frontiers in Genetics

ORIGINAL RESEARCH
published: 12 June 2019
doi: 10.3389/fgene.2019.00543

Optimizing Genomic Prediction of Host Resistance to Koi Herpesvirus Disease in Carp

Christos Palaiokostas[1,2], Tomas Vesely[3], Martin Kocour[4], Martin Prchal[4], Dagmar Pokorova[3], Veronika Piackova[4], Lubomir Pojezdal[3] and Ross D. Houston[1]**

[1] Royal (Dick) School of Veterinary Studies, The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom, [2] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [3] Veterinary Research Institute, Brno, Czechia, [4] Faculty of Fisheries and Protection of Waters, South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, University of South Bohemia České Budějovice, Vodňany, Czechia

OPEN ACCESS

Edited by:
Johann Sölkner,
University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Wanchang Zhang,
Nanchang University, China
Paulino Martínez,
University of Santiago de Compostela, Spain

***Correspondence:**
Christos Palaiokostas
christos.palaiokostas@slu.se
Ross D. Houston
ross.houston@roslin.ed.ac.uk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 03 December 2018
Accepted: 22 May 2019
Published: 12 June 2019

Citation:
Palaiokostas C, Vesely T, Kocour M, Prchal M, Pokorova D, Piackova V, Pojezdal L and Houston RD (2019) Optimizing Genomic Prediction of Host Resistance to Koi Herpesvirus Disease in Carp. Front. Genet. 10:543. doi: 10.3389/fgene.2019.00543

Genomic selection (GS) is increasingly applied in breeding programs of major aquaculture species, enabling improved prediction accuracy and genetic gain compared to pedigree-based approaches. Koi Herpesvirus disease (KHVD) is notifiable by the World Organization for Animal Health and the European Union, causing major economic losses to carp production. GS has potential to breed carp with improved resistance to KHVD, thereby contributing to disease control. In the current study, Restriction-site Associated DNA sequencing (RAD-seq) was applied on a population of 1,425 common carp juveniles which had been challenged with Koi herpes virus, followed by sampling of survivors and mortalities. GS was tested on a wide range of scenarios by varying both SNP densities and the genetic relationships between training and validation sets. The accuracy of correctly identifying KHVD resistant animals using GS was between 8 and 18% higher than pedigree best linear unbiased predictor (pBLUP) depending on the tested scenario. Furthermore, minor decreases in prediction accuracy were observed with decreased SNP density. However, the genetic relationship between the training and validation sets was a key factor in the efficacy of genomic prediction of KHVD resistance in carp, with substantially lower prediction accuracy when the relationships between the training and validation sets did not contain close relatives.

Keywords: KHVD, carp, RAD-seq, genomic selection, aquaculture breeding

INTRODUCTION

Genomic selection (GS) has become a cornerstone of genetic improvement in both plant and livestock breeding, enabling improved prediction accuracy, control of inbreeding, and (in some cases) reduction in generation interval compared to traditional pedigree-based approaches (Meuwissen et al., 2016; Hickey et al., 2017). The landmark paper of Meuwissen et al. (2001) highlighted the concept of breeding value prediction based on the joint merit of all markers distributed throughout the genome, and the advent of high-throughput DNA sequencing and development of SNP arrays in the subsequent decade made this concept a practical reality. While the application of genomics in aquaculture breeding has traditionally lagged behind the plant and terrestrial livestock sector, it is gaining momentum with reference genome assemblies and

SNP arrays now available for most of the key aquaculture species (Robledo et al., 2017; Yue and Wang, 2017). Both simulation and empirical studies suggest that considerable improvement in breeding value prediction accuracy is plausible, even with relatively modest SNP marker densities (Sonesson and Meuwissen, 2009; Lillehammer et al., 2013; Ødegård et al., 2014; Tsai et al., 2015; Correa et al., 2017; Vallejo et al., 2017, 2018; Robledo et al., 2018).

Infectious diseases present a major and persistent threat to sustainable aquaculture production, and breeding for improved host resistance is an increasingly important component of mitigation (Houston et al., 2017). Common carp (*Cyprinus carpio*) is one of the world's most important freshwater aquaculture species, particularly in Asia and Europe. However, koi herpesvirus disease (KHVD), also known as Cyprinid herpesvirus-3 (CyHV-3) disease is a major threat to carp farming and is listed as a notifiable disease by the European Union (Taylor et al., 2010) and the World Organization for Animal Health (OIE, 2018). Encouragingly, resistance to KHVD has been shown to be a highly heritable trait with estimates ranging between 0.50 and 0.79 (Ødegård et al., 2010; Palaiokostas et al., 2018a). The potential of selective breeding for improved KHVD resistance in carp (utilizing information from challenge trials) has been illustrated by several studies which demonstrated large variation in survival both between-family (Dixon et al., 2009; Tadmor-Levi et al., 2017) and between strain (Shapira et al., 2005; Piačková et al., 2013). Further, a significant QTL associated with resistance to KHVD has been identified (Palaiokostas et al., 2018a). Nevertheless, the potential of GS for improving KHVD resistance in carp has not yet been studied.

While SNP arrays are available for several aquaculture species, and are commonly used in some of the most advanced commercial breeding programs (e.g., Atlantic salmon), they tend to be relatively expensive and can suffer from ascertainment bias (Robledo et al., 2017). Genotyping by sequencing technology, such as RAD-seq (Baird et al., 2008) and subsequent variants, have also been effective in studying complex traits such as disease resistance in aquaculture species, and testing GS (Vallejo et al., 2016; Barría et al., 2018; Palaiokostas et al., 2018b; Aslam et al., 2018). Disease resistance is particularly amenable to GS, because typically it is not possible to record on selection candidates themselves (Yáñez et al., 2014), and is typically measured on their close relatives (e.g., full siblings) in aquaculture breeding programs (Gjedrem and Rye, 2016). While effective, the limitations of current GS methods in aquaculture include (i) that the genotyping is typically expensive, partially due to the high-density marker genotyping, and (ii) the accuracy of prediction drops rapidly when the genetic relationship between the training and validation populations decreases (e.g., Tsai et al., 2016).

Family-based breeding programs are at a formative stage in common carp, including a program focused on the Amur mirror carp breed in Europe (Prchal et al., 2018a,b), where improvement of disease resistance is a major breeding goal. The main aim of the current study was to investigate the potential of GS to predict host resistance to KHVD in common carp using genome-wide SNP markers generated by RAD sequencing. An additional aim was to investigate the importance of SNP marker density in genomic prediction accuracy, with a view to future low-density SNP panels for cost-effective GS. Finally, the impact of genetic relationship between the training and validation sets was assessed by comparing prediction accuracy in groups of closely and distantly related fish.

MATERIALS AND METHODS

Population Origin and Disease Challenge

The origin of the samples and the details of the disease challenge experiment have been fully described previously (Palaiokostas et al., 2018a). In brief, the study was performed on a population of Amur mirror carp that was created at the University of South Bohemia in České Budějovice, Czech Republic in May 2014 using an artificial insemination method (Vandeputte et al., 2004). The population was the result of four factorial crosses of five dams x ten sires (20 dams and 40 sires in total). A cohabitation KHV challenge was performed on randomly sampled progeny of these crosses. Mortality of individual fish was recorded for a period of 35 days post infection (dpi), by which stage the mortality level had returned to baseline. In total, phenotypic records regarding survival/mortality were documented for 1,425 animals. Presence of KHV in a sample of dead fish (n = 100) was confirmed by PCR according to guidelines by the Centre for Environment, Fisheries and Aquaculture Science, United Kingdom (Cefas) (Pokorova et al., 2010). The entire experiment was conducted in accordance with the law on the protection of animals against cruelty (Act no. 246/1992 Coll. of the Czech Republic) upon its approval by Institutional Animal Care and Use Committee (IACUC).

RAD Sequencing and Parentage Assignment

The RAD library preparation protocol followed the methodology originally described in Baird et al. (2008), presented in detail in Palaiokostas et al. (2018c). In brief, RAD libraries were sequenced by BMR Genomics (Padova, Italy) in fourteen lanes of an Illumina NextSeq 500, using 75 base paired-end reads (v2 chemistry). Reads missing the restriction site, with ambiguous barcodes and PCR duplicates were identified and discarded using the Stacks v2.0 software (Catchen et al., 2011). Remained sequenced reads were aligned to the common carp reference genome assembly version *GCA_000951615.2* (Xu et al., 2014) using bowtie2 (Langmead and Salzberg, 2012). Uniquely aligned reads were retained for downstream analysis. The aligned reads were sorted into RAD loci and SNPs were identified from both P1 and P2 reads using the Stacks software v2.0 (Catchen et al., 2011). Opposed to our previous study (Palaiokostas et al., 2018b) variant calling in Stacks v2.0 and above utilizes information from both P1 and P2 ends, while prior versions were using only P1 ends. SNPs were detected using gstacks *(–var-alpha 0.001 –gt-alpha 0.001 –min-mapq 40)*. Only single SNPs from each individual RAD locus where considered for downstream analysis to minimize the possibility of genotypic errors. SNPs with minor allele frequency (MAF) below 0.05, greater than 25% missing data were discarded. The R package hsphase (Ferdosi et al., 2014)

was used for parentage assignment allowing for a maximum genotyping error of 2%. The aligned reads in the format of bam files were deposited in the National Centre for Biotechnology Information (NCBI) repository under project ID PRJNA414021.

Genomic Prediction Models

Overall binary survival (0 = dead, 1 = alive) was used as the phenotype to assess the potential of GS for improved resistance to KHVD in common carp. Several commonly used GS models were tested on the data using the R package BGLR for binary traits (Pérez and de los Campos, 2014): specifically rrBLUP, BayesA, BayesB (Meuwissen et al., 2001) and BayesC (Habier et al., 2011). In addition, pedigree-based BLUP (Henderson, 1975) was evaluated using the same software. The general form of the fitted models was:

$$\mathbf{l} = \mathbf{Xb} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, \quad (1)$$

where $\mathbf{l}$ is the vector of latent variables, $\mathbf{b}$ is the vector of the fixed effects (intercept, standard length), $\mathbf{X}$ is the incidence matrix relating phenotypes with the fixed effects, $\mathbf{Z}$ the incidence matrix relating the underlying liability with the genotypes, $\boldsymbol{\alpha}$ the vector of SNP effects using the corresponding prior distribution for each of the aforementioned Bayesian models and $\mathbf{e}$ the vector of residuals. The parameters of each model were estimated by Markov chain Monte Carlo (MCMC) using Gibbs sampling (110,000 iterations; burn-in: 10,000; thin: 10). Convergence of the resulting posterior distributions was assessed both visually (inspecting the resulting MCMC plots) and analytically using the R package coda v0.19-1 software (Plummer et al., 2006).

Prediction Metrics for KHVD Resistance

The prediction performance of the utilized models was tested using the following metrics:

- Accuracy
- Receiver operator characteristic (ROC) curves

The prediction accuracy was approximated as:

$$r = (\mathbf{GEBV}, \mathbf{y})/h, \quad (2)$$

where $\mathbf{y}$ is the vector of recorded phenotypes, **(G) EBV** is the (genomic) estimated breeding values and h is the square root of the heritability (h^2 = 0.50 using the genomic relationship matrix as described in Palaiokostas et al., 2018a).

Receiver operator characteristic curves were used to assess the efficacy of classifying the animals as resistant or susceptible, using either the pedigree- or the genomic-based models. The area under the curve (AUC) metric (Hanley and McNeil, 1982; Wray et al., 2010) was used to interpret the performance of the genomic prediction models, with values of 1 representing the perfect classifier.

Genomic Prediction With Varying SNP Densities

Genomic prediction models were applied using datasets of varying SNP density using either MAF or linkage disequilibrium (LD) values as thresholds for filtering. In particular, to obtain the reduced density SNP panels for genomic prediction, a strategy of retaining SNPs surpassing a sequentially increased MAF threshold was applied, as described in Robledo et al. (2018). These MAF thresholds were 0.1 (3,993 SNPs), 0.25 (1,619 SNPs) and 0.35 (802 SNPs).

In addition, reduced density SNP datasets were obtained by applying filtering based on LD values. LD amongst SNP pairs was calculated using SNPrune (Calus and Vandenplas, 2018). Thereafter, only SNP pairs below a sequentially increased LD value were retained. The LD thresholds were 0.15 (1,006 SNPs), 0.25 (2,895 SNPs), 0.35 (5,118 SNPs).

Five-fold cross-validation was performed for all the density varying SNP datasets in order to test the efficiency of correctly classifying animals in the validation set as resistant or susceptible. The dataset was randomly split into sequential training (n = 1008) and validation sets (n = 251). The number of resistant and susceptible animals in each validation set was proportional to the overall survival of the challenged population. In the validation sets, the phenotypes of the animals were masked, and their (genomic) estimated breeding values – (G)EBV – were estimated based on the prediction model derived from the training set. This cross-validation procedure was repeated five times to minimize potential bias.

Testing the Impact of Genetic Relationship on Genomic Prediction

Four different scenarios were tested for evaluating the impact of genetic relationships between training and validation sets. In scenario 1 (S1), the formation of training and validation sets required the existence of full-siblings in both sets for each family. For scenario 2 (S2) the formation of validation and training sets allowed the existence of only half siblings between the two sets (and no full siblings). Both in S1 and S2 the cross validation procedure was repeated five times in order to reduce potential bias, while the size of the validation set was 290 animals on each replicate. In scenario (S3) the genomic prediction models were tested by sequentially assigning each of the four factorial crosses (mean = 315 animals; sd = 81 animals) as a validation set, using the remaining three as a training set. This approach resulted in relatively unrelated training and validation sets, since it avoided the inclusion of full/half sibs in both the training and the validation sets. The genomic prediction models were tested on the dataset comprised of the full SNP data. Since pedigree information was not available for prior generations, pBLUP could not be used for obtaining meaningful predictions across the factorial cross groups. Finally, a scenario 4 (S4) was performed as control where no restrictions were applied in the formation of training and validation sets (i.e., they were taken at random). Cross validation in S4 was performed five times with the size of the validation sets being set to 290 animals. The S4 scenario was in fact similar with the approaches tested in the previous section regarding varying SNP densities with the only difference being the size of the validation set. The full SNP dataset was used for all the tested scenarios.

RESULTS

Disease Challenge

The mean weight of the genotyped carp juveniles was 16.3 g (SD 4.6) and the mean standard length (SL) was 77 mm (SD 7.1). Mortalities began at 12 dpi reaching a maximum daily rate between 21 and 24 dpi (98 – 130 mortalities per day) decreasing thereafter (**Supplementary File S2**). Observed mortalities displaying typical KHVD symptoms (weakness, lethargy, loss of equilibrium, erratic swimming, sunken eyes, excessive mucous production, discoloration, and hemorrhagic lesions on the skin and gills).

RAD Sequencing and Parentage Assignment

2.8 billion paired-end reads were uniquely aligned to the common carp genome assembly (GenBank assembly accession *GCA_000951615.2*) representing approximately 82% of reads passing initial quality filters (missing restriction site, ambiguous barcodes and PCR duplicates). Approximately 5% of those reads had a mapping quality below 40 and were discarded. In total 397,047 putative RAD loci were identified with a mean coverage of 21X (SD = 7.6, min = 1.3X, max = 58.5X). 15,615 SNPs found in more than 75% of the genotyped animals and with a MAF above 0.05 were retained for downstream analysis (**Supplementary File S1**).

The carp progeny were assigned to unique parental pairs allowing for a maximum genotypic error rate of 2%. In total 1,259 offspring were uniquely assigned (**Supplementary File S3**), comprising 195 full-sib families (40 sires, 20 dams) ranging from 1 to 21 animals per family with a mean size of 6 (SD 4). The individual dam contribution to the population ranged from 9 to 99 animals with a mean of 61 (SD 23), while the sire contribution ranged from 7 to 53 animals with a mean of 30 (SD 12). In addition, the mean weight and length per full-sib family were approximately 16 (SD 2.8) g and 76 (SD 4.5) mm respectively. Finally, mean survival per full sib family was 34% (**Figure 1**).

Impact of SNP Density on Genomic Prediction

Datasets of varying genotyping density were comprised of 15,615 SNPs (D1; full dataset; **Supplementary File S4**) and in the case of MAF as the filtering criterion of 3,993 (D2; MAF 0.1), 1,619 (D3; MAF 0.25) and 802 (D4; MAF 0.35) SNPs. The accuracy of genomic prediction of breeding values was assessed and compared to prediction using a pedigree-based approach. Prediction accuracy with pBLUP was 0.49, compared to 0.53 – 0.54 for the genomic prediction models applied using D1 (**Table 1**). Prediction accuracies for D2 ranged between 0.52 and 0.53, while in the case of D3 and D4 prediction accuracy for all genomic models was 0.49 and 0.46 respectively (**Figure 2A**). Following estimation of ROC curves, the genomic models for D1 had a maximum AUC estimate of 0.74 as opposed to 0.71 using

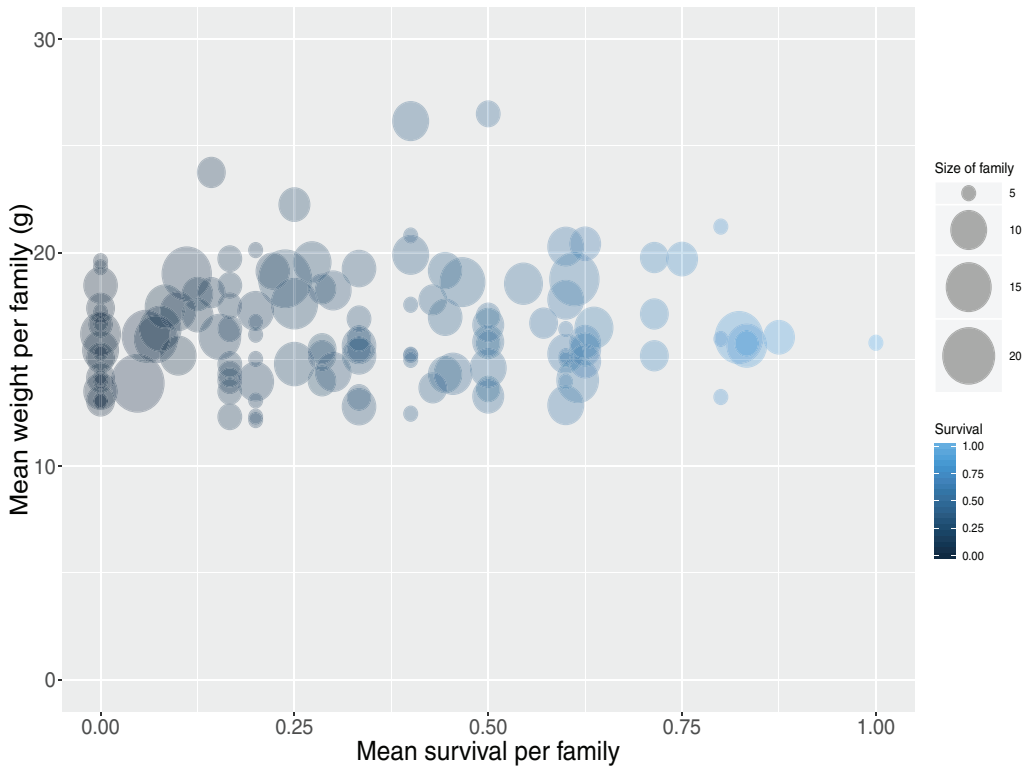

FIGURE 1 | Mean weight and survival for all assigned full-sib families.

TABLE 1 | Mean survival accuracy for D1[1] (5-fold cross validation; 5 replicates).

Reps	pBLUP	rrBLUP	BayesA	BayesB	BayesC
1st	0.47	0.51	0.52	0.52	0.52
2nd	0.49	0.54	0.55	0.55	0.54
3rd	0.48	0.52	0.53	0.54	0.53
4th	0.50	0.54	0.54	0.55	0.55
5th	0.49	0.52	0.53	0.54	0.52
Mean	0.49	0.53	0.53	0.54	0.53

[1] *15,615 SNPs.*

pBLUP. AUC for D2 was 0.73 for all genomic models. In the case of D3 and D4 the AUC for all genomic models was 0.71 and 0.70 respectively.

Regarding the reduced density SNP datasets obtained using LD pruning, the number of SNPs in the sets with the LD thresholds of 0.15, 0.25, and 0.35 were 1,006 (LD1), 2,895 (LD2) and 5,118 (LD3) respectively. The genomic prediction accuracy obtained for LD1 was very slightly higher than pBLUP using the BayesB and BayesC models (<1% increase), while the AUC was the same. In the case of rrBLUP and BayesA for the same SNP dataset the estimates were 2 and 1% lower compared to pBLUP for accuracy and AUC respectively. Using datasets of higher SNP density resulted in the increase of both the accuracy and the AUC metrics as observed previously for the reduced density datasets filtered by MAF. In particular, accuracy for LD2 and LD3 ranged between 0.52 and 0.54 and AUC between 0.72 and 0.74 (**Figure 2B**), which were very similar to the accuracy and AUC values obtained for the full SNP dataset (15,615 SNPs).

Impact of Genetic Relationship on Genomic Prediction

For the scenario S1, where all animals in the validation set had full sibs in the training set the genomic prediction accuracy was approximately 0.56, which was marginally higher (∼ 4% increase) than the random allocation of animals into training and validation sets described above. In S2 where the design of the validation set allowed the inclusion of only corresponding half sibs in the training and validation set, the genomic prediction accuracy fell to ∼ 0.53. In S3 where the training and validation sets were set up to correspond to separate factorial crosses, the mean accuracy for the genomic models was markedly lower, and ranged between 0.16 and 0.20. Finally, in the scenario where training and validation sets were set up without posing any restrictions estimated, such that close relatives are likely to be included in both sets, accuracy ranged between 0.52 and 0.54 for the genomic prediction models and 0.49 for pBLUP (**Table 2**).

The obtained AUC values from the ROC curves were 0.74 (BayesB; **Figure 3**) and 0.72 for S1 and S2 for the genomic prediction models, while the corresponding AUC values from pBLUP were 0.72 and 0.69 respectively. For S3 the estimated AUC values for the genomic models were again substantially lower and ranged between 0.57 and 0.58. In S4, where no restrictions were applied regarding the inclusion of full/half sibs on both training and validation sets, the AUC values were between 0.72 and 0.74, comparable to S1 and S2 (**Table 2**).

DISCUSSION

In the current study, genotyping by sequencing was applied to study genomic prediction of resistance to KHVD in carp, including testing the impact of SNP marker density and genetic relationship between training and validation sets. While genomic data in the form of genetic markers can be a valuable addition to selective breeding for disease resistance, the methods of applying the data depend on the underlying genetic architecture of the trait. In the case of major QTL such as resistance to Infectious Pancreatic Necrosis in salmon (Houston et al., 2008; Moen et al., 2009), it may be most effective to use QTL-targeted marker-assisted selection, and in the case of polygenic traits GS is likely to be preferable. In our previous study we identified a QTL associated with KHVD resistance in common carp located on chromosome 33 (Palaiokostas et al., 2018a). However, this QTL accounted for approximately 7% of the genetic variation in the trait, highlighting that multiple additional loci are involved. Further, using genomic prediction models that incorporate variable selection – i.e., allow for the existence of QTL of large effect – did not result in significant improvement in prediction accuracy compared to ridge regression BLUP, which supports the involvement of many genomic regions in the trait (Meuwissen et al., 2001; Kizilkaya et al., 2010; Habier et al., 2013).

Since genotyping cost is generally related to SNP marker density, determining the lowest SNP density that retains maximum genomic prediction accuracy is a logical goal. In the current study, reducing SNP density from 15,615 to 2,895 resulted in minor decreases in prediction accuracy, with 1,000–1,600 SNPs giving approximately the same accuracy as pBLUP. Furthermore, the LD-pruned dataset of approximately 5,000 SNPs resulted in the same prediction accuracy performance as the full dataset (15,615 SNPs). A more drastic impact of genetic relationship between training and validation sets on prediction accuracy was observed. The highest prediction efficiency was observed in scenario S1 where animals in the validation set had full siblings in the training set. Prediction efficiency decreased 6–8% in the scenario allowing for only the inclusion of half-siblings (and no full siblings) in the training and validation sets but was still comparable to the results when the sets were established at random. Interestingly, the impact of the lower genetic relationships on pBLUP accuracy was greater, and it dropped by approximately 16% between S1 and S2. This may indicate that genomic prediction models have the potential to utilize distant relationships compared to pBLUP, especially in the current set up where there was only a two generation pedigree. Furthermore, when the training set comprised three of the factorial cross groups and the validation set comprised the fourth, thus resulting in no shared full/half sibs between the two sets, the accuracy dropped massively to 0.16–0.17 (15,615 SNPs). The decrease in prediction accuracy with more distant relationships is to be expected, thus close relationships between training and validation sets is a necessary prerequisite for successfully implementing GS (Meuwissen et al., 2013), and it highlights the importance of obtaining genotype and phenotype records on close relatives of selection candidates in future carp breeding programs using genomic (and pedigree) selection.

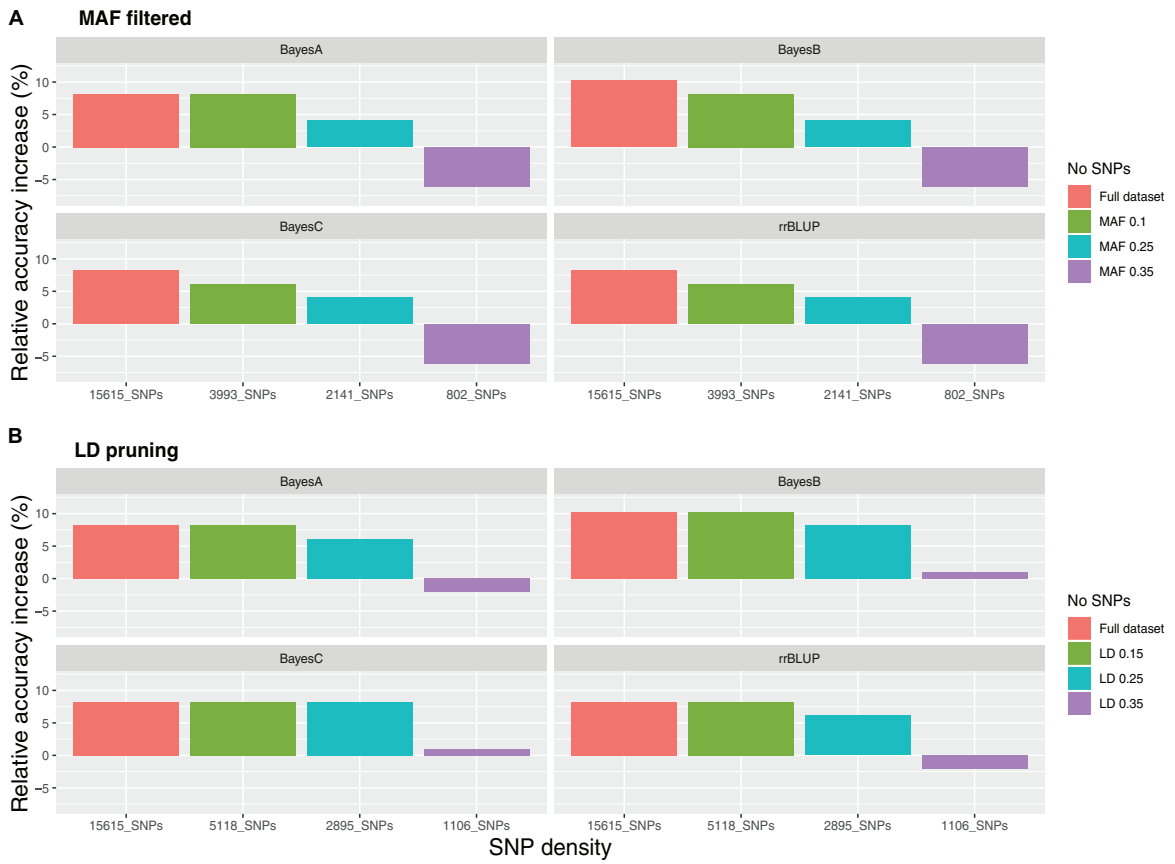

FIGURE 2 | Relative accuracy of genomic prediction models compared to pedigree BLUP for varying SNP densities. **(A)** SNP filtering based on minor allele frequency and **(B)** SNP filtering based on linkage disequilibrium.

TABLE 2 | Prediction metrics for varying genetic relationships in the validation set.

Scenario	pBLUP		rrBLUP		BayesA		BayesB		BayesC	
	Acc[1]	AUC[2]	Acc[1]	AUC[2]	Acc[1]	AUC[2]	Acc[1]	AUC[2]	Acc[1]	AUC[2]
S1	0.52	0.72	0.56	0.74	0.57	0.74	0.57	0.74	0.56	0.74
S2	0.45	0.69	0.53	0.72	0.52	0.72	0.52	0.72	0.52	0.72
S3	N/A	N/A	0.17	0.57	0.16	0.57	0.18	0.58	0.20	0.57
S4	0.49	0.71	0.52	0.72	0.53	0.74	0.54	0.74	0.52	0.72

[1]Accuracy; [2]Area under curve. Genetic relationships for the 4 tested scenarios were for S1: Full sibs on the training set for all animals of the validation set (n = 290), for S2: Half sibs on the training set for all animals in the validation set (n = 290), for S3: cross validation performed for each of the breeding cross. No full/half sibs on the training set for any of the animals in the validation set (n = 315) and for S4: No restrictions applied (n = 290) for genetic relationship between training and validation set.

Testing genomic prediction on binary traits such as survival, presents a challenge to define a suitable test metric for selecting the best performing model, especially when survival deviates significantly from 50%. Solely relying on correlation derived accuracy for model assessment in this case could result in suboptimal selection decisions. Suitable metrics for evaluating prediction efficiency in binary traits and thus selecting the best performing models for estimating breeding values include the AUC from ROC curves.

The AUC values provide a commonly used metric for assessing the prediction efficacy of binary classifiers, taking into consideration both the rate of false positives and false negatives with values of one suggesting 100% successful classification. This approach has been routinely used to test the efficacy of prediction models in disease resistance studies both in humans (Wray et al., 2010), livestock (Tsairidou et al., 2014) and aquaculture (Palaiokostas et al., 2018b) amongst others. In the current study, genomic prediction using the marker density scenarios of ∼ 3,000 SNPs and above resulted in a slight improvement (∼ 4%) of AUC compared to pBLUP. Performing predictions using approximately 1,000 SNPs resulted in the same AUC value (0.71) as pBLUP, while when using approximately 800 SNP the

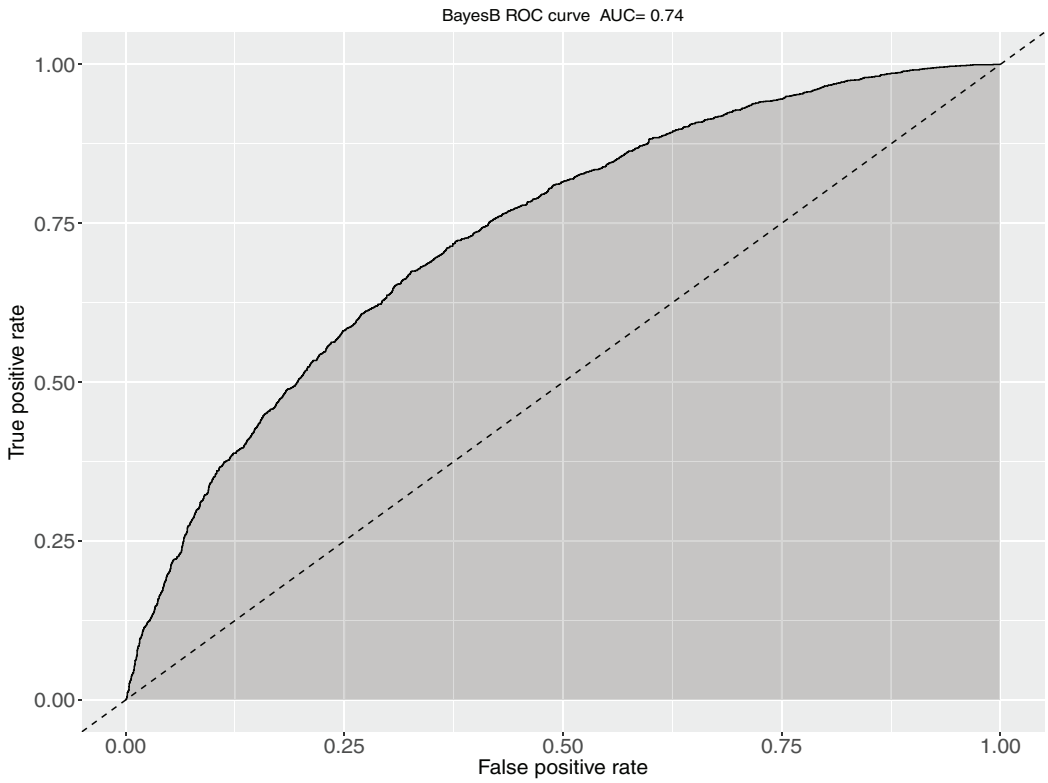

FIGURE 3 | The ROC curve and corresponding AUC metrics for BayesB-based predictions of KHV survival. The plot was obtained from aggregation of a 5-fold cross validation scheme when full sibs existed in the training set for every animal of the validation set.

estimated AUC value was 0.70 which is slightly inferior. A gradual decrease was observed regarding the estimated AUC values for the scenarios of varying genetic relationship as was also the case for the prediction accuracy metric. As expected highest values were obtained in the scenario of highest relationships between training and validation sets (S1). Most striking effect of the impact of genetic relationships between the above sets, however, was observed in the scenario where the training and validation sets were set up to be most distantly related, where the estimated AUC values ranged between 0.56 and 0.57, which are substantially lower than all other tested scenarios, but still useful.

In summary the results from the current study demonstrate that GS was more efficient than pBLUP in predicting for KHVD resistant carp. The consistency of improvement in prediction accuracy versus pedigree-based accuracy across multiple scenarios highlights flexibility and robustness to different approaches, and it may allow circumvention of limitations posed by incomplete pedigree records. Of major importance is the fact that relatively low density marker panels could be of value for genomic prediction without loss of accuracy. However, close relationships between training and validation sets are key, with substantial loss of prediction accuracy in the scenario where the sets were relatively unrelated. Pedigree-based prediction was also efficient in scenarios with recorded relationships between training and validation sets, possibly partly because KHVD resistance is a high-heritable trait (h^2 = 0.5 – 0.79), but genetic markers were required to assign the pedigree in the factorial crosses. Future studies testing the efficiency of single-step BLUP approaches (Aguilar et al., 2011; Legarra et al., 2014) could potentially prove beneficial by allowing genomic predictions based on larger datasets (only a portion of the dataset would be genotyped, thus reducing costs). Overall our results help inform the use of genetic markers in carp breeding to enable improvement of disease resistance, with downstream benefits of helping prevent KHVD outbreaks in carp aquaculture.

ETHICS STATEMENT

The entire experiment was conducted in accordance with the law on the protection of animals against cruelty (Act No. 246/1992 Coll. of the Czech Republic) upon its approval by Institutional Animal Care and Use Committee (IACUC).

AUTHOR CONTRIBUTIONS

TV, MK, MP, VP, and RH conceived the study and contributed to designing the experimental structure. TV, DP, and LP carried out the challenge experiment. CP carried out DNA extractions,

RD library preparation, and sequence data processing. CP and RH carried out parentage assignment and the quantitative genetic analyses. All authors contributed to drafting the manuscript.

FUNDING

The authors are supported by funding from the European Union's Seventh Framework Programme (FP7 2007–2013) under grant agreement no. 613611 (FISHBOOST). CP and RH gratefully acknowledge Institute Strategic Funding Grants to The Roslin Institute (Grant Nos. BBS/E/D/20002172, BBS/E/D/30002275, and BBS/E/D/10002070). MK and MP were also supported by project, Biodiverzita (CZ.02.1.01/0.0/0.0/16_025/0007370). VP was also supported by project PROFISH (CZ.02.1.01/0.0/0.0/16_019/0000869) both under the Ministry of Education, Youth and Sports of the Czech Republic. TV, DP, and LP were also supported by Ministry of Agriculture of the Czech Republic (Project MZE-RO 0518).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00543/full#supplementary-material

REFERENCES

Aguilar, I., Misztal, I., Legarra, A., and Tsuruta, S. (2011). Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128, 422–428. doi: 10.1111/j.1439-0388.2010.00912.x

Aslam, M. L., Carraro, R., Bestin, A., Cariou, S., Sonesson, A. K., Haffray, P., et al. (2018). Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *BMC Genet.* 19:43. doi: 10.1186/s12863-018-0631-x

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376

Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2018). Genomic predictions and genome-wide association study of resistance against piscirickettsia salmonis in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3* 8, 1183–1194. doi: 10.1534/g3.118.200053

Calus, M. P. L., and Vandenplas, J. (2018). SNPrune: an efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. *Genet. Sel. Evol.* 50:34. doi: 10.1186/s12711-018-0404-z

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

Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., and Yáñez, J. M. (2017). The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi)* resistance in Atlantic salmon (*Salmo salar*). *Genet. Sel. Evol.* 49:15. doi: 10.1186/s12711-017-0291-8

Dixon, P. F., Joiner, C. L., Way, K., Reese, R. A., Jeney, G., Jeney, Z., et al. (2009). Comparison of the resistance of selected families of common carp, *Cyprinus carpio* L., to koi herpesvirus: preliminary study. *J. Fish Dis.* 32, 1035–1039. doi: 10.1111/j.1365-2761.2009.01081.x

Ferdosi, M. H., Kinghorn, B. P., van der Werf, J. H. J., Lee, S. H., and Gondro, C. (2014). hsphase: an R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. *BMC Bioinformatics* 15:172. doi: 10.1186/1471-2105-15-172

Gjedrem, T., and Rye, M. (2016). Selection response in fish and shellfish: a review. *Rev. Aquacult.* 10, 168–179. doi: 10.1111/raq.12154

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194, 597–607. doi: 10.1534/genetics.113.152207

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186

Hanley, A. J., and McNeil, J. B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) Curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447.

Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* 49, 1297–1303. doi: 10.1038/ng.3920

Houston, R. D., Brasileira, R., and De Zootecnia. (2017). Invited review future directions in breeding for disease resistance in aquaculture species. *Bras. Zootec.* 46, 545–551. doi: 10.1590/s1806-92902017000600010

Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., et al. (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178, 1109–1115. doi: 10.1534/genetics.107.082974

Kizilkaya, K., Fernando, R. L., and Garrick, D. J. (2010). Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J. Anim. Sci.* 88, 544–551. doi: 10.2527/jas.2009-2064

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci.* 166, 54–65. doi: 10.1016/j.livsci.2014.04.029

Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45:39. doi: 10.1186/1297-9686-45-39

Meuwissen, T., Hayes, B., and Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1, 221–237. doi: 10.1146/annurev-animal-031412-103705

Meuwissen, T., Hayes, B., and Goddard, M. (2016). Genomic selection: a paradigm shift in animal breeding. *Anim. Front.* 6, 6–14. doi: 10.2527/af.2016-0002

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368

Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., Meuwissen, T. H., et al. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet* 5:402. doi: 10.3389/fgene.2014.00402

Ødegård, J., Olesen, I., Dixon, P., Jeney, Z., Nielsen, H. M., Way, K., et al. (2010). Genetic analysis of common carp (*Cyprinus carpio*) strains. II: resistance to koi herpesvirus and Aeromonas hydrophila and their relationship with pond survival. *Aquaculture* 304, 7–13. doi: 10.1016/j.aquaculture.2010.03.017

OIE (2018). *OIE-Listed Diseases, Infections and Infestations in Force in 2018*. Available at: http://www.oie.int/animal-health-in-the-world/oie-listed-diseases-2018/ (accessed March 26, 2018).

Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J. S., Haffray, P., Morin, T., et al. (2018c). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet. Sel. Evol.* 50:30. doi: 10.1186/s12711-018-0401-2

Palaiokostas, C., Kocour, M., Prchal, M., and Houston, R. D. (2018b). Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. *Front. Genet.* 9:82. doi: 10.3389/fgene.2018.00082

Palaiokostas, C., Robledo, D., Vesely, T., Prchal, M., Pokorova, D., Piackova, V., et al. (2018a). Mapping and sequencing of a significant quantitative trait locus affecting resistance to koi herpesvirus in common carp. *G3* 8, 3507–3513. doi: 10.1534/g3.118.200593

Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442

Piačková, V., Flajšhans, M., Pokorová, D., Reschová, S., Gela, D., Čížek, A., et al. (2013). Sensitivity of common carp, *Cyprinus carpio* L., strains and crossbreeds reared in the Czech Republic to infection by cyprinid herpesvirus 3 (CyHV-3; KHV). *J. Fish Dis.* 36, 75–80. doi: 10.1111/jfd.12007

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.

Pokorova, D., Reschova, S., Hulova, J., Vicenova, M., Vesely, T., and Piackova, V. (2010). Detection of cyprinid herpesvirus-3 in field samples of common and koi carp by various single-round and nested PCR methods. *J. World Aqua. Soc.* 41, 773–779. doi: 10.1111/j.1749-7345.2010.00419.x

Prchal, M., Kause, A., Vandeputte, M., Gela, D., Allamellou, J.-M., et al. (2018a). The genetics of overwintering performance in two-year old common carp and its relation to performance until market size. *PLoS One* 13:e0191624. doi: 10.1371/journal.pone.0191624

Prchal, M., Bugeon, J., Vandeputte, M., Kause, A., Vergnet, A., Zhao, J., et al. (2018b). Potential for genetic improvement of the main slaughter yields in common carp with *in vivo* morphological predictors. *Front. Genet.* 9:283. doi: 10.3389/fgene.2018.00283

Robledo, D., Matika, O., Hamilton, A., and Houston, R. D. (2018). Genome-wide association and genomic selection for resistance to amoebic gill disease in atlantic salmon. *G3* 8, 1195–1203. doi: 10.1534/g3.118.200075

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* 10, 670-682.

Shapira, Y., Magen, Y., Zak, T., Kotler, M., Hulata, G., and Berta Levavi, S. (2005). Differential resistance to koi herpes virus (KHV)/carp interstitial nephritis and gill necrosis virus (CNGV) among common carp (*Cyprinus carpio* L.) strains and crossbreds. *Aquaculture* 245, 1–11. doi: 10.1016/j.aquaculture.2004.11.038

Sonesson, A. K., and Meuwissen, T. H. E. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* 41:37. doi: 10.1186/1297-9686-41-37

Tadmor-Levi, R., Asoulin, E., Hulata, G., and David, L. (2017). Studying the genetics of resistance to CyHV-3 disease using introgression from feral to cultured common carp strains. *Front. Genet.* 8:24. doi: 10.3389/fgene.2017.00024

Taylor, N. G. H., Dixon, P. F., Jeffery, K. R., Peeler, E. J., Denham, K. L., Way, K., et al. (2010). Koi herpesvirus: distribution and prospects for control in England and Wales. *J. Fish Dis.* 33, 221–230. doi: 10.1111/j.1365-2761.2009.01111.x

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48:47. doi: 10.1186/s12711-016-0226-9

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., et al. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 16:969. doi: 10.1186/s12864-015-2117-9

Tsairidou, S., Woolliams, J. A., Allen, A. R., Skuce, R. A., McBride, S. H., Wright, D. M., et al. (2014). Genomic prediction for tuberculosis resistance in dairy cattle. *PLoS One* 9:e96728. doi: 10.1371/journal.pone.0096728

Vallejo, R. L., Leeds, T. D., Fragomeni, B. O., Gao, G., Hernandez, A. G., Misztal, I., et al. (2016). Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. *Front. Genet.* 7:96. doi: 10.3389/fgene.2016.00096

Vallejo, R. L., Liu, S., Gao, G., Fragomeni, B. O., Hernandez, A. G., Leeds, T. D., et al. (2017). Similar genetic architecture with shared and unique quantitative trait loci for bacterial cold water disease resistance in two rainbow trout breeding populations. *Front. Genet.* 8:156. doi: 10.3389/fgene.2017.00156

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

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Gela, D., et al. (2004). Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). *Aquaculture* 235, 223–236. doi: 10.1016/j.aquaculture.2003.12.019

Wray, N. R., Yang, J., Goddard, M. E., Visscher, P. M., and Kimberly, R. (2010). The genetic interpretation of area under the ROC curve in genomic profiling (Schork, N. J., Ed.). *PLoS Genet.* 6:e1000864. doi: 10.1371/journal.pgen.1000864

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., et al. (2014). Genome sequence and genetic diversity of the common carp, Cyprinus carpio. *Nat. Genet.* 46, 1212–1219. doi: 10.1038/ng.3098

Yáñez, J. M., Houston, R. D., and Newman, S. (2014). Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5:415. doi: 10.3389/fgene.2014.00415

Yue, G. H., and Wang, L. (2017). Current status of genome sequencing and its applications in aquaculture. *Aquaculture* 468, 337–347. doi: 10.3390/ijms19041083

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PM declared a past co-authorship with several of the authors CP and RH to the handling Editor.

Copyright © 2019 Palaiokostas, Vesely, Kocour, Prchal, Pokorova, Piackova, Pojezdal and Houston. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.