Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout

Grazyella M. Yoshida*,†,1, Rama Bangera§,1, Roberto Carvalheiro†, Katharina Correa‡, René Figueroa‡, Jean P. Lhorente‡, José M. Yáñez*,‡,**,2

* Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago 8820808, Chile.
† Animal Science Department, School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), Campus of Jaboticabal, 14884-900, Jaboticabal, Brazil.
§ Akvaforsk Genetics, Auragata 3, 6600 Sunndalsora, Norway.
‡ Aquainnovo, Puerto Montt, Chile.
** Núcleo Milenio INVASAL, Concepción, Chile.
1 Both authors contributed equally to this work.

ORCID IDs: 0000-0002-6788-7369 (GM.Y.); 0000-0002-5394-2427 (RB.); 0000-0002-4506-0555 (RC.); 0000-0003-1386-8522 (K.C.); 0000-0002-1645-9399 (RF.); 0000-0002-9157-4231 (JP.L.); 0000-0002-6612-4087 (JM.Y.)
Genomic selection for *P. salmonis*

Keywords: disease resistance; genomic selection; *Oncorhynchus mykiss*; reliability

Corresponding author: Facultad de Ciencias Agronómicas, Universidad de Chile, 11735 Santa Rosa avenue, La Pintana, 8820808, Santiago, Chile

Phone number: +56 2 2978 55 33

E-mail: jmayanez@uchile.cl
ABSTRACT

Salmonid Rickettsial Syndrome (SRS), caused by the intracellular bacterium *Piscirickettsia salmonis*, is one of the main diseases affecting rainbow trout (*Oncorhynchus mykiss*) farming. To accelerate genetic progress, genomic selection methods can be used as an effective approach to control the disease. The aim of this study were: “(i) to compare the accuracy of estimated breeding values (EBV) using pedigree-based best linear unbiased prediction (PBLUP) with genomic BLUP (GBLUP), single-step GBLUP (ssGBLUP), Bayes C and Bayesian Lasso (LASSO), and (ii) to test the accuracy of genomic prediction and pedigree-based BLUP using different marker densities (0.5, 3, 10, 20 and 27K) for resistance against *P. salmonis* in rainbow trout. Phenotypes were recorded as number of days to death (DD) and binary survival (BS) from 2,416 fish challenged with *P. salmonis*. A total of 1,934 fish were genotyped using 57K single nucleotide polymorphism (SNP) array. All genomic prediction methods achieved higher accuracies than PBLUP. The relative increase in accuracy for different genomic models ranged from 28 to 41% for both DD and BS at 27K SNP. Between different genomic models, the highest relative increase in accuracy was obtained with Bayes C (~ 40%), where 3K SNP was enough to achieve a similar accuracy as the 27K SNP for both traits. For resistance against *P. salmonis* in rainbow trout we showed that genomic predictions using GBLUP, ssGBLUP, Bayes C and LASSO can increase accuracy compared to PBLUP. Moreover, it is possible to use relatively low-density SNP panels for genomic prediction without compromising accuracy predictions for resistance against *P. salmonis* in rainbow trout.
INTRODUCTION

In 1989, *Piscirickettsia salmonis* was identified as a pathogenic bacteria causing the Salmonid Rickettsial Syndrome (SRS) in farmed coho salmon (*Oncorhynchus kisutch*) in Chile (Cvitanich et al. 1991; Branson and Diaz-Munoz 1991). Since then, *P. salmonis* has been confirmed as the causative agent for SRS in coho salmon, Atlantic salmon (*Salmo salar*) and rainbow trout (*Onchorhyncus mykiss*) in several countries including Norway, Canada, Scotland, Ireland and Chile (Fryer and Hedrick 2003; Rozas and Enríquez 2014). The economic losses related to SRS in Chile in the year 2012 were US$450 million due to mortality, antibiotic treatment and vaccinations (Camussetti et al. 2015).

Currently, treatment for bacterial diseases in the aquaculture industry is predominantly based on antibiotics (Peña et al. 2016). Although several vaccines are available for prevention of SRS, none of them provide complete protection against *P. salmonis* in field conditions (Kuzyk et al. 2001; Tobar et al. 2011). In addition, selective breeding can be used to alleviate disease problems. The levels of genetic variation for resistance to *P. salmonis*, with heritability values ranging from 0.11 to 0.41, have demonstrated the feasibility to improve the trait by means of artificial selection in salmon breeding populations (Yañez et al. 2013; 2014; 2016a; Lhorente et al. 2014).

With the recent advances in genotyping methods and the development of single nucleotide polymorphism (SNP) panels for salmonids (Houston et al. 2014; Palti et al. 2015; Yáñez et al. 2016; Macqueen et al. 2017), genetic markers linked with quantitative trait loci (QTL) can be identified and implemented in breeding programs through marker-assisted selection (MAS) (Yáñez et al. 2014). For example, in Atlantic
salmon one major QTL for infectious pancreatic necrosis virus (IPNV) resistance was
detected, explaining 29% and 83% of the phenotypic and genetic variances, respectively
(Gheyas et al. 2010; Houston et al. 2010; Houston et al. 2008a; Houston et al. 2008b).
This QTL have been successfully used in MAS programs in this species (Moen et al.
2015). However, genome-wide association studies (GWAS) in Atlantic salmon
suggested that resistance against P. salmonis is a trait with moderate polygenic control,
with many markers explaining a small proportion of the genetic variance (Correa et al.
2015). The complexity of this trait and the absence of QTL with major effects suggest
that the implementation of MAS could be not successful in this particular case. In
contrast, the genomic selection (GS) will be the most appropriate way to incorporate the
genomic information to accelerate the genetic progress for traits were the markers have
small effect.

Genomic evaluations using dense SNP markers have been shown to increase
accuracy of estimated breeding values (EBV) compared to pedigree-based methods for
different economically important traits in Atlantic salmon (Ødegård et al. 2014; Tsai et
al. 2015, 2016; Sae-Lim et al. 2017; Bangera et al. 2017; Correa et al. 2017) and
rainbow trout (Vallejo et al. 2016, 2017). Different GS methods have been tested and
prediction accuracy varies depending on the method used, which mainly differ with
respect to the assumption about markers effects and genetic relationship matrix
calculation. The genomic best linear unbiased predictor (GBLUP) assumes that all
markers effects come from a normal distribution (Meuwissen et al. 2001; VanRaden
2008) and the relationship matrix is calculated using genomic information only. The
ssGBLUP assumes the same normal distribution for marker effects; however, uses of a
combination of pedigree and genomic information to determine the additive genetic
relationship matrix (Aguilar et al. 2010). In general, Bayesian methods assume more
flexible and non-normal distributed marker effects. For instance, Bayes C method assumes that SNP effects have independent and identical mixture distributions (Habier et al. 2011), while the Bayesian Lasso (LASSO) assumes a double exponential prior distribution for variances of SNP marker effects (Aguilar et al. 2010).

The performance of the different GS methods have been tested for different livestock species and traits (Hayes et al. 2010; Colombani et al. 2013; Neves et al. 2014; Chen et al. 2014). The best method in terms of accuracy will depend on some factors, such as the number of phenotyped animals, heritability, effective population size, size of the genome, marker density and genetic architecture of the trait (Daetwyler et al. 2008; Meuwissen 2009; Goddard 2009). In general, Bayesian methods usually outperform GBLUP method for the traits which are affected by a few large QTL, while for traits that are affected by many QTL with small effects, GBLUP would likely perform better than or similar as the Bayesian methods (Chen et al. 2014). Furthermore, Hayes et al. (2010) suggested that the results obtained from cattle may not be relevant for other species, due the larger linkage disequilibrium (LD) blocks in bovine than other species.

Therefore, it is valuable to compare the accuracy of different GS methodologies to identify the method that will result in the highest accuracy for the genetic evaluation of resistance to one of the most important bacterial diseases affecting sea rearing of rainbow trout, which turns is one of most widely distributed aquaculture species in the world. In addition, GS can also be implemented using a cost-effective individual genotyping strategy using low-density panels without much loss in information (Cleveland and Hickey 2014). Recent empirical studies demonstrated that low density panels are sufficient to get higher accuracy of genomic estimated breeding values (GEBVs) than EBV obtained from pedigree best linear unbiased prediction (PBLUP).
for resistance against *P. salmonis* (Bangera *et al.* 2017) and sea lice (Tsai *et al.* 2016; Correa *et al.* 2017) in Atlantic salmon.

The objective of this study were: (i) to compare the accuracy of estimated breeding values (EBV) using pedigree-based best linear unbiased prediction (PBLUP) with genomic BLUP (GBLUP), single-step GBLUP (ssGBLUP), Bayes C and Bayesian Lasso (LASSO), and (ii) to test the accuracy of genomic prediction and pedigree-based BLUP using different marker densities (0.5, 3, 10, 20 and 27K) for resistance against *P. salmonis* in rainbow trout.

**MATERIALS AND METHODS**

**Challenge test and phenotypes**

The rainbow trout (*O. mykiss*) used in this study were obtained from the breeding nucleus of Aguas Claras S.A. (Puerto Montt, Chile) and were challenge tested for resistance against *P. salmonis* at the Aquaculture Technology Center Patagonia/Aquainnovo, Puerto Montt, Chile (Flores-Mara *et al.* 2017). The fish used in this study were from the year-class 2011, which has undergone three generations of selection for growth, carcass quality and appearance traits. Juveniles from 105 families (representing progeny from 105 dams and 48 sires) were reared in separate tanks until individually tagged using a Passive Integrated Transponder-tag (PIT-tag) at an average weight of 7g. After tagging the animals were communally reared in a single tank for about seven months before being transferred to Aquainnovo's Research Station (Lenca River, Xth Region, Chile). The fish were subjected to acclimation period during 20 days at the research station. After this period, a total of 2,416 juveniles (with an average of 23 fish per family and ranging from 15 to 30 individuals) were experimentally challenged with *P. salmonis*. Before the challenge test, all fish were proven to be negative to the
presence of Infectious Salmon Anemia virus, Infectious Pancreatic Necrosis virus and *Renibacterium salmoninarum* by RT-PCR and *Flavobacterium spp.* by culture. Fish were infected by injecting 0.2 ml of a LD50 inoculum of *P. salmonis* through intra-peritoneal (IP) injection. Post IP injection, infected fish were equally distributed by family into three different tank replicates (used as fixed effect for PBLUP and genomics models). The challenge test continued for 32 days and mortality and weight at the end of the experiment were recorded in all fish. All surviving fish at day 32 were anesthetized and euthanized. Tissue samples (fin clips) for genomic DNA isolation were taken from all dead and surviving fish and preserved in 95% ethanol at −80 °C.

Resistance to salmon rickettsial syndrome (SRS) was defined as the number of days to death (DD), with values ranging from 5 to 32, and as binary survival (BS), scored as 1 if the fish died during the challenge test and 0 if the fish survived until the end of the challenge test.

**Genotypes**

The genotyped individuals were selected to obtain a balanced number of animals per family (mean = 19, range from 12 to 26) and maintain the phenotypic variance. Genomic DNA was extracted from fin clip samples from 2,130 fish (average of 19 fish per family, range from 12 to 26 fish) using the commercial DNeasy Blood & Tissue Kit, Qiagen, following the manufacturer's instructions. The fish were genotyped using a commercially available 57K Affymetrix® Axiom® SNP array, designed by The National Center for Cool and Cold Water Aquaculture at the United States Department of Agriculture (USA) (Palti *et al.* 2015).

The genotypes were quality controlled (QC) using Affymetrix’s Software AXIOM Analysis Suite using the default settings (Dish QC ≥ 0.82 and genotype call
rate $\geq 97\%$ for each sample). Additional quality control steps were conducted by filtering out SNPs and samples with a Hardy–Weinberg equilibrium test $p$-value less than 0.00001, SNP call rate lower than 0.90 and a minor allele frequency lower than 0.01.

Statistical Models

Pedigree-based BLUP

The pedigree-based variance components and EBV were estimated using BLUP and were compared with genomic evaluations. The model used was as follows:

$$ y = X\beta + Zg + e \quad \text{(M1)} $$

where $y$ is a vector of phenotypes (DD or BS), $\beta$ is a vector of fixed effects (tank and body weight), $g$ is a vector of random additive polygenic genetic effects that follows a normal distribution $\sim N(0, A\sigma^2_g)$, $X$ and $Z$ are incidence matrices, $A$ is the additive relationship matrix, and $e$ is the random residual error with a distribution $\sim N(0, I\sigma^2_e)$ and $I$ is the identity matrix (Lynch and Walsh 1998). Body weight was included as a covariate in the analysis given that is significantly ($p$-value < 0.05) affecting both traits. This is most likely due to the fact that inoculum was IP injected in the same dose for all fish, disregarding their initial size.

Genomic BLUP

The SNP based variance components and GEBV were estimated using GBLUP, similar to the PBLUP model (M1), as implemented in the BLUPF90 software package (Misztal et al. 2016). GBLUP model is a modification of the PBLUP method, where $g$ is a vector of random additive genetic polygenic effects with a distribution $\sim N(0, G\sigma^2_g)$ and $G$ is the genomic relationship matrix as described by VanRaden (2008). The $G$ matrix is
constructed based on all markers and it can differ from the pedigree-based numerator relationship matrix \(A\), in which the former can potentially have some negative off-diagonal values when individuals are molecularly less related than average pairs of animals in the sense of identity by state if the population were in Hardy-Weinberg equilibrium (Toro et al. 2002). The variance components, PBLUP and GBLUP solutions for the breeding values were obtained using restricted maximum likelihood (REML) method implemented in AIREMLF90 from BLUPF90 family programs (Misztal et al. 2016)

**Single-step GBLUP**

The ssGBLUP model is similar to the PBLUP model (M1) except for the use of a combined genomic and pedigree relationship. The kinship matrix used was \(H\) (Aguilar et al. 2010), in which genotype and pedigree data are combined, the inverse of the matrix \(H\) is:

\[
H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}, \quad (E1)
\]

where, \(A^{-1}\) is the inverse numerator relationship matrix for all animals, \(A_{22}^{-1}\) is the inverse of a pedigree-based relationship matrix for genotyped animals only; and \(G^{-1}\) is the inverse genomic relationship matrix.

The EBV and the GEBV for DD was analyzed as a linear trait using AIREMLF90 and BLUPF90. BS was analyzed using a threshold model (including probit link function to transform event incidence to liability) by means of a Bayesian approach implemented in THRGIBBS1F90 module from BLUPF90 family of programs (Misztal et al. 2016). For Bayesian analysis (THRGIBBS1F90) 200,000 iterations were used in the Gibbs sampling, with a burn-in period of 20,000 iterations and samples were
saved every 50 cycles. Visual inspection of trace plots of the posterior variance components generated by POSTGIBBSF90 were used for QC purposes regarding convergence.

**Bayes C**

The Bayes C fits a mixture model that assumes some known fraction of markers has zero effects and it has been shown that Bayes C is less sensitive to prior assumptions than, e.g. Bayes B (Habier et al. 2011). All model parameters for Bayes C are defined as in M1, except the elements of vector $g$ which was calculated for each fish as:

$$
\sum_{i=1}^{n} g_i a_i \delta_i, \quad (M2)
$$

where $g_i$ is the vector of the genotypes for the $i^{th}$ SNP for each animal; $a_i$ is the random allele substitution effect of the $i^{th}$ SNP; $\delta_i$ is an indicator variable (0,1) sampled from a binomial distribution with parameters determined such that 1% of the markers were included in the model. The prior assumption is that SNP effects have independent and identical mixture distributions, where each marker has a point mass at zero with probability $\pi$ and a univariate normal distribution with probability $1 - \pi$ having a null mean and variance $\sigma^2$, which in turn has a scaled inverse chi-squared prior, with $v_a = 4$ and $v_e = 10$ degrees of freedom and scale parameter $s_a^2$ (or $s_e^2$) (Fernando and Garrick 2013). For the additive variance, the df = 4 was used for the data not to overwhelm the prior if many loci are fitted, considering that for Bayes C, a common locus variance is assumed and estimated by combining information from the prior and the data and each fitted locus contributes to estimation of the common locus variance from the data (Fernando and Garrick 2013). The residual variance degrees of freedom were chosen based values used on previous studies (Peters et al. 2012; Santana et al. 2016; Wolc et al. 2016; Yoshida et al. 2017).
Bayesian Lasso

The Bayesian Lasso (Legarra et al. 2010) appears to be an interesting alternative method for performing regression on markers, suggesting that a double exponential prior may be a better choice than Bayes A method, when most markers do not have effect. The parameters for LASSO method are defined as above in M1, except for a priori distribution of individual SNP effects ($a_i$) which was calculated as:

$$\text{Pr}(a_i | \tau_i^2) \sim \text{N}(1, \tau_i^2) \quad \text{and} \quad \text{Pr}(a_i | \tau_i^2) = \frac{\lambda^2}{2} \exp(-\frac{\lambda^2}{2} | \tau_i^2), \quad (M3)$$

where, $\tau_i^2$ is the individual variance for each SNP, estimated conditionally on a regularization parameter $\lambda$ (initial value was $\lambda^2 = 2/\sigma_g^2$), which was estimated using a priori gamma distribution bounded between 0 and $10^7$.

The Bayes C and Bayesian Lasso analyses were performed using GS3 software (Legarra et al. 2010). A total of 200,000 iterations were used in the Gibbs sampling, with a burn-in period of 20,000 cycles where results were saved every 50 cycles. Convergence and autocorrelation were assessed by visual inspection of trace plots of the posterior variance components.

Genetic parameters and heritability

The total additive genetic variance ($\sigma_g^2$) was estimated using relation matrix A, G and H for PBLUP, GBLUP and ssGBLUP, respectively. For both, days to death and binary survival, the heritabilities were computed using the following equation:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2}, \quad (E2)$$
For Bayesian models, the total additive genetic variance \( (V_A) \) was estimated as the sum of additive marker \( (2\sigma_a^2\pi\sum p_i q_i) \) and the polygenic pedigree \( (\sigma_g^2) \) based additive genetic variance \( (V_A = 2\sigma_a^2\pi\sum p_i q_i + \sigma_g^2) \) and the heritability were computed as:

\[
h^2 = \frac{V_A}{V_A + \sigma_e^2}, \quad (E3)
\]

**Prediction accuracy**

The predictive ability of different models was assessed using a five-fold cross validation scheme. Briefly, all phenotyped and genotyped animals were randomly separated into five validations sets. The genomic predictions of the validation data sets were determined one at a time where the phenotypic records of the validation fish (20% of the population) were set to missing and all remaining individuals with phenotypes and genotypes (80% of the population) were used as training dataset. For ssGBLUP training and validation datasets were separated as described above, with the addition of 100% of the animals with only phenotypes \( (n = 482) \) into the training set.

Accuracy was used to assess the performance of each model for validation set and was estimated as:

\[
r_{GEBV,BV} = \frac{r_{GEBV,y}}{h}, \quad (E3)
\]

where \( r_{GEBV,y} \) is the correlation between the EBV or GEBV of a given model (predicted for the validation set using information from the training set) and the record phenotype, while \( h \) is the square root of the pedigree-based estimate of heritability (Legarra et al. 2008; Ødegård et al. 2014).

In addition, the prediction accuracies obtained using different SNP densities were tested for all the methods. The 0.5K, 3K, 10K and 20K SNPs densities were randomly selected five times for each test method from the ~27K SNP that passed QC.
The bias of EBV prediction obtained as the regression coefficient of phenotyped animal on EBV or GEBV for PBLUP and genomics methods (GBLUP, ssGBLUP, Bayes C and Bayesian Lasso) in the validation data.

**Data availability**

All phenotypic and genotypic data used in the current study can be found at the Figshare public repository (https://figshare.com/s/5219597a19f23873fda3).

**RESULTS**

**Descriptive statistics and genetic parameters**

Summary statistics for both traits and covariate (body weight at the end of the challenge test) are presented in Table 1. The average days to death ranged from 22 to 24 and from 23 to 25 days between tanks for phenotyped (n = 2,320) and genotyped (n = 1,844) animals, respectively. The proportion of cumulative mortality ranged from 0.59 to 0.65 and from 0.52 to 0.60 days between tanks for phenotyped and genotyped animals, respectively. The average body weight at the end of the challenge test was 165.3 g (SD = 40.44 g) and 168.8 g (SD = 41.37 g) for phenotyped and genotyped fish, respectively. A total of 1,934 animals and 27,490 SNP (27K) passed in QC.

Variance components estimates for all the models are presented in Table 2. For both DD and BS the additive genetic variance and heritability were higher for genomic methods compared to PBLUP. For PBLUP the heritability were 0.38 and 0.54 for DD and BS, respectively. For genomic prediction methods the heritability values ranged from 0.45 to 0.57 and from 0.54 to 0.62 for DD and BS, respectively. For both traits, the lowest and the highest heritability estimates when using genomic prediction methods were obtained from GBLUP and Bayes C methods, respectively.
Accuracy of different methods and marker densities

Based on the five-fold cross validation, the prediction accuracy for GEBV from genomic methods outperformed the EBV from PBLUP (Table 3). Within all genomic methods, the accuracy predicted for DD was higher than those for BS with a low standard error of the estimate (Table 3).

The relative increase in accuracy of predicted GEBV compared with EBV from PBLUP varied moderately between models and traits at 27 K marker density (Figure 1). For both traits, the Bayes C method resulted in higher relative improvement in accuracy (> 40%). On the other hand, the LASSO and GBLUP resulted in the lowest relative increase in accuracy and were the same (28%) for DD and similar for BS (LASSO = 36 % and GBLUP = 37%) (Figure 1).

For marker density equal to 20K, the Bayes C and ssGBLUP method was most favorable in terms of relative increase in accuracy for DD and BS, respectively (Figure 1). At marker densities of 3K and 10K, the ssGBLUP and GBLUP resulted the same relative increase in accuracy for BS (Figure 1). In contrast, for DD the Bayesian methods had better perform. The ssGBLUP performed slightly better than the others genomics methods at the lowest marker density (0.5K), especially compared to Bayes C, that showed the lower increase in accuracy for both traits (<11%). Nevertheless, the relative increase in accuracy of predicted GEBV from all genomic models were superior to EBV from PBLUP even at the lowest marker density of 0.5 K for both traits. In general, the relative increase in accuracy was considerably more evident for BS than DD.

The GEBV estimated using GBLUP had the smallest departure from the unity for DD. In contrast, the Bayes C method resulted in the most biased estimate (1.035). The bias values for EBV and GEBVs for BS were considerably lower than 1.0 for all
methods and ranged from 0.24 to 0.27, which indicates that all results for BS are upward biased (Table 3).

**DISCUSSION**

**Heritability for pedigree based and genomic models**

Low to moderate heritability estimates (from 0.16 to 0.24) were reported for SRS resistance in Atlantic salmon (Yáñez *et al.* 2013 and Yáñez *et al.* 2014) and coho salmon (Yáñez *et al.* 2016) using a pedigree-based method to analyze the trait defined similarly to DD and BS. The comparatively higher estimates of heritability reported using genomic information compared to PBLUP in our study is in accordance with what has been reported in other fish species (Tsai *et al.* 2016; Vallejo *et al.* 2017; Bangera *et al.* 2017; Correa *et al.* 2017). Vallejo *et al.* (2016, 2017) also estimated similar range of heritability using genomic models (0.26 - 0.54) and PBLUP (0.31 - 0.48) for bacterial cold water disease resistance in rainbow trout.

**Prediction accuracy**

The relatively high accuracy were achieved in present study for genomic methods, we suggested that the high relationship between the animals in training and validation dataset, and small effective population size of this breeding population could be contributed for the accuracies values. This in turn can result in extensive linkage disequilibrium (LD) and a smaller number of effective chromosome segments to be estimated. The GEBV prediction accuracy for resistance against cold water disease in rainbow trout was estimated using different methods by Vallejo *et al.* (2017) and accuracies reported were similar of magnitude for survival days (0.63 to 0.71) and survival status (0.66 to 0.71).
In Atlantic salmon, Bangera et al. (2017) and Correa et al. (2017) showed that the relative increase in GEBV prediction accuracies from different models compared to PBLUP was higher by up to 30% and 22% for resistance against SRS and *Caligus rogercresseyi*, respectively. However, improvement in accuracy values in the current study varied from 28 to 41%; which was still lower than values reported by Vallejo et al. (2017), which ranged from 83 to 109% for bacterial cold water disease resistance in rainbow trout. We speculate that in the study of Vallejo et al. (2017), the use of a large number of animals with phenotype in the training dataset (7,893 vs 2,417) resulted in a higher relative increase in accuracy. Furthermore, Piyasatian et al. (2007) suggested that high heritability of the trait (> 0.45 in the present study) reduced the benefit of the genomic selection over PBLUP.

**Effect of marker density on accuracy**

Genotyping of large number of selection candidates with high-density panels may not be cost-effective if the economic benefit per animal is low compared to the cost of genotyping (Habier et al. 2009) as in aquaculture species. The use of low-density panels, with considerable reduction in cost of genotyping, is a potential cost-effective approach to implement genomic selection. Previous studies in Atlantic salmon reported that low-density panels between 5 to 10K were sufficient to obtain reliable increases in accuracy (even close to maximal accuracy of high density panel) compared to PBLUP (Tsai et al. 2015; Tsai, et al. 2016; Correa et al. 2017). The lowest density SNP panel (0.5K) used in our study resulted in lowest accuracies, mainly for DD, as a result of insufficient LD between the markers due the large distance between the randomly selected low-density markers (Bangera et al. 2017).
We suggest that the considerably high gain in GEBV accuracy obtained in different genomic prediction methods using markers above 10 K, was because of the high LD between the randomly selected markers. All low-density panels showed improved GEBV accuracy over PBLUP (Figure 1), higher accuracy of genomic prediction can be obtained by using high-density panels as also shown by Ødegård et al. (2014) and Bangera et al. (2017). Therefore, to implement cost-effective genomic selection, a combination of genotyping the selection candidates with a low-density panel (e.g. 500 SNPs) and then imputation to high-density (e.g. 50K) could be used (Tsai et al. 2017). The strategy of imputing from 0.25K and 0.5K to a high-density panel, and then use of imputed genotypes for genomic prediction was shown to achieve a similar level of accuracy compared to using true genotypes in Atlantic salmon (Tsai et al. 2017).

**Comparison of models at different marker density**

The GBLUP approach assumes a polygenic control of the trait and makes use of all genotyped SNPs for calculating the genomic relationship matrix. In contrast, Bayesian models assume that few markers explain the genetic variance of a trait (Habier et al. 2007; Hayes et al. 2009; de los Campos et al. 2013). Thus, Bayesian methods are expected to perform better than GBLUP when several moderate to large effect QTL are controlling the trait. In this study, two GBLUP and two Bayesian methods were tested to compare the accuracy of genomic predictions from different GS models to those obtained with ordinary pedigree-based BLUP.

All genomic prediction methods outperformed the PBLUP at different SNP densities (Figure 1). For both traits the Bayes C method had the highest accuracy (> 40% relative increase over PBLUP) at the highest SNP density (27K). The GBLUP and
ssGBLUP method, had a constant relative increase in accuracy from 3K to 27K SNP panels mainly for BS. Interestingly, for the 0.5K SNP panel, the ssGBLUP resulted in the highest accuracies for both traits, suggesting that for very low density panels, the use of additional animals with only phenotypes in the training set can improve the accuracy of predictions. Furthermore, ssGBLUP could be used as strategy to reduce the genotyping costs and still achieve higher GEBV accuracies compared to PBLUP. As it has been reported previously, the use of information from genotyped and non-genotyped individuals (Lourenco et al. 2014) and the increase in accuracy when compared to PBLUP (Chen et al. 2011; Christensen et al. 2012) are some of the advantages in use ssGBLUP.

The use of progressively more markers in the GBLUP method might have resulted in better capturing of genetic relationships, whereas, Bayes C was more effective in capturing LD between markers and QTL when more markers were used (Bermingham et al. 2015). Furthermore, fitting 1% of the SNPs with larger effect in Bayes C method resulted in the highest relative increase of accuracy. This is most likely due to the genetic architecture of *P. salmonis* resistance in rainbow trout. In a previous genome-wide association study in the same population, *P. salmonis* resistance was suggested to be under oligogenic control (data not published), with few SNPs showing moderate to large effect (top 10 SNPs explained more than 50% of the genetic variance, results not show).

Bayes C outperformed ssGBLUP at 20 and 27K SNP density for DD and was slightly lower for BS. However, for lower SNP densities, the Bayes C method had lower accuracy (Table 3 and Figure 1). The large distance between the low-density SNPs results in lower LD between the markers and QTL. The possibility of exclusion of the
SNPs with moderate to high effect during the process of random selection might have resulted in lower relative accuracies in Bayes C.

Several other studies also reported that GEBVs estimated by Bayesian methods outperformed EBVs estimated using pedigree-based methods, and even other genomic methods (i.e. GBLUP and ssGBLUP) (Neves et al. 2014; Vallejo et al. 2016, 2017; Bangera et al. 2017; Correa et al. 2017). A disadvantage in using Bayesian methodologies (e.g. Bayes C) is a considerably higher computational time, which could increase linearly depending on the number of markers fitted in the model (Bermingham et al. 2015). Considering the similarity in accuracies between Bayes C and ssGBLUP, and the highest accuracies for the low-density panels (Table 3 and Figure 1), the ssGBLUP method may be a more flexible and computationally efficient alternative.

**Bias**

GEBV bias was calculated as the regression of EBV on GEBV. A regression coefficient equal to one is indicative of predictions that are on a scale similar to that of the GEBV, a regression less than one or greater than one indicated GEBV is overestimated or underestimated, respectively. Here we found bias values somewhat below the unity for BS, indicating that GEBVs were under regressed than EBV, suggested that the genetics trends could be underestimated, and having negative impact in selection schemes. Other studies reported bias values less than one for BS and similar bias values compared to present work for DD (Vallejo et al. 2016, 2017; Bangera et al. 2017).

**Implications**

Our results showed that using the genomic information for estimating breeding values achieved higher accuracies compared to using only pedigree information for both days
to death and binary survival. Using 20K and 3K SNP panels for DD and BS, respectively, was enough to improve accuracy to similar values obtained for 27K using PBLUP. Given the economic importance of resistance against *P. salmonis* to rainbow trout, and the efficacy of genomic prediction over pedigree-based methods, we suggest that selective breeding using genomic information will be an important component to control SRS and reduce losses in aquaculture systems.

**Acknowledgements**

Aguas Claras S.A. provided funding for the experimental challenge test and fish used in this study. This work has been partially funded by a grant from CORFO (11IEI-12843) and FONDECYT REGULAR No 1171720 and Núcleo Milenio INVASAL. GMY acknowledge Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP process numbers 2014/20626-4 and 2015/25232-7) for doctoral fellowship. RC acknowledge for CNPq fellowship (process number 308636/2014-7).

**Conflicts of interest**

The authors declare that they have no conflict of interest.

**References**


Correa, K., R. Bangera, R. Figueroa, J. P. Lhorente, and J. M. Yáñez, 2017 The use of


Flores-Mara, R., F. H. Rodríguez, R. Bangera, J. P. Lhorente, R. Neira *et al.*, 2017 Resistance against infectious pancreatic necrosis exhibits significant genetic variation and is not genetically correlated with harvest weight in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 479: 155–160.


Houston, R. D., C. S. Haley, A. Hamilton, D. R. Guy, A. E. Tinch et al., 2008 Major
quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). Genetics 178: 1109–15.


Misztal, I., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar et al., 2016 Manual for BLUPF90 family of programs.


Peña, B., A. Isla, D. Haussmann, and J. Figueroa, 2016 Immunostimulatory effect of
association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. BMC Genomics 1–9.


Table 1. Summary statistics for resistance against Piscirickettsia salmonis for phenotyped and genotyped rainbow trout.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used as covariable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Estimates of residual variance ($\sigma^2_e$), total additive genetic variance ($V'_a$) and heritability ($h^2$) for resistance against Piscirickettsia salmonis in rainbow trout.
Total additive genetic variance; for PBLUP, ssGBLUP and GBLUP was $\sigma_g^2$; LASSO and BAYESC was $2\sigma_a^2 + \sum p_i q_i + \sigma_p^2$ ($\sigma_p^2 = \text{polygenic effect}$).

Standard error or standard deviation for Bayesians methods.

Table 3. Mean accuracy, bias and standard errors (SE) of EBV and GEBV for resistance against *Piscirickettsia salmonis* using a 27 K SNP panel.

Regression for the EBV obtained by PBLUP and GEBV predicted with the different genomic methods.

Figure 1. Relative increase in accuracy of different genomic selection methods for trait days to death and binary survival compared with PBLUP in rainbow trout using different SNP chip densities (0.5, 3, 10, 20 and 27 K).
<table>
<thead>
<tr>
<th>Traits</th>
<th>Tank</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Mean</th>
<th>SD&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Min&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Max&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phenotyped fish</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days to death (days)</td>
<td>1</td>
<td>819</td>
<td>23.59</td>
<td>8.07</td>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>805</td>
<td>22.82</td>
<td>8.03</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>792</td>
<td>22.13</td>
<td>8.27</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>Binary survival (1 or 0)</td>
<td>1</td>
<td>819</td>
<td>0.59</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>805</td>
<td>0.65</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>792</td>
<td>0.65</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Final challenge weight (g)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>-</td>
<td>2,320</td>
<td>165.30</td>
<td>40.44</td>
<td>46</td>
<td>295</td>
</tr>
<tr>
<td><strong>Genotyped fish</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days to death (days)</td>
<td>1</td>
<td>669</td>
<td>24.92</td>
<td>7.64</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>641</td>
<td>24.01</td>
<td>7.78</td>
<td>11</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>624</td>
<td>23.25</td>
<td>8.07</td>
<td>11</td>
<td>32</td>
</tr>
<tr>
<td>Binary survival (1 or 0)</td>
<td>1</td>
<td>669</td>
<td>0.52</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>641</td>
<td>0.59</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>624</td>
<td>0.60</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Final challenge weight (g)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>-</td>
<td>1,844</td>
<td>168.80</td>
<td>41.37</td>
<td>66</td>
<td>295</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>Days to death</th>
<th>Binary survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V'_a$&lt;sup&gt;a&lt;/sup&gt;</td>
<td>$\sigma_e^2$&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>PBLUP</td>
<td>23.017</td>
<td>37.375</td>
</tr>
<tr>
<td>ssGBLUP</td>
<td>34.585</td>
<td>34.376</td>
</tr>
<tr>
<td>LASSO</td>
<td>29.031</td>
<td>32.840</td>
</tr>
<tr>
<td>GBLUP</td>
<td>27.313</td>
<td>33.813</td>
</tr>
<tr>
<td>BAYESC</td>
<td>41.580</td>
<td>31.030</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>Days to death</th>
<th>Binary survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>SE</td>
</tr>
<tr>
<td>PBLUP</td>
<td>0.613</td>
<td>0.097</td>
</tr>
<tr>
<td>ssGBLUP</td>
<td>0.724</td>
<td>0.034</td>
</tr>
<tr>
<td>Method</td>
<td>0.784</td>
<td>0.069</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>LASSO</td>
<td>0.785</td>
<td>0.064</td>
</tr>
<tr>
<td>GBLUP</td>
<td>0.859</td>
<td>0.061</td>
</tr>
</tbody>
</table>
Days to death

<table>
<thead>
<tr>
<th>SNP densities</th>
<th>LASSO</th>
<th>GBLUP</th>
<th>ssGBLUP</th>
<th>BAYESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5K</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>3K</td>
<td>27</td>
<td>26</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>10K</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>20K</td>
<td>28</td>
<td>28</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>27K</td>
<td>40</td>
<td>39</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

Binary survival

<table>
<thead>
<tr>
<th>SNP densities</th>
<th>LASSO</th>
<th>GBLUP</th>
<th>ssGBLUP</th>
<th>BAYESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5K</td>
<td>20</td>
<td>21</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>3K</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>10K</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>20K</td>
<td>33</td>
<td>38</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>27K</td>
<td>36</td>
<td>37</td>
<td>40</td>
<td>41</td>
</tr>
</tbody>
</table>