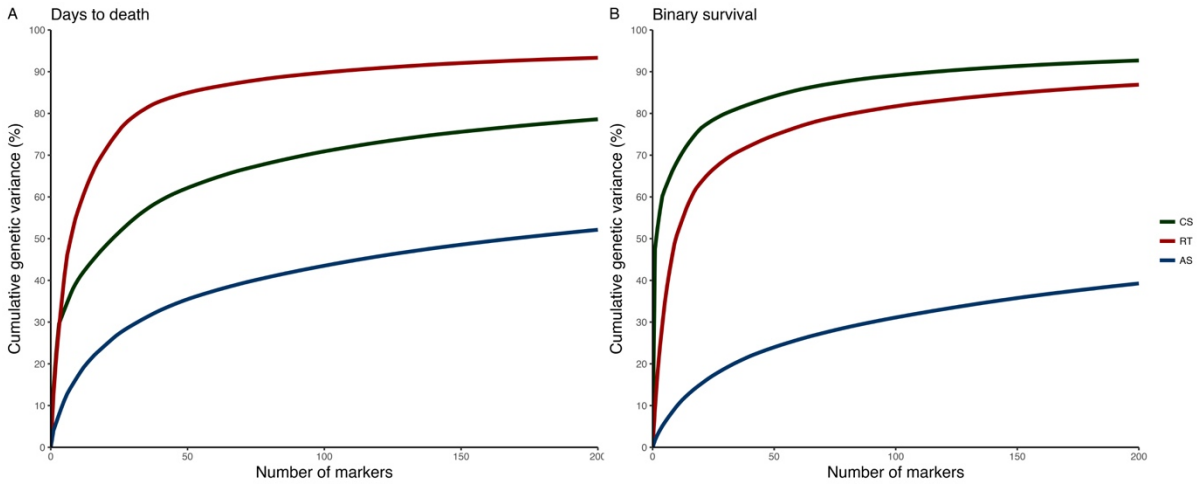
908

909

910

911

912



913 **Figure 3**. Cumulative percentage of the genetic variance explained (GEV) by the top 200
914 markers from Bayesian GWAS for resistance to *P. salmonis* measured as days to death (DD)
915 (A) and binary survival (BS) (B) in Coho salmon (CS), Rainbow trout (RT) and Atlantic
916 salmon (AS).

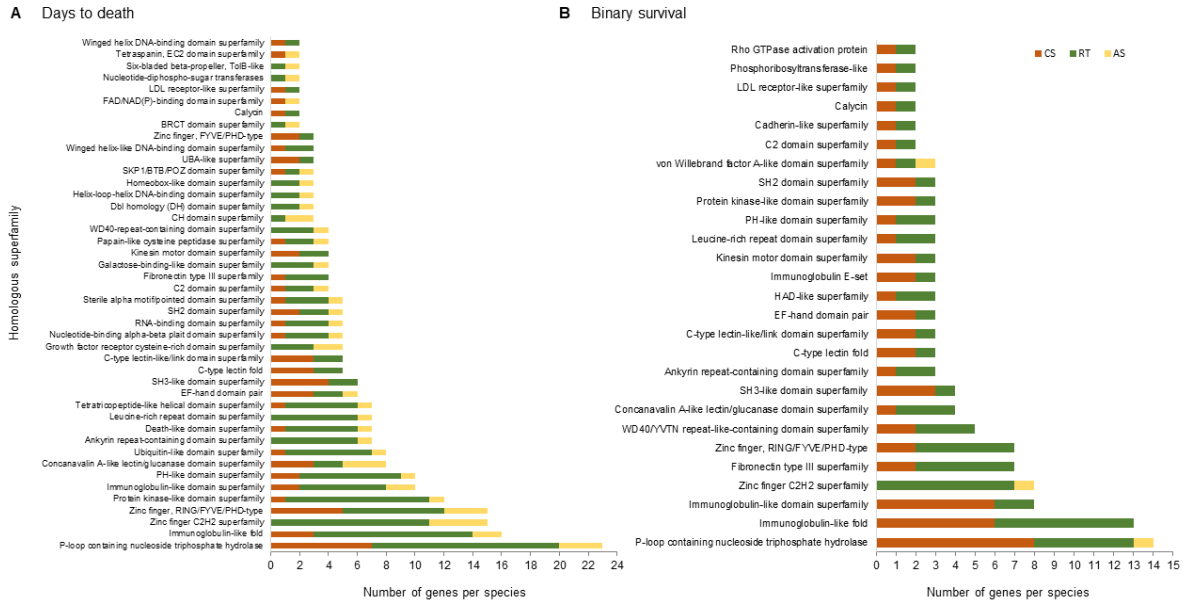
917

918

919

920

921



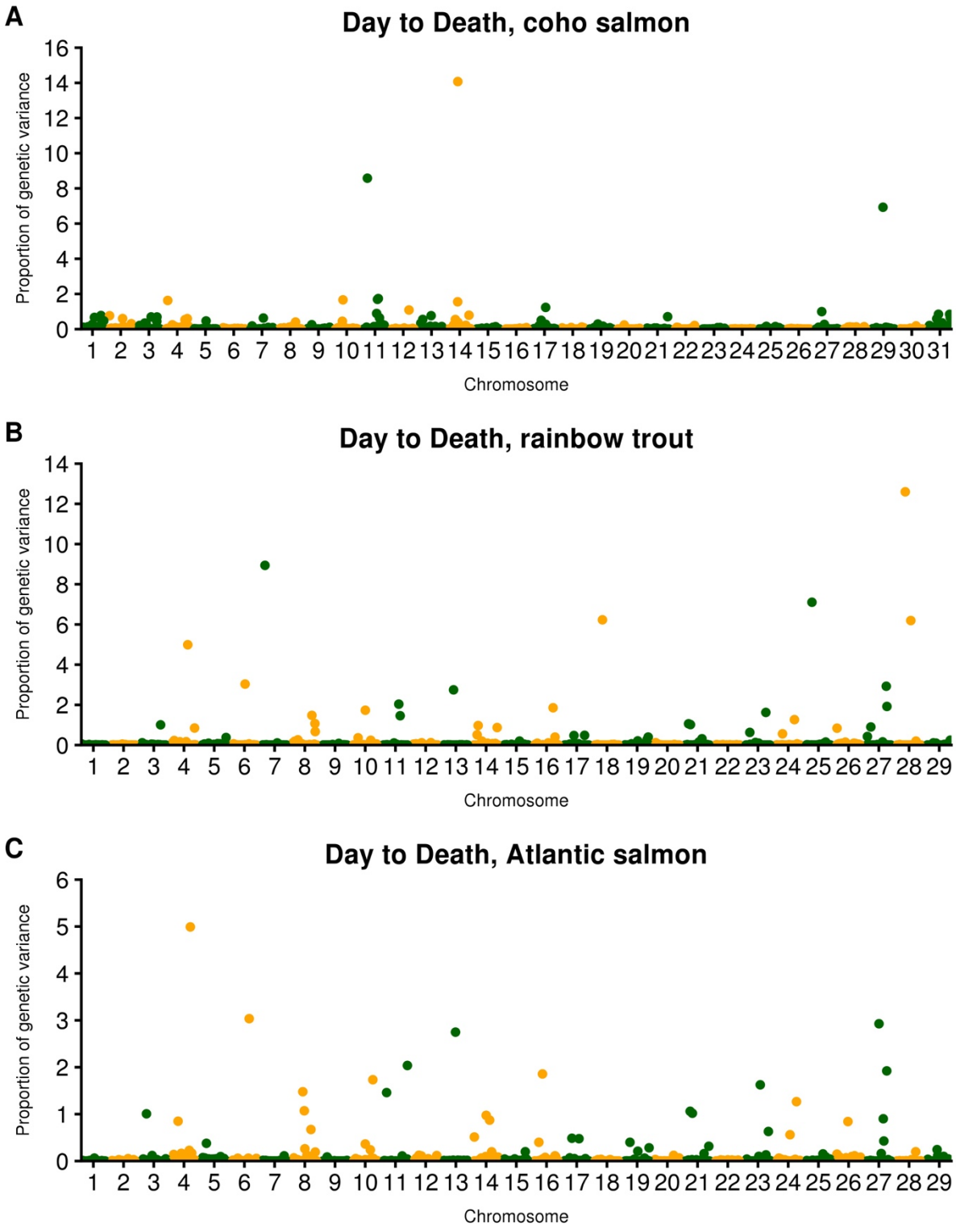
922

923 **Figure 4**. Homologous superfamilies (InterPro) adjacent to the complete set of SNPs that
924 explain over 1% of the genetic variance of resistance to SRS measured as days to death (DD)
925 (A) and binary survival (BS) (B). Bars represent the abundance of genes in each homologous
926 superfamily present in at least two salmonids species. Coho salmon (CS), rainbow trout (RT)
927 and Atlantic salmon (AS).

928

929

930



931

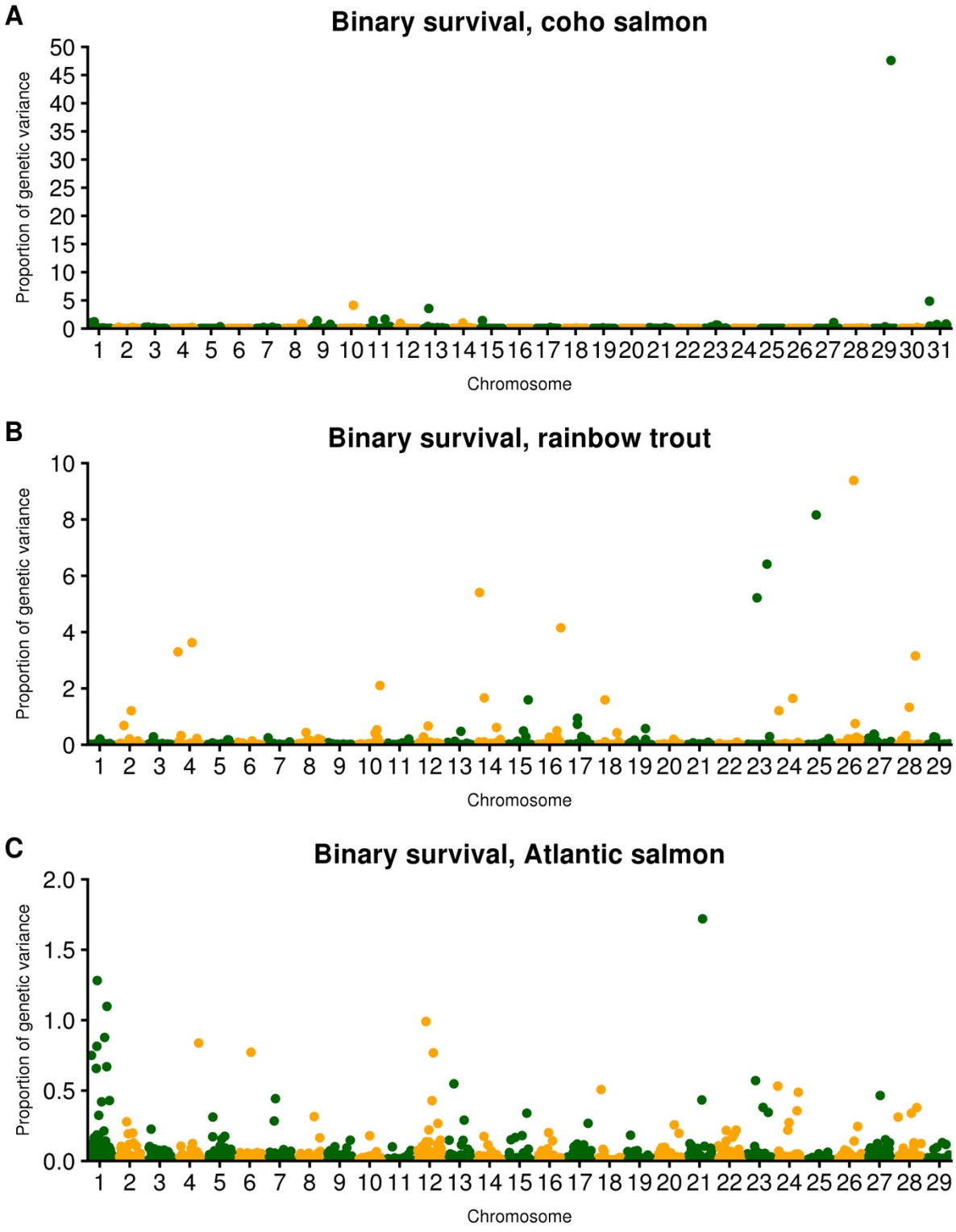
932 **Figure S1**. Manhattan plots for resistance to *P. salmonis* measured as day of death (DD) in

933 coho salmon (CS), rainbow trout (RT) and Atlantic salmon (AS). Y-axis represents percentage

934 of the genetic variance explained by each marker.

935

936



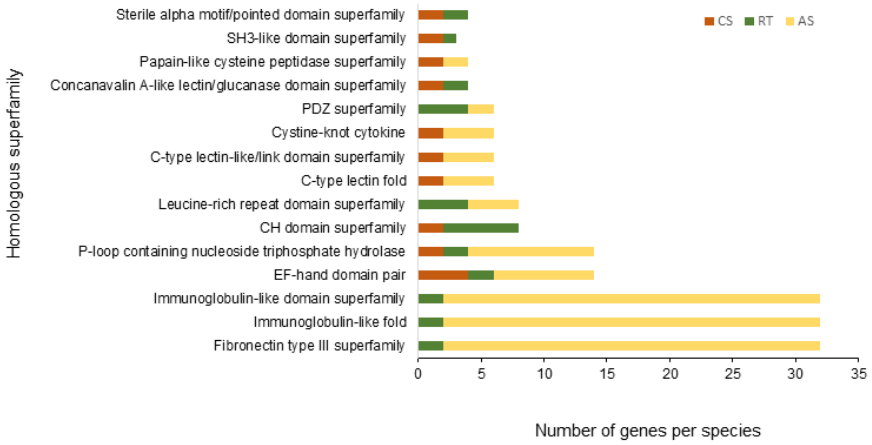
937

938 **Figure S2**. Manhattan plots for resistance to *P. salmonis* measured as binary survival (BS) in

939 coho salmon (CS), rainbow trout (RT) and Atlantic salmon (AS). Y-axis represents percentage

940 of the genetic variance explained by each marker.

941



942

943 **Figure S3.** Homologous superfamilies (InterPro) associated to 100 random selected proteins

944 from coho salmon (CS), rainbow trout (RT) and Atlantic salmon (AS) genomes.

945