GSE Genetics Selection Evolution

RESEARCH ARTICLE

CrossMark

Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture

Roger L. Vallejo[1*], Timothy D. Leeds[1], Guangtu Gao[1], James E. Parsons[2], Kyle E. Martin[2], Jason P. Evenhuis[1], Breno O. Fragomeni[3], Gregory D. Wiens[1] and Yniv Palti[1]

Abstract

Background: Previously, we have shown that bacterial cold water disease (BCWD) resistance in rainbow trout can be improved using traditional family-based selection, but progress has been limited to exploiting only between-family genetic variation. Genomic selection (GS) is a new alternative that enables exploitation of within-family genetic variation.

Methods: We compared three GS models [single-step genomic best linear unbiased prediction (ssGBLUP), weighted ssGBLUP (wssGBLUP), and BayesB] to predict genomic-enabled breeding values (GEBV) for BCWD resistance in a commercial rainbow trout population, and compared the accuracy of GEBV to traditional estimates of breeding values (EBV) from a pedigree-based BLUP (P-BLUP) model. We also assessed the impact of sampling design on the accuracy of GEBV predictions. For these comparisons, we used BCWD survival phenotypes recorded on 7893 fish from 102 families, of which 1473 fish from 50 families had genotypes [57 K single nucleotide polymorphism (SNP) array]. Naïve siblings of the training fish ($n = 930$ testing fish) were genotyped to predict their GEBV and mated to produce 138 progeny testing families. In the following generation, 9968 progeny were phenotyped to empirically assess the accuracy of GEBV predictions made on their non-phenotyped parents.

Results: The accuracy of GEBV from all tested GS models were substantially higher than the P-BLUP model EBV. The highest increase in accuracy relative to the P-BLUP model was achieved with BayesB (97.2 to 108.8%), followed by wssGBLUP at iteration 2 (94.4 to 97.1%) and 3 (88.9 to 91.2%) and ssGBLUP (83.3 to 85.3%). Reducing the training sample size to $n = {\sim}1000$ had no negative impact on the accuracy (0.67 to 0.72), but with $n = {\sim}500$ the accuracy dropped to 0.53 to 0.61 if the training and testing fish were full-sibs, and even substantially lower, to 0.22 to 0.25, when they were not full-sibs.

Conclusions: Using progeny performance data, we showed that the accuracy of genomic predictions is substantially higher than estimates obtained from the traditional pedigree-based BLUP model for BCWD resistance. Overall, we found that using a much smaller training sample size compared to similar studies in livestock, GS can substantially improve the selection accuracy and genetic gains for this trait in a commercial rainbow trout breeding population.

*Correspondence: roger.vallejo@ars.usda.gov
[1] National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV, USA
Full list of author information is available at the end of the article

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 2 of 13

Background

Bacterial cold water disease (BCWD) causes significant mortality and economic losses in salmonid aquaculture [1, 2]. The etiological agent of BCWD is a gram-negative bacterium, *Flavobacterium psychrophilum* (*Fp*) and current methods for control of BCWD are limited. At the National Center for Cool and Cold Water Aquaculture (NCCWA), we have pursued a selective breeding program to increase genetic resistance of rainbow trout to BCWD and have shown that BCWD resistance is a moderately heritable trait that responds to selection [3]. Furthermore, we have revealed a complex genetic architecture of BCWD resistance [4] and identified several major quantitative trait loci (QTL) for this trait in the NCCWA odd- and even-year rainbow trout selective-breeding populations [5–8]. Although those QTL can be evaluated for marker-assisted selection (MAS) in this population, following fine-mapping to identify tightly linked markers to the BCWD resistance QTL, the complex genetic architecture of BCWD resistance and the high genetic variability that we detected in past studies [3, 5, 6, 8] suggest that a genomic selection (GS) approach will likely be more effective than MAS for improving BCWD resistance in rainbow trout.

Genomic selection is a methodology [9] that is revolutionizing animal and plant breeding. This methodology uses dense marker genotypes that cover the genome, combined with phenotypic data to predict breeding values of all genotyped individuals. In GS, a reference population is genotyped and recorded for the trait to train the GS model and estimate the effects of each single nucleotide polymorphism (SNP). Selection candidates are also genotyped, and by combining their genotypes with the estimated SNP effects, genomic-enabled breeding value (GEBV) are estimated for the selection candidates. The GS approach does not necessarily require pedigree recording and the selection candidates do not need phenotypes. Thus, the GS methodology is particularly relevant for traits that cannot be measured directly on selection candidates, including carcass traits, sex-limited traits, and disease resistance, and has been demonstrated to be very effective in commercial dairy cattle [10–13]. For aquaculture species, the main advantage of GS is that it enables exploitation of within-family genetic variation for traits that cannot be measured directly on selection candidates. In addition to increasing accuracy of selection, GS is expected to reduce rates of inbreeding because the increased accuracy of Mendelian sampling terms in GS allows for identification and selection of elite breeding candidates from more families, with lower co-selection of sibs [14, 15].

The genomic best linear unbiased prediction BLUP (GBLUP) method assumes that the trait has a polygenic architecture and considers the contribution of all genotyped markers in construction of the genomic relationship matrix (G). In contrast, Bayesian variable selection models assume that the genetic variance of a trait is explained by a reduced number of markers [16–19]. Based on this assumption, GBLUP is not expected to perform as well as Bayesian variable selection models when the trait is controlled by several QTL with moderate-to-large effects. The GBLUP method has been recently extended to the single-step GBLUP method, which allows the incorporation of both pedigree- and genomic-derived relationships into a single relationship matrix [20, 21], and to the weighted single-step GBLUP method, which emulates the Bayesian variable selection models by fitting in the model only those SNPs that explain a fraction of the trait genetic variance [22].

The genetic architecture of the trait and the population structure may have a significant impact on the accuracy of genomic predictions. Therefore, when evaluating a trait for the first time in a population, it is important to compare the accuracy of GEBV predictions from several GS models to those obtained with pedigree-based BLUP.

In a recent post hoc study [23] that was conducted on a research population maintained at the NCCWA, we did not find improved accuracy using GS models compared with the pedigree-based BLUP model for predicting genetic merit of BCWD resistance in an experimental rainbow trout breeding population. However, the training sample size and the number of fish and families used for testing in that study were insufficient and the representation of families in the testing sample was imbalanced. Thus, the current study was conducted to assess the feasibility of GS for improving BCWD resistance in the rainbow trout aquaculture industry and compare its accuracy with traditional family-based selective breeding using a larger sample size and a more balanced mating design from a larger number of full-sib families. Another major difference with the pilot study [23] is that it was conducted using historic archived samples, while in the current study the mating design for progeny testing was based on the GEBV of the potential breeders, which provided a much more accurate assessment of the potential impact of GS on commercial breeding in rainbow trout aquaculture, as well as the feasibility of its real-time implementation into current commercial breeding schemes. Thus, the objectives of this study were to (1) predict GEBV for BCWD resistance in a commercial breeding population that has been selected primarily on growth; (2) compare the accuracy of pedigree-based EBV with that of GEBV from three GS models using actual progeny performance data; and (3) assess the impact of the study design on the accuracy of genomic predictions using different sampling schemes.

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 3 of 13

Methods

Fish rearing and disease challenge

All fish work was conducted in accordance with national and international guidelines. The protocol for this study was approved by the Institutional Animal Care and Use Committee (IACUC; Protocol # 053) of the US Department of Agriculture, Agricultural Research Service, the National Center for Cool and Cold Water Aquaculture. All efforts were made to ensure fish welfare and to minimize suffering.

Details of the fish rearing conditions and the 21-day survival trial following intraperitoneal injection with *F. psychrophilum* (*Fp*), the causative agent of BCWD, have been reported elsewhere [3, 24]. Mortalities were removed and recorded daily and fin clipped. Fish that survived to day 21 post-infection were euthanized in 200 mg L^{-1} of tricaine methanesulfonate, MS 222 (Sigma) for at least 10 min prior to sampling of fin clips. Fin clips from all mortalities and survivors were individually kept in 95% ethanol until DNA was extracted using established protocols [25].
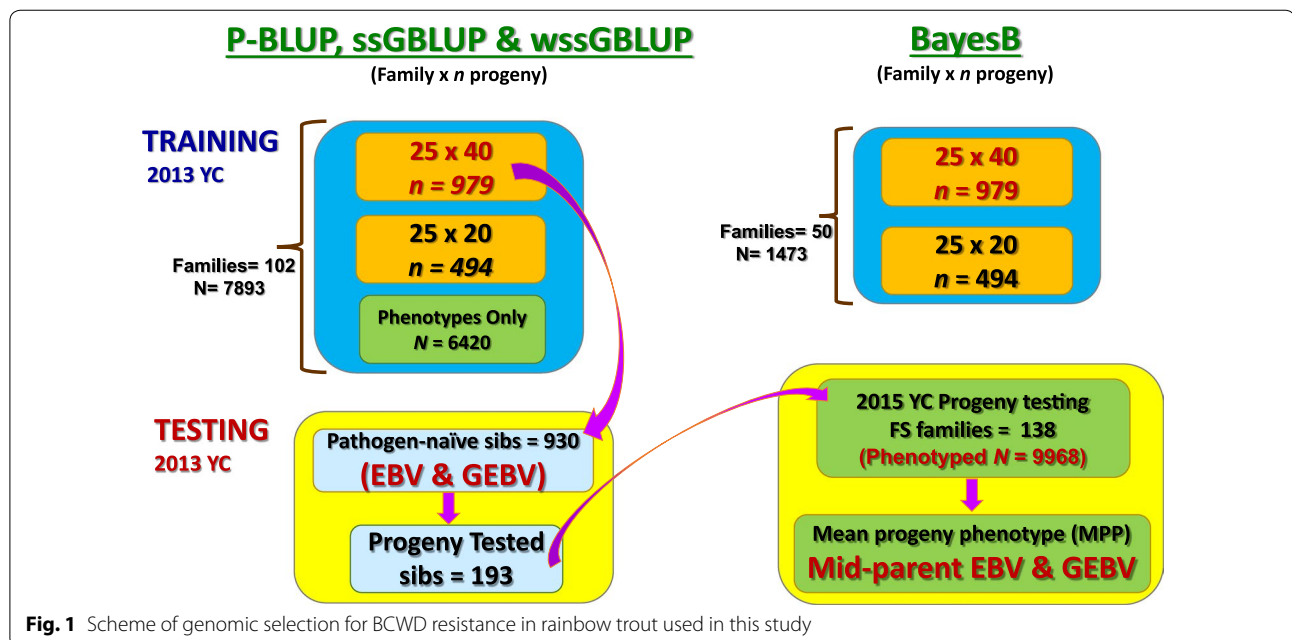
Training and testing data

The training sample included 102 pedigreed full-sib (FS) families from year-class (YC) 2013 of the Troutlodge, Inc., all-female, May-spawning population (Fig. 1). The 102 YC 2013 families represented a nucleus breeding population undergoing selection for growth, and thus had not previously been selected for BCWD resistance. The fish from YC 2013 families were evaluated in the laboratory BCWD challenge in two tanks per family, with an initial stocking of 40 fish per tank (total phenotyped fish $n = 7893$). The original study design was to sample $n = 1500$ fish with phenotypes and genotypes from 50 FS families. Of the 50 FS families, 25 were full-sibs of the testing sample and 25 were least related to the testing sample families based on pedigree records. We sampled ~40 fish from the 25 FS families that were closely related to the testing sample and ~20 fish per family from the other less related 25 families. In practice, we sampled $n = 1473$ fish with phenotypes and genotypes from those 50 families ($n = 17$ to 40 per family). Thus from the 7893 BCWD evaluated fish, 1473 fish had genotype data.

The testing sample included 930 potential breeders or selection candidates (sires and dams) that were disease naïve fish sampled from 25 families ($n = 31$ to 44 testing fish per family). The testing fish had family-based EBV for survival days (DAYS) and survival status (STATUS) that were estimated with a pedigree-based BLUP model (described below) using BCWD survival records measured on their siblings and any collateral relatives among the 102 FS families ($n = 7893$). Each of these testing fish also had predicted GEBV from GS models (also described below).

To assess the accuracy of the GEBV, we generated 138 next-generation YC 2015 FS progeny testing families (PTF) from crosses that involved 193 of the YC 2013 testing fish (Fig. 1). These 138 YC 2015 PTF were phenotyped in 2015 for BCWD survival ($n = 9968$) to calculate the mean progeny phenotype (MPP) for each PTF.

Fig. 1 Scheme of genomic selection for BCWD resistance in rainbow trout used in this study

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 4 of 13

BCWD resistance phenotypes

Survival DAYS, the number of days to death post-challenge, were recorded for a total of 21 days post-challenge, with survivors being assigned a value of 21 days post-challenge. Each fish also had a binary survival STATUS record. The binary STATUS had two classes: 2 for fish that were alive on day 21 post-challenge and 1 for fish that died during the 21 days post challenge evaluation period. In the GS analysis, we used DAYS and STATUS records from training sample fish to estimate marker effects to then predict GEBV for DAYS and STATUS for each of the testing sample fish.

SNP genotyping platform

Genotyping was performed by a commercial genotyping service provider (Neogen, Inc., Lincoln, NE) using the Rainbow Trout Axiom® 57 K SNP array, as previously described in [26]. Our quality control (QC) bioinformatics pipeline filtered out SNPs with significant distortion from the expected Mendelian segregation in each FS family (Bonferroni adjusted to $P < 0.10$) and also removed two training fish that did not have genotypes that matched the parents based on the pedigree records. After genotype data QC, a total of 41,868 SNPs were included in the genotyping dataset.

Before training the GS models, all genotyped SNPs were further filtered using QC algorithms that are implemented in the computer program BLUPF90 [27]. The QC retained SNPs with a genotype calling rate higher than 0.90, minor allele frequency higher than 0.05, and departures from Hardy–Weinberg equilibrium less than 0.15 (difference between expected and observed frequency of heterozygotes). Parent-progeny pairs were tested for discrepant homozygous SNPs, those SNPS with a conflict rate of more than 1% were discarded. After this final QC step, 35,636 SNPs remained for the GS analysis.

Estimation of pedigree-based EBV

For the testing fish, we estimated EBV for BCWD resistance phenotypes using a pedigree-based BLUP (P-BLUP) model. Family-based EBV were estimated using BCWD survival records measured on siblings of the testing fish and any collateral relatives. The phenotypic dataset included records from $n = 7893$ fish from 102 FS families (39 paternal half-sib families, no maternal half-sib families, and 24 families not nested within a half-sib family). The pedigree dataset included 32,279 fish from seven generations.

Based on past genetic analyses for estimating EBV for BCWD resistance in rainbow trout [3, 23], we decided to use an animal model that included a population mean, random animal genetic and random residual effects. The records of the BCWD survival phenotypes DAYS

and STATUS were fit into P-BLUP linear and threshold models, respectively, using the computer application BLUPF90 [27]. Family was not included in the model because we have often found that genetic variance is downward-biased when the family effect is included in the animal model [23]. The challenge tank effect was also not included in the model because the fish were too small for individual tagging at the time of disease challenge and hence the fish were challenged and reared in individual family tanks, which confounded tank with family effects. Likewise, body weight was not included in the model because, for this high-throughput disease challenge study using non-tagged fish, pre-challenge body weight data are only available as an average body weight for each challenge tank (i.e., bulk weight of fish divided by number of fish) and are, therefore, confounded with family.

Estimates of the heritability for the binary trait STATUS obtained with the pedigree-based BLUP (and GS models below) on the underlying scale of liability were transformed to the observed scale of survival STATUS using this expression:

$$h^2_{observed} = \left(h^2_{liability} i^2 p\right)/(1 - p);$$

where i is the mean deviation of affected individuals from their group mean, and p is the incidence of mortality [28].

Estimation of GEBV with Bayesian variable selection models

The SNP genotype data from the training fish (YC 2013 families), with their corresponding BCWD phenotypic records, were used to train the prediction model and estimate marker effects using the Bayesian variable selection model BayesB implemented in the software GENSEL [29] as previously described in [23]; an animal model was used that included a population mean, random marker and random error effects. The mixture parameter π was assumed to be known and defined to meet the condition $k \le n$; where n is the number of training fish. After testing $\pi = 0.96, 0.97,$ and 0.98 by performing fivefold cross-validation analyses (results not presented), we decided to use $\pi = 0.97$ in the final GS analysis with BayesB because it yielded the best accuracy predictions.

The software GENSEL uses a Gibbs sampling approach in the BayesB analysis [30]. In this study, DAYS and STATUS were analyzed using 210,000 Markov chain Monte Carlo (MCMC) iterations, of which the first 10,000 samples were discarded as burn-in. From the remaining 200,000 samples, we saved one from every 40 samples, thus a total of 5000 samples were used in the analysis. We assessed the proper mixing and convergence of the MCMC iterations using the R package CODA [31] to

Vallejo *et al. Genet Sel Evol (2017) 49:17*

Page 5 of 13

ensure that the MCMC samples were drawn from the full posterior distributions.

Estimation of GEBV with single-step GBLUP models

The SNP genotype data from training fish and pedigree information on all fish included in this GS study were used to estimate GEBV for the testing sample fish ($n = 930$ full-sibs of training fish that were not disease challenged) using two methods: (1) single-step genomic BLUP (ssGBLUP) [20, 32]; and (2) weighted ssGBLUP (wssGBLUP), as previously described [23]. In wssGBLUP, the weights for each SNP are 1s for the first iteration, which means that all SNPs have the same weight (i.e., standard ssGBLUP). For the next iterations (2nd, 3rd, etc.), the weights are individual SNP variances that are calculated using both the SNP effects estimated in the previous iteration and their corresponding allele frequencies [22]. In contrast to the BayesB model, the ssGLUP and wssGBLUP models included also fish from YC 2013 families in the analysis, which had only BCWD resistance phenotype records ($n = 6420$) without marker genotype data: full-sibs of training fish (50 FS families) and 52 additional FS families from the same breeding population as the training and testing fish (Fig. 1). The linear and threshold models to estimate GEBV for DAYS and STATUS, respectively, included a population mean, random animal genetic effects, and random error effects and were fitted as previously described in [23] using the software BLUPF90 [27].

Before performing the GS analysis with ssGBLUP and wssGBLUP, we estimated genetic parameters to use as priors in the Bayesian analysis of the binary trait STATUS as previously described in [23]. The MCMC Gibbs sampling scheme included a total of 210,000 iterations; the first 10,000 iterations were discarded as burn-in iterations. Then, from the remaining 200,000 samples, one from every 40 samples was saved for analysis. This Gibbs sampling scheme collected 5000 independent samples for analysis. The proper mixing and convergence of these MCMC iterations were also assessed using the R package CODA [31].

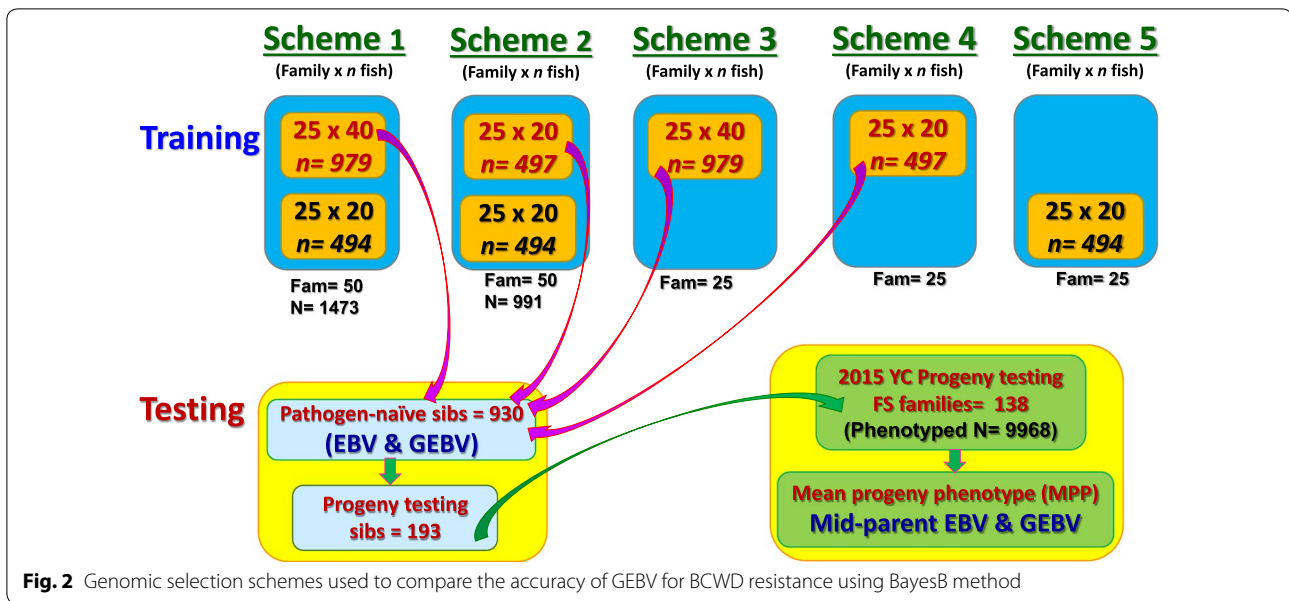### Predictive ability and bias of EBV and GEBV

The predictive ability (PA) of EBV and GEBV, which are both estimates of additive genetic effects, was estimated under the assumption that the correlation of mid-parent EBV or GEBV with the mean progeny performance (MPP) for each PTF is an estimate of the accuracy of the estimated breeding values [23, 33, 34]. We used the mid-parent EBV or GEBV instead of the individual EBV or GEBV of each parent, because the testing fish were mated to each other to generate the 138 PTF, rather than mating each testing fish to a large random sample of fish

from a common genetic background, as is often done in GS studies with terrestrial agricultural animals and birds.

Bias of the EBV was estimated as the regression coefficient of MPP on predicted mid-parent EBV ($\beta_{MPP.EBV}$). Similarly, bias of the GEBV was estimated as the regression coefficient of MPP on predicted mid-parent GEBV ($\beta_{MPP.GEBV}$). A value of 1.0 for the regression of true breeding value, performance phenotype or MPP on predicted EBV or GEBV is theoretically expected for unbiased estimates of BV; and a deviation from 1.0 can be interpreted as prediction bias [35]. Before estimating the regression coefficients, the predicted EBV and GEBV for STATUS, which were estimated on the underlying scale of liability, were transformed to the observed scale. Categorical data analysis performed with the software programs BLUPF90 and GENSEL uses a probit link function; therefore, the EBV and GEBV were transformed to the standard normal cumulative distribution function (CDF) to estimate the probability of survival [36, 37].

Impact of GS study design on accuracy of GEBV

To evaluate the impact of sample size and relatedness between the training and testing fish on the accuracy of GEBV predictions, we used five GS schemes that were developed using the genotype and phenotype records collected in this study, as outlined in Fig. 2 and Table 3. The following study design variables were evaluated: size of the training data (~500, ~1000 or ~1500 fish); number of training families (25 or 50 families); size of the training families (20 or 40 fish per family); and proportion of fish in the training data that were full-sibs (FS) of the testing fish (Table 3). For the latter variable, scheme 1 = 0.66 means that 66% of the fish in the training data were FS of fish in the testing data; scheme 2 = 0.50 means that 50% of the fish in the training data were FS of fish in the testing data; schemes 3–4 = 1.0 means that all fish in the training data were FS of fish in the testing data; and scheme 5 = 0.0 means that none of the fish in the training data were FS of fish in the testing data (i.e., fish in the training and testing data were sampled from different families from the same breeding population). In scheme 1, there were two distinct groups of training families: (1) a set of 25 families with ~40 progeny each ($n = 979$) that also contributed fish to the testing data; and (2) a set of 25 families with ~20 progeny each ($n = 494$) that did not contribute fish to the testing data (Fig. 2). In scheme 2, we used both groups again, but reduced the number of fish sampled per family in group (1) to ~20 ($n = 497$). In scheme 3, we only sampled group (1) ($n = 979$). In scheme 4, we only sampled group (1) again, but reduced the number of fish sampled per family to ~20 ($n = 497$). In scheme 5, we only sampled group (2).

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 6 of 13

Fig. 2 Genomic selection schemes used to compare the accuracy of GEBV for BCWD resistance using BayesB method

Table 1 Accuracy of genomic prediction for BCWD survival DAYS in rainbow trout

Model[a]	Training sample				Testing sample		
	Phenotyped fish	Genotyped fish	Effective SNPs	h^{2b}	Genotyped fish	Predictive ability[c]	Bias[d]
P-BLUP	7893	0	0	0.37	0	0.34	0.86
ssGBLUP	7893	1473	35,636	0.33	930	0.63	0.99
wssGBLUP2	7893	1473	35,623	0.33	930	0.67	0.71
wssGBLUP3	7893	1473	35,623	0.33	930	0.65	0.65
BayesB	1473	1473	35,636	0.23	930	0.71	1.16

[a] The estimated breeding values (EBV) were estimated with a pedigree-based animal model (P-BLUP); and the genomic EBV (GEBV) were estimated with three genomic selection (GS) models: single-step GBLUP (ssGBLUP), weighted ssGBLUP (wssGBLUP) and Bayesian method BayesB. The wssGBLUP2 and wssGBLUP3 corresponds to iteration 2 and 3, respectively

[b] For the GS models, h^2 is the proportion of phenotypic variance explained by the markers. For the P-BLUP model, h^2 is the trait narrow-sense heritability estimated from pedigree and phenotypic records

[c] The predictive ability of EBV (PA_{EBV}) or GEBV (PA_{GEBV}) was defined as the correlation of mid-parent EBV or GEBV with MPP from each PTF:
$PA_{EBV} = CORR(MPP, Midparent\ EBV); PA_{GEBV} = CORR(MPP, Midparent\ GEBV)$

[d] The bias of EBV ($Bias_{EBV}$) or GEBV ($Bias_{GEBV}$) was defined as the regression coefficient of performance MPP on predicted mid-parent EBV or GEBV:
$Bias_{EBV} = REGRES(MPP, Midparent\ EBV); \ Bias_{GEBV} = REGRES(MPP, Midparent\ GEBV)$

We used only GEBV that were estimated with the Bayesian variable selection model BayesB for this evaluation of the impact of the GS study design on the accuracy of predictions because it resulted in the highest accuracy of GEBV in scheme 1, which is the scheme with the largest training sample size. The BayesB model was run using three mixture parameters ($\pi = 0.97, 0.98,$ and 0.987), which were chosen accordingly based on the training data size of the tested GS scheme (Table 3).

Results

Mean progeny phenotype and mid-parent EBV or GEBV

For BCWD resistance phenotypes DAYS and STATUS, the mean progeny phenotype (MPP) and the mid-parent EBV and GEBV estimated for each of the 138 progeny testing families (PTF) are in Additional file 1: Table S1.

Heritability of BCWD resistance

Estimates of the narrow-sense heritability for DAYS and STATUS were equal to 0.37 and 0.35, respectively, using the BLUP model without genomic data (Tables 1, 2). Similarly, the proportion of phenotypic variance explained by the markers for DAYS and STATUS ranged from 0.23 to 0.33 and 0.25 to 0.35, respectively, using the GS models.

Accuracy and bias of EBV

The prediction accuracy (PA) of EBV for DAYS ($PA_{EBV} = 0.34$) was marginally lower than the PA of EBV

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 7 of 13

Table 2 Accuracy of genomic prediction for BCWD survival STATUS in rainbow trout

Model[a]	Training sample				Testing sample		
	Phenotyped fish	Genotyped fish	Effective SNPs	h^{2b}	Genotyped fish	Predictive ability[c]	Bias[d]
P-BLUP	7893	0	0	0.35	0	0.36	0.67
ssGBLUP	7893	1473	35,636	0.35	930	0.66	0.86
wssGBLUP2	7893	1473	35,623	0.35	930	0.70	0.68
wssGBLUP3	7893	1473	35,623	0.35	930	0.68	0.64
BayesB	1473	1473	35,636	0.25	930	0.71	1.01

[a] The estimated breeding values (EBV) were estimated with a pedigree-based animal model (P-BLUP); and the genomic EBV (GEBV) were estimated with three genomic selection (GS) models: single-step GBLUP (ssGBLUP), weighted ssGBLUP (wssGBLUP) and Bayesian method BayesB. The wssGBLUP2 and wssGBLUP3 corresponds to iteration 2 and 3, respectively

[b] For the GS models, h^2 is the proportion of phenotypic variance explained by the markers. For the P-BLUP model, h^2 is the trait narrow-sense heritability estimated from pedigree and phenotypic records. The heritability estimated on the underlying scale of liability was transformed to the observed scale of survival STATUS

[c] The predictive ability of EBV (PA_{EBV}) or GEBV (PA_{GEBV}) was defined as the correlation of mid-parent EBV or GEBV with MPP from each PTF: $PA_{EBV} = CORR(MPP, Midparent\ EBV); PA_{GEBV} = CORR(MPP, Midparent\ GEBV)$

[d] The bias of EBV ($Bias_{EBV}$) or GEBV ($Bias_{GEBV}$) was defined as the regression coefficient of performance MPP on predicted mid-parent EBV or GEBV: $Bias_{EBV} = REGRES(MPP, Midparent\ EBV); Bias_{GEBV} = REGRES(MPP, Midparent\ GEBV)$. The predicted EBV and GEBV for STATUS estimated on the underlying scale of liability were transformed to the observed scale (probability of survival)

for STATUS ($PA_{EBV} = 0.36$) (Tables 1, 2). The bias of EBV for DAYS ($\beta_{MPP.EBV} = 0.86$) deviated less from 1.0 than the bias of EBV for STATUS ($\beta_{MPP.EBV} = 0.67$).

Accuracy and bias of GEBV

The PA of GEBV (PA_{GEBV}) for DAYS ranged from 0.63 to 0.71 and the BayesB model had genomic predictions with the highest accuracy (Table 1; see Additional file 2: Figure S1). The bias of the GEBV for DAYS ($\beta_{MPP.GEBV}$) ranged from 0.65 to 1.16 and the predictions obtained with ssGBLUP were the least biased.

The PA_{GEBV} for STATUS ranged from 0.66 to 0.71 and BayesB resulted in genomic predictions with the highest accuracy (Table 2; see Additional file 3: Figure S2). The bias of GEBV for STATUS, $\beta_{MPP.GEBV}$, ranged from 0.64 to 1.01 and the predictions obtained with BayesB were the least biased.

Overall, across GS models, the accuracy of genomic predictions for STATUS were marginally higher than those for DAYS (Tables 1, 2). However, the predictions for DAYS were marginally less biased or closer to 1.0 than those for STATUS.

Comparison of accuracies of EBV and GEBV

The relative increase in accuracy of GEBV from GS models over those estimated with the classical P-BLUP model is shown in Fig. 3. Overall, the GS models substantially outperformed the P-BLUP model. The highest increase in accuracy of prediction was achieved with BayesB (DAYS = 108.8%; STATUS = 97.2%) followed by wssGBLUP at iteration 2 (wssGBLUP2) (DAYS = 97.1%; STATUS = 94.4%). The wssGBLUP2 outperformed the wssGBLUP at iteration 3 (wssGBLUP3) (DAYS = 91.2%;

STATUS = 88.9%). The lowest increase in accuracy of prediction was achieved with ssGBLUP (DAYS = 85.3%; STATUS = 83.3%).
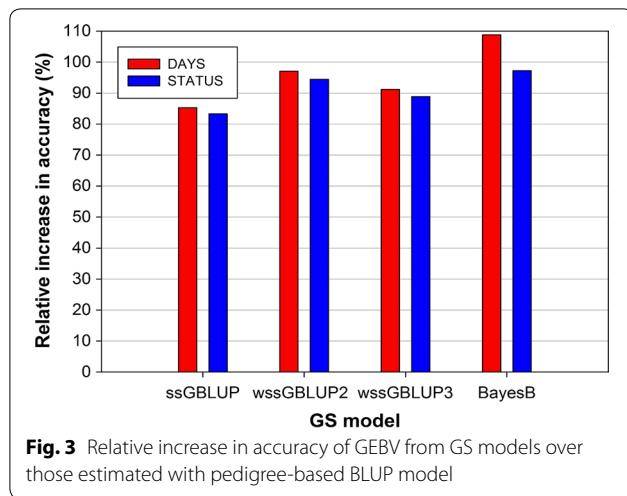
Accuracy and bias of GEBV in the five GS schemes

The MPP and the mid-parent GEBV for DAYS and STATUS for each of the 138 PTF in five GS schemes are in Additional file 4: Table S2. The accuracies of GEBV for DAYS and STATUS obtained with BayesB using the five GS schemes are in Table 3. Schemes 1 and 3 had the highest prediction accuracies (0.69 to 0.72), followed by scheme 2 (0.67) (Fig. 4). Scheme 4 GEBV had moderate accuracies (0.53 to 0.61) that were substantially lower than those for schemes 1–3. Scheme 5 had the lowest accuracies (0.22 to 0.25) among the tested GS schemes. The accuracies of GEBV from scheme 5 were even lower than the classical pedigree-based BLUP model accuracies (0.34 to 0.36) (Tables 1, 2).

For DAYS, schemes 1 and 3 had the least biased GEBV (1.16 and 1.26) and scheme 5 had the most down-biased GEBV (3.33) (Table 3). For STATUS, scheme 1 had the least biased GEBV (1.01), schemes 2 to 5 had down-biased GEBV (1.23 to 5.08), and scheme 5 had the most down-biased GEBV (5.08).

Discussion

To our knowledge, this is the first report that demonstrates that the accuracy of GEBV is higher than that of pedigree-based EBV using actual progeny performance data from a commercial finfish aquaculture species. In other fish species, the accuracy of GEBV predictions has been assessed only by cross-validation analysis using phenotypes of training animals [38, 39].

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 8 of 13

Fig. 3 Relative increase in accuracy of GEBV from GS models over those estimated with pedigree-based BLUP model

The accuracy of GEBV for DAYS and STATUS were similar when using the Bayesian method BayesB ($PA_{GEBV} = 0.71$) and higher than those estimated with ssGBLUP and wssGBLUP (Tables 1, 2). However, the accuracies of GEBV for STATUS ($PA_{GEBV} = 0.66 - 0.70$) were slightly higher than those estimated for DAYS ($PA_{GEBV} = 0.63 - 0.67$) when using ssGBLUP and wss-GBLUP methods, which may be due to (1) a better fit of the binary trait STATUS with a threshold model than the discrete data DAYS with a linear model, (2) our imprecise measure of DAYS for fish that survived the challenge (arbitrarily assigned 21 days of survival), and (3)

the resulting slightly higher heritability of STATUS compared to DAYS.

In this study, the accuracy of genomic predictions for BCWD resistance ranged from 0.63 to 0.72, which is substantially higher than accuracies of EBV estimated with the classical P- BLUP model ($PA_{EBV} = 0.34 - 0.36$) and is also significantly higher than the 0.55 maximum realized accuracy of EBV prediction using pedigree and phenotype data with a P-BLUP model given a heritability of 0.30 for BCWD resistance [40].

Given the training sample size used here ($n = 1473$) and the heritability of 0.30 for BCWD resistance, based on the deterministic expression of [41], genomic predictions with an accuracy of 0.68 can be expected if BCWD resistance is controlled by more than 500 independent loci; which is close to the accuracy of GEBV for BCWD resistance phenotypes achieved here with BayesB. Thus, assuming that BCWD resistance is controlled by more than 500 independent loci (with few genes with a moderate to large effect and many genes with a small effect) and given a heritability of 0.30, with training datasets of $n = 3000$ and $n = 10,000$ fish, we can expect to predict GEBV with an accuracy of about 0.80 and 0.93, respectively.

The accuracy of genomic prediction in dairy cattle exceeded 0.8 for milk production traits and 0.7 for health-related traits using large reference populations that included progeny-tested bulls with highly accurate phenotypes based on average daughter performance [12, 42]. In this study, it was remarkable to have genomic

Table 3 Accuracy of genomic prediction for BCWD resistance with BayesB using progeny testing families in five GS schemes

GS scheme	Family		Training size	Training–testing relationship[b]	π[c]	SNPs[d]	DAYS[e]		STATUS[e]	
	Number	Size					PA_{GEBV}[f]	$Bias_{GEBV}$[g]	PA_{GEBV}[f]	$Bias_{GEBV}$[g]
1	50	20-40[a]	1473	0.66	0.97	1069	0.71	1.16	0.71	1.01
2	50	20	991	0.50	0.98	713	0.67	1.55	0.67	1.51
3	25	40	979	1.00	0.98	713	0.69	1.26	0.72	1.23
4	25	20	497	1.00	0.987	463	0.53	1.37	0.61	1.66
5	25	20	494	0.00	0.987	463	0.25	3.33	0.22	5.08

A sample of 193 testing fish (from total $n = 930$ testing fish) were inter-mated to develop 138 progeny testing families (PTF). After disease evaluation of progeny from the 138 PTF ($n = 9968$), we estimated the mean progeny phenotype (MPP) for each PTF

[a] In scheme1, there were two groups of training families: (1) A set of 25 families with 40 offspring each that contributed fish to the testing sample; and (2) A set of 25 families with 20 offspring each that did not contribute fish to the testing sample

[b] Proportion of training fish that were full-sibs (FS) of testing fish: scheme 1 = 0.66 indicates that 66% of training fish were FS of testing fish; scheme 2 = 0.50 indicates that 50% of training fish were FS of testing fish; schemes 3 and 4 = 1.0 indicates that ALL training fish were FS of testing fish; and scheme 5 = 0.0 indicates that NONE of training fish were FS of testing fish (i.e., training and testing fish were sampled from different families)

[c] BayesB method uses a mixture parameter π that specifies the proportion of loci with zero effect, and the analyses included 35,636 effective SNPs

[d] Number of SNPs that are sampled as having non-zero effect $(1 - \pi)$ and fitted simultaneously in the multiple regression model

[e] Bacterial cold water disease (BCWD) resistance phenotypes: BCWD survival days (DAYS) and survival status (STATUS)

[f] Predictive ability of GEBV (PA_{GEBV}) was defined as the correlation of MPP with mid-parent GEBV from each PTF: $PA_{GEBV} = CORR(MPP, Midparent\ GEBV)$

[g] Bias of GEBV ($Bias_{GEBV}$) was defined as the regression coefficient of performance MPP on predicted mid-parent GEBV: $Bias_{GEBV} = REGRES(MPP, Midparent\ GEBV)$. The predicted GEBV for STATUS estimated on the underlying scale of liability were transformed to the observed scale (probability of survival)

Fig. 4 Accuracy of GEBV for BCWD resistance estimated with BayesB in five GS schemes
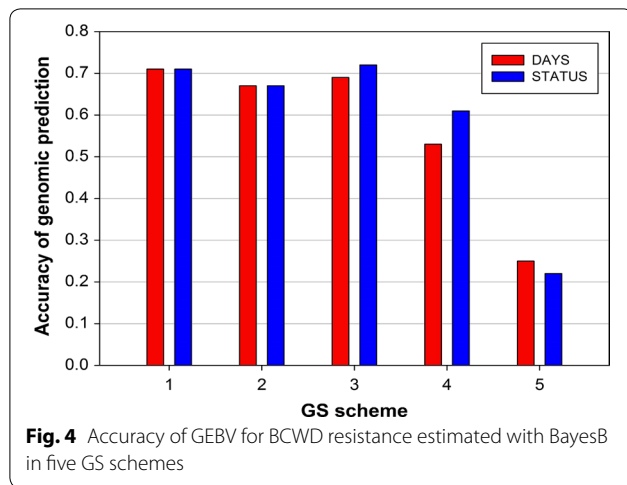
evaluations with an accuracy of 0.71 using a relatively small training dataset ($n = 1473$) in comparison to those used in dairy cattle. We hypothesize that the relatively high accuracy achieved in the current study was due to the high relationship between the training and testing fish, the small effective population size of this farmed rainbow trout population, which leads to extensive linkage disequilibrium (LD) and a substantially smaller number of effective chromosome segment effects to be estimated, hence better predictions and higher accuracies [43], and the high extent of long-range LD observed in admixed salmonid populations [38, 44]. This high extent of long-range LD is generated by the high level of admixture in the population, which also reduces the short-range LD in the population. Admixture of genetically divergent populations can result in significantly elevated LD over large genomic regions and for many generations [45–47]. Due to recombination, the decay of the admixture-generated LD (ALD) among unlinked genes is rapid (within two to four generations), but the ALD between linked genes decays more slowly. For loci that are separated by 1 cM, about 90 and 82% of the ALD will remain after 10 and 20 generations, respectively [48–50]. A similar phenomenon of high extent of long-range ALD that enabled efficient GS with relatively low marker density and a small training dataset ($n = 1963$) was reported in farmed salmonids [38].

The training and testing sample sizes used in this study were larger than the sample size that we used in a previous GS study [23], which resulted in a much better accuracy of the GEBV prediction in the current study. The number of fish in the training data and the number of progeny-tested FS families in the testing data might be close to optimal for this commercial population because we achieved a high genomic prediction accuracy of 0.72. Nonetheless, based on theoretical predictions, we can expect to further increase the accuracy of the genomic

predictions by substantially increasing the training sample size to $n = 3000$ or even $n = 10,000$.

The heritability of the trait has a significant impact on the accuracy of the predicted GEBV [51]. The heritability estimated with the P-BLUP model and the proportion of phenotypic variance explained by the markers estimated with GS models for DAYS and STATUS in this commercial population were close to the previously reported heritability for BCWD resistance in the NCCCWA breeding population [3, 23, 24].

Comparison of GS models

The differences in accuracy between the GS models that we tested here were small and all of them outperformed the classical P-BLUP model. For DAYS and STATUS, the GEBV obtained with BayesB had the highest accuracy (0.71), and the GEBV obtained with ssGBLUP had the lowest accuracy of predictions (0.63 to 0.66). The wssGBLUP2 model outperformed ssGBLUP by 0.04 units of accuracy for both BCWD phenotypes. The Bayesian method BayesB outperformed wssGBLUP2 marginally by 0.01 and 0.04 units of accuracy for STATUS and DAYS, respectively.

For DAYS and STATUS, the GEBV obtained with ssGLBUP and BayesB were the least biased and had the smallest departure from 1.0. In contrast, the GEBV obtained with wssGBLUP2 and wssGBLUP3 were the most biased and had the largest departure from 1.0. Regardless of the GS model used, DAYS had genomic predictions with a marginally lower bias than STATUS; a plausible explanation for this outcome is that the estimated additive genetic variance for the binary trait STATUS is inflated when analyzed in the underlying scale of disease liability with a threshold model, and consequently the EBV and GEBV of STATUS have marginally higher bias than those of DAYS, which is analyzed with a linear animal model.

Previously, in a different rainbow trout population, we showed that BCWD resistance is controlled by oligogenic inheritance of a few QTL with a moderate to large effect and many genes/loci each with a small effect [4–6]. Thus, given this genetic architecture, variable selection models [9, 19, 30] that fit markers with mostly moderate to large effects can yield GEBV with higher accuracy than GS models that use pedigree and phenotype records with marker genotype data in a single-step GS BLUP analysis [20, 21, 32]. Thus, our finding that wssGBLUP2, which fits only SNPs with an effect different from zero and weighted by their genetic variance [22] to emulate Bayesian variable selection models, predicts GEBV with higher accuracy than ssGBLUP and remarkably close to BayesB was largely expected.

Overall, our results are in agreement with previous reports on GS in livestock, which highlights that for quantitative traits of oligogenic inheritance that are controlled

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 10 of 13

by a few genes/loci with a moderate to large effect and many loci with a small effect, variable selection models such as BayesB and its emulator wssGBLUP outperform GBLUP-based models [52–54]. Conversely, if the quantitative trait has polygenic inheritance and follows the infinitesimal model then the GBLUP-based models, which assume a normal distribution with equal variance for all SNP effects, performs as well as variable selection models. Thus, the model used for genomic prediction is important and the relative performance of the GS model depends on the genetic architecture that underlies the trait [43].

Comparison of GS study designs

The most interesting result from comparing the different GS sampling schemes was that the accuracy of scheme 3 was similar to that of scheme 1, in spite of the smaller training data size of scheme 3 ($n = 979$ vs. $n = 1473$), which is likely due to the higher relationship between the training and testing fish in scheme 3 (1.0 vs. 0.66), and also because the average relationship among the fish in the training data was higher in scheme 3 than in scheme 1 (Fig. 2; Table 3). These results validate the notion that if the main breeding objective is to obtain high accuracy GEBV only for selection candidates (not for the entire population), then we should design GS studies that ensure a high genetic relationship between the training and testing fish, and also a high average relationship among the training fish [55, 56]. Likewise, the prediction accuracy of scheme 3 was better than that of scheme 2 (0.69 to 0.72 vs. 0.67) because of the following two design characteristics: (1) a higher relationship between training and testing fish

in scheme 3 than in scheme 2 (1.0 vs. 0.5); and (2) a higher average relationship among training fish in scheme 3 than in scheme 2 due to the larger family size of scheme 3.

The substantial superiority of scheme 3 over scheme 4 on accuracy of predictions (0.69 to 0.72 vs. 0.53 to 0.61), in spite of the same level of relationships between the training and testing fish in those two schemes, was due to the overall larger sample size and larger family size in scheme 3. Scheme 5 had genomic predictions with the lowest accuracy (0.22 to 0.25) because the relationship between training and testing fish was the lowest and the overall sample size was the lowest of the five GS training fish sampling schemes.

These results have important implications on the design of effective GS studies in finfish aquaculture using similar SNP array densities for genotyping, because they highlight the following: (1) the importance of a high relationship between training and testing fish for genomic prediction, i.e. the accuracy of predictions will drastically drop if the training and testing fish are sampled from different families within a population; and (2) the accuracy of GEBV from GS across populations will be relatively low, i.e. training sample from one population and testing from another. These results also suggest that the prediction model has to be retrained at each generation to maintain the accuracy of genomic predictions at a constant desired level across generations [57, 58].

Additional remarks

In comparison to dairy cattle and other livestock species, one of the main challenges of implementing GS in
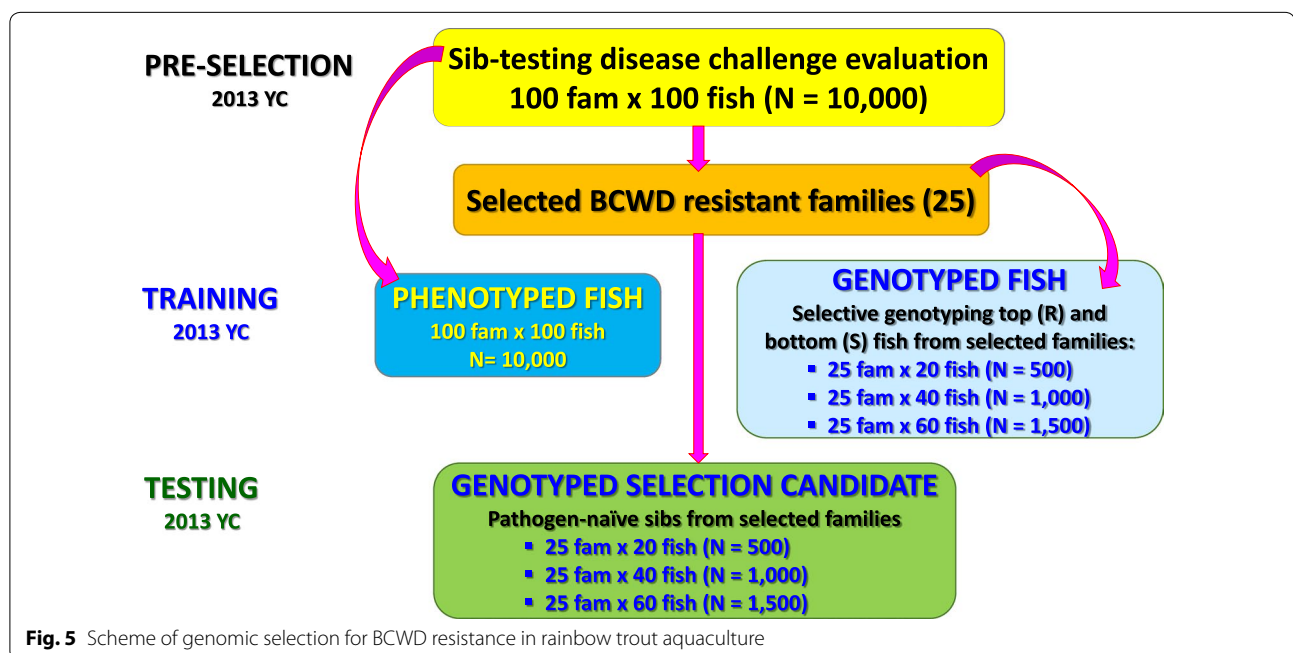
Fig. 5 Scheme of genomic selection for BCWD resistance in rainbow trout aquaculture

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 11 of 13

traditional family-based breeding programs with salmonid species is the large number of selection candidates and the limited value of the individual candidates compared to the genotyping cost. Nevertheless, the classical sib-testing scheme used in disease resistance breeding programs with salmonids can be redesigned to capitalize on the ability of GS to increase the accuracy of breeding value prediction and rate of genetic progress. To this end, for implementing GS for BCWD resistance in sib-selection schemes in the rainbow trout industry, we suggest combining a first step of sib-testing disease challenge evaluations to pre-select families for disease resistance, as suggested elsewhere [59–61], with a second step of selective genotyping individuals from the disease resistance pre-selected families to reduce genotyping costs (Fig. 5). In this GS scheme, the disease phenotype and marker genotype records from the pre-selected families can be used to train the prediction model; and then in a third step to predict GEBV for each genotyped selection candidate from families that were pre-selected at the first step. This strategy will incorporate genomic information into traditional family-based selective breeding programs and cost-effectively exploit within-family genetic variation to maximize the accuracy of genomic evaluations.

The unique features of genomic information, such as increasing the accuracy of breeding value prediction and response to selection while not increasing, or even decreasing, rates of inbreeding, are one of the main advantages of GS in livestock species. The ability of GS to reduce rates of inbreeding was reported for poultry [62] and a much larger reduction in rate of inbreeding was reported in aquaculture breeding programs due to sib-testing for both sexes [59]. The main reason for the reduction of rates of inbreeding with GS is that genomic data provide information on the Mendelian sampling terms, which reduces the emphasis placed on family selection and consequently reduces the correlations of EBV among family members and probabilities of co-selection of relatives [15]. Furthermore, since the Bulmer effect reduces between-family variation in the population, the ability to use within-family genetic variation is even more important over multiple generations of selection. Thus, because classical sib-selection with salmonid species does not exploit within-family variation, the relative advantage of GS is expected to increase if selective breeding is applied over multiple generations [61].

A major challenge for implementing GS in applied aquaculture breeding programs is assembling the large training population that is required to accurately estimate SNP effects. In this study, the number of training fish was still rather limited. Regardless, we have found that genomic predictions for BCWD resistance can be obtained with a high accuracy in the tested rainbow trout commercial breeding population by using a relatively small training sample size of $n = 1000$ and the accuracy was substantially better than the traditional P-BLUP EBV with a training sample size of only $n = 500$.

Conclusions

So far, to the best of our knowledge, this is the first study that assesses the accuracy of genomic predictions for BCWD resistance using progeny performance data and empirically tests the potential of GS to exploit within-family genetic variation in sib-selection breeding schemes in the rainbow trout industry. In this study, we have shown that (1) the accuracy of genomic predictions is substantially higher than those from a classical P-BLUP model; (2) high and near-optimal accuracy of genomic predictions for BCWD resistance can be obtained in the rainbow trout commercial population that was evaluated in this study using a relatively small training sample size of $n = 1000$; and (3) the accuracy of GEBV estimated with BayesB is higher than those from wssGBLUP3 and ssG-BLUP, followed by the accuracy of wssGBLUP2. Finally, this study provides guidelines for the implementation of GS in the rainbow trout industry.

Additional files

Additional file 1: Table S1. Mean progeny phenotype (MPP) and predictions for BCWD resistance estimated with pedigree-based BLUP (P-BLUP) and genomic selection (GS) models in progeny testing families (PTF) of rainbow trout. The data provided represent the mean progeny phenotype (MPP) and predictions (EBV and GEBV) for BCWD resistance (DAYS and STATUS) using pedigree-based BLUP (P-BLUP) and genomic selection (GS) models (ssGBLUP, wssGBLUP and BayesB) in progeny testing families (PTF) of rainbow trout. The EBV and GEBV for the binary trait STATUS estimated on the underlying scale of liability using a threshold animal model are shown as transformed to the observed scale of probability of survival using a standard normal cumulative distribution function (CDF).

Additional file 2: Figure S1. Correlation between mean progeny phenotype (MPP) and mid-parent EBV or GEBV for BCWD survival DAYS from progeny testing families (PTF). This figure represents the correlation between mean progeny phenotype (MPP) and mid-parent EBV or GEBV for BCWD survival DAYS from 138 progeny testing families (PTF) of rainbow trout.

Additional file 3: Figure S2. Correlation between mean progeny phenotype (MPP) and mid-parent EBV or GEBV for BCWD survival STATUS from progeny testing families (PTF). This figure represents the correlation between mean progeny phenotype (MPP) and mid-parent EBV or GEBV for BCWD survival STATUS from 138 progeny testing families (PTF) of rainbow trout.

Additional file 4: Table S2. Mean progeny phenotype (MPP) and genomic predictions for BCWD resistance estimated with BayesB method using five genomic selection (GS) schemes in progeny testing families (PTF) of rainbow trout. The data provided represent the mean progeny phenotype (MPP) and genomic predictions (GEBV) for BCWD resistance (DAYS and STATUS) estimated with BayesB method using five genomic selection (GS) schemes in progeny testing families (PTF) of rainbow trout. The GEBV for the binary trait STATUS estimated on the underlying scale of liability using a threshold animal model are shown as transformed to the observed scale of probability of survival using a standard normal cumulative distribution function (CDF).

Vallejo *et al. Genet Sel Evol (2017) 49:17*

Page 12 of 13

Authors' contributions

YP, TL and RV conceived and planned the study; YP, GW and JE coordinated and performed the disease challenges and collection of samples; YP coordinated and performed sample processing and genotyping; GG performed genotype data quality control and bioinformatics filtering and developed a database pipeline to assemble genotype and phenotype records; TL, JP and KM provided pedigree records and identified the fish for genotyping based on pedigree and phenotype records; RV performed statistical data analysis and genomic selection analysis with single-step GBLUP-based methods and Bayesian variable selection models, and wrote the first draft of the manuscript; BF provided support in performing BLUP, single-step and weighted single-step GBLUP analyses. All authors read and approved the final manuscript.

Author details

[1] National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV, USA. [2] Troutlodge, Inc., P.O. Box 1290, Sumner, WA, USA. [3] Animal and Dairy Science Department, University of Georgia, Athens, GA, USA.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets supporting the conclusions of this research article are included within the article and its additional files.

References

1. Nematollahi A, Decostere A, Pasmans F, Haesebrouck F. *Flavobacterium psychrophilum* infections in salmonid fish. J Fish Dis. 2003;26:563–74.
2. Barnes M, Brown M. A review of *Flavobacterium psychrophilum* biology, clinical signs, and bacterial cold water disease prevention and treatment. Open Fish Sci J. 2011;411:40–8.
3. Leeds TD, Silverstein JT, Weber GM, Vallejo RL, Palti Y, Rexroad CE, et al. Response to selection for bacterial cold water disease resistance in rainbow trout. J Anim Sci. 2010;88:1936–46.
4. Vallejo RL, Wiens GD, Rexroad CE 3rd, Welch TJ, Evenhuis JP, Leeds TD, et al. Evidence of major genes affecting resistance to bacterial cold water disease in rainbow trout using Bayesian methods of segregation analysis. J Anim Sci. 2010;88:3814–32.
5. Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE 3rd, et al. Detection and validation of QTL affecting bacterial cold water disease resistance in Rainbow trout using restriction-site associated DNA sequencing. PLoS One. 2015;10:e0138435.
6. Vallejo RL, Palti Y, Liu S, Evenhuis JP, Gao G, Rexroad CE 3rd, et al. Detection of QTL in Rainbow trout affecting survival when challenged with *Flavobacterium psychrophilum*. Mar Biotechnol. 2014;16:349–60.
7. Wiens GD, Vallejo RL, Leeds TD, Palti Y, Hadidi S, Liu S, et al. Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout: identification of QTL on chromosome Omy19. PLoS One. 2013;8:e75749.
8. Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG, et al. Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. Front Genet. 2015;6:298.
9. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker map. Genetics. 2001;157:1819–29.
10. Goddard ME, Hayes BJ, Meuwissen THE. Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breed Genet. 2011;128:409–21.
11. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci. 2009;92:433–43.
12. Wiggans GR, VanRaden PM, Cooper TA. The genomic evaluation system in the United States: past, present, future. J Dairy Sci. 2011;94:3202–11.
13. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci. 2009;92:16–24.
14. Dekkers JC. Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet. 2007;124:331–41.
15. Daetwyler HD, Villanueva B, Bijma P, Woolliams JA. Inbreeding in genome-wide selection. J Anim Breed Genet. 2007;124:369–76.
16. Habier D, Fernando RL, Dekkers JC. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;177:2389–97.
17. Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res (Camb). 2009;91:47–60.
18. de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genet. 2013;9:e1003608.
19. Fernando RL, Garrick DJ. Bayesian methods applied to GWAS. In: Gondro C, van der Werf J, Hayes B, editors. Genome-wide association studies and genomic prediction. New York: Humana Press; 2013. p. 237–74.
20. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 2010;93:743–52.
21. Legarra A, Christensen OF, Aguilar I, Misztal I. Single step, a general approach for genomic selection. Livest Sci. 2014;166:54–65.
22. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res (Camb). 2012;94:73–83.
23. Vallejo RL, Leeds TD, Fragomeni BO, Gao G, Hernandez AG, Misztal I, et al. Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. Front Genet. 2016;7:96.
24. Silverstein JT, Vallejo RL, Palti Y, Leeds TD, Rexroad CE 3rd, Welch TJ, et al. Rainbow trout resistance to bacterial cold-water disease is moderately heritable and is not adversely correlated with growth. J Anim Sci. 2009;87:860–7.
25. Palti Y, Silverstein JT, Wieman H, Phillips JG, Barrows FT, Parsons JE. Evaluation of family growth response to fishmeal and gluten-based diets in rainbow trout (*Oncorhynchus mykiss*). Aquaculture. 2006;255:548–56.
26. Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57 K single nucleotide polymorphism array for rainbow trout. Mol Ecol Resour. 2015;15:662–72.
27. Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. 2015. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf. Accessed 23 Jan 2017.
28. Falconer DS, Mackay TFK. Introduction to quantitative genetics. 4th ed. Harlow: Pearson Education Ltd.; 1996.
29. Fernando RL, Garrick DJ. GenSel—user manual for a portfolio of genomic selection related analyses. Animal Breeding and Genetics, Iowa State University Ames. 2009. http://www.biomedcentral.com/content/supplementary/1471-2105-12-186-S1.PDF. Accessed 23 Jan 2017.

Vallejo *et al. Genet Sel Evol* (2017) 49:17

Page 13 of 13

30. Garrick DJ, Fernando RL. Implementing a QTL detection study (GWAS) using genomic prediction methodology. In: Gondro C, van der Werf J, Hayes B, editors. Genome-wide association studies and genomic prediction. New York: Humana Press; 2013. p. 275–98.
31. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6:7–11.
32. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. Genet Sel Evol. 2010;42:2.
33. Odegard J, Olesen I, Gjerde B, Klemetsdal G. Evaluation of statistical models for genetic analysis of challenge-test data on ISA resistance in Atlantic salmon (*Salmo salar*): prediction of progeny survival. Aquaculture. 2007;266:70–6.
34. Cheng HH, Perumbakkam S, Pyrkosz AB, Dunn JR, Legarra A, Muir WM. Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek's disease is predominantly determined by transcriptional regulation. BMC Genomics. 2015;16:816.
35. Saatchi M, Ward J, Garrick DJ. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. J Anim Sci. 2013;91:1538–51.
36. Van Melis MH, Eler JP, Oliveira HN, Rosa GJM, Silva JAIIV, Ferraz JBS, et al. Study of stayability in Nellore cows using a threshold mode. J Anim Sci. 2007;85:1780–6.
37. Agresti A. Categorical data analysis. New York: Wiley; 1990.
38. Odegard J, Moen T, Santi N, Korsvoll SA, Kjoglum S, Meuwissen TH. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). Front Genet. 2014;5:402.
39. Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. BMC Genomics. 2015;16:969.
40. Van Vleck LD, Pollak EJ, Oltenacu EA. Genetics for the animal sciences. New York: W H Freeman & Co; 1987.
41. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One. 2008;3:e3395.
42. Lund MS, Roos AP, Vries AG, Druet T, Ducrocq V, Fritz S, et al. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet Sel Evol. 2011;43:43.
43. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. The impact of genetic architecture on genome-wide evaluation methods. Genetics. 2010;185:1021–31.
44. Rexroad CE 3rd, Vallejo RL. Estimates of linkage disequilibrium and effective population size in rainbow trout. BMC Genet. 2009;10:83.
45. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA. 1988;85:9119–23.
46. Stephens JC, Briscoe D, O'Brien SJ. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet. 1994;55:809–24.
47. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am J Hum Genet. 1998;63:241–51.
48. Wilson JF, David B, Goldstein DB. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-semitic hybrid population. Am J Hum Genet. 2000;67:926–35.
49. Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genet. 2001;68:198–207.
50. Moen T, Hayes B, Baranski M, Berg PR, Kjoglum S, Koop BF, et al. A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. BMC Genomics. 2008;9:223.
51. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica. 2009;136:245–57.
52. Fernando RL, Dekkers JCM, Garrick DJ. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet Sel Evol. 2014;46:50.
53. Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaber RL, Decker JE, et al. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. Genet Sel Evol. 2015;47:23.
54. Meuwissen THE, Hayes B, Goddard M. Genomic selection: a paradigm shift in animal breeding. Anim Front. 2016;6:6–14.
55. Pszczola M, Strabel T, Mulder HA, Calus MPL. Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci. 2012;95:389–400.
56. Wu X, Lund MS, Sun D, Zhang Q, Su G. Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. J Anim Breed Genet. 2015;132:366–75.
57. Pszczola M, Calus MPL. Updating the reference population to achieve constant genomic prediction reliability across generations. Animal. 2016;10:1018–24.
58. Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Implementation of genomic selection in the poultry industry. Anim Front. 2016;6:23–31.
59. Sonesson AK, Meuwissen TH. Testing strategies for genomic selection in aquaculture breeding programs. Genet Sel Evol. 2009;41:37.
60. Lillehammer M, Meuwissen THE, Sonesson AK. A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. Genet Sel Evol. 2013;45:39.
61. Odegard J, Meuwissen THE. Identity-by-descent genomic selection using selective and sparse genotyping. Genet Sel Evol. 2014;46:3.
62. Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Response and inbreeding from a genomic selection experiment in layer chickens. Genet Sel Evol. 2015;47:59.