

RESEARCH ARTICLE

Open Access



Viral nervous necrosis resistance in gilthead sea bream (*Sparus aurata*) at the larval stage: heritability and accuracy of genomic prediction with different training and testing settings

Sara Faggion^{1*} , Paolo Carnier¹, Raffaella Franch¹, Massimiliano Babbucci¹, Francesco Pascoli², Giulia Dalla Rovere¹, Massimo Caggiano³, Hervé Chavanne³, Anna Toffan² and Luca Bargelloni¹

Abstract

Background The gilthead sea bream (*Sparus aurata*) has long been considered resistant to viral nervous necrosis (VNN), until recently, when significant mortalities caused by a reassortant nervous necrosis virus (NNV) strain were reported. Selective breeding to enhance resistance against NNV might be a preventive action. In this study, 972 sea bream larvae were subjected to a NNV challenge test and the symptomatology was recorded. All the experimental fish and their parents were genotyped using a genome-wide single nucleotide polymorphism (SNP) array consisting of over 26,000 markers.

Results Estimates of pedigree-based and genomic heritabilities of VNN symptomatology were consistent with each other (0.21, highest posterior density interval at 95% (HPD95%): 0.1–0.4; 0.19, HPD95%: 0.1–0.3, respectively). The genome-wide association study suggested one genomic region, i.e., in linkage group (LG) 23 that might be involved in sea bream VNN resistance, although it was far from the genome-wide significance threshold. The accuracies (r) of the predicted estimated breeding values (EBV) provided by three Bayesian genomic regression models (Bayes B, Bayes C, and Ridge Regression) were consistent and on average were equal to 0.90 when assessed in a set of cross-validation (CV) procedures. When genomic relationships between training and testing sets were minimized, accuracy decreased greatly ($r = 0.53$ for a validation based on genomic clustering, $r = 0.12$ for a validation based on a leave-one-family-out approach focused on the parents of the challenged fish). Classification of the phenotype using the genomic predictions of the phenotype or using the genomic predictions of the pedigree-based, all data included, EBV as classifiers was moderately accurate (area under the ROC curve 0.60 and 0.66, respectively).

Conclusions The estimate of the heritability for VNN symptomatology indicates that it is feasible to implement selective breeding programs for increased resistance to VNN of sea bream larvae/juveniles. Exploiting genomic information offers the opportunity of developing prediction tools for VNN resistance, and genomic models can be trained on EBV using all data or phenotypes, with minimal differences in classification performance of the trait phenotype. In a long-term view, the weakening of the genomic ties between animals in the training and test sets leads to decreased genomic prediction accuracies, thus periodical update of the reference population with new data is mandatory.

*Correspondence:

Sara Faggion

sara.faggion@unipd.it

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Viral nervous necrosis (VNN), known also as viral encephalopathy and retinopathy (VER), is one of the most important diseases in aquaculture. VNN is endemic in the Mediterranean area [1] and the occurrence of the disease has been widely reported in both wild and farmed fish [2]. The pathogenesis of VNN consists in the vacuolation and necrosis of the nervous cells of the brain, spinal cord and retina [3] followed by clinical signs of neurological damages (abnormal swimming patterns, lethargy, skin darkening, and loss of appetite) [4].

The causative agent of VNN is Nodavirus [or nervous necrosis virus (NNV)], a non-enveloped icosahedral virus of the Nodaviridae family, genus Betanodavirus. In fish, four Betanodavirus genotypes have been detected according to a partial nucleotide sequence of the coat protein gene: striped jack nervous necrosis virus (SJNNV, the first isolated strain), tiger puffer nervous necrosis virus (TPNNV), barfin flounder nervous necrosis virus (BFNNV) and red-spotted grouper nervous necrosis virus (RGNNV) [5].

Gilthead sea bream (*Sparus aurata*) has long been considered not susceptible to VNN and a potential asymptomatic vector of the infection [6, 7]. After several mass mortality outbreaks observed in Mediterranean sea bream hatcheries [8–10], it was discovered that the NNV reassortant strain RGNNV/SJNNV possesses a particular tropism to sea bream and poses a new and unexpected threat to the sea bream aquaculture industry. Larval stages are the most susceptible to NNV, with rather severe outcomes (100% mortality rates) [9, 10]. The outcome of the infection seems to be less predictable in juveniles, and to depend on age and environmental conditions; however, mortality rates can still be moderate to high (10 to 100%) [9, 11]. During natural outbreaks in sea bream hatcheries, infected larvae showed a clearly recognizable symptomatology, implying clinical and behavioural signs such as lethargy, apathy, anorexia, abnormal swimming patterns, loss of balance and swim bladder hyperinflation [8–10]. Those clinical and behavioural signs were reported to always precede the death [8–10], and thus can be used as indicator traits of VNN mortality in sea bream.

Vaccines or alternative therapeutics against NNV infection are not available; and anyway, the protection of the larval stages by vaccination is precluded by the immaturity of the immune system. In sea bream, well-developed anterior kidney, thymus and spleen are observed at 1, 22–29 and 12 d post-hatching, respectively, but lymphocytes do not appear before 47–54 d post-hatching [12].

The genetic improvement of individual resistance against the reassortant NNV strain is thus a major goal in sea bream breeding. In other commercially important

species (European sea bass, Asian sea bass and Atlantic cod), significant additive genetic variation for VNN resistance has been reported [13–15], which is promising for sea bream. Along with a long-term control of the disease and a cumulative and permanent increase in resistance over generations [16], selective breeding for enhanced resistance offers the opportunity to reduce disease outbreaks within offspring batches, and to enhance the sustainability and profitability of aquaculture production [17, 18].

Phenotyping for disease resistance based on challenge trials is difficult, due to the complexity of experimental infections, and limited to the relatives of breeding candidates due to biosecurity reasons. In this context, the implementation of genomic selection procedures is of great interest and is made feasible by the recent availability of high-density marker arrays for sea bream [19]. Genomic selection relies on genotyping and phenotyping, in challenge trials, a reference population consisting of full- and half-sibs of the breeding candidates. Genotypes and phenotypes of the reference population are used to estimate allele substitution effects for a dense panel of genetic markers, which are exploited in the prediction of genomic estimated breeding values (GEBV). Such estimated effects are validated in a “testing” population and then used in the routine prediction of the genetic merit of future breeding candidates, which need to be genotyped only [20].

Testing the consistency of prediction equations in a long-term view is of great importance: in commercial breeding, genomic models are first trained using a reference population consisting of full- and half-sibs of the candidate breeders. Random k-fold CV have been demonstrated to provide overly optimistic genomic prediction accuracies [21–25], thus alternative strategies to assess the accuracy of prediction equations should consider the decrease in the genomic ties between animals in training and testing sets (i.e., the animals to be predicted are distant relatives of the animals included in the reference population). For livestock species, the proportion of close relatives in training and testing sets of CV is well-known to affect the accuracy of genomic predictions both for traits with high and low heritability [22, 26, 27].

To date, in gilthead sea bream, neither the genetic variation in nor the genomic prediction accuracy for VNN resistance have been investigated. Hence, the aims of this study were to estimate the genetic and genomic parameters of VNN resistance in challenged gilthead sea bream larvae, to estimate the effects of single nucleotide polymorphisms (SNPs) on the VNN resistance phenotype through a genome-wide association study (GWAS) and to assess the accuracy of genomic predictions of the trait or EBV using a novel

SNP array for sea bream. The accuracy of genomic prediction models was assessed using both random CV and two validation settings that mimic a reduction of the genomic relatedness between training and testing sets: a CV based on genomic clustering and a leave-one-family-out validation focused on the parents of the challenged fish.

Methods

Production of the experimental fish and NNV challenge test

The experimental fish were produced in a commercial hatchery (Panittica Italia Società Agricola srl, Torre Canne di Fasano, Brindisi, Italy) in three independent full-factorial mating designs (10 sires \times 12 dams, 10 sires \times 7 dams, 9 sires \times 6 dams). At 16 d post-hatching, nearly 1300 sea bream larvae, NNV-free tested, were transferred to the Istituto Zooprofilattico Sperimentale delle Venezie (IZSVE, Legnaro, Padova). Larvae were distributed into one close-system 250-L tank filled with artificial saltwater at a salinity of 35–37‰ (i.e., UV and carbon filtered tap water with artificial balanced salt, Instant Ocean- Aquarium System), a temperature of 19 ± 2 °C and oxygen 6–8 ppm, and exposed to an artificial photoperiod of 8 h of light and 16 h of darkness. After five days of acclimation, larvae were infected by immersion using the reassortant betanodavirus strain (RGNNV/SJNNV) isolated from animals belonging to a fish farm during a disease outbreak (VNNV/S.aurata/Farm1/461-1/Nov2014) [8]. Two hundred and fifty mL of viral suspension ($10^{7.80}$ TCID₅₀/mL) (TCID₅₀ for median tissue culture infectious dose 50%) was added to the tank. The final infectious titre, obtained by back titration of the water, was $10^{5.45}$ TCID₅₀/mL. Larvae were exposed to the same amount of virus at the same time. A few hours after distributing the viral suspension, the water in the tank was discarded and replaced with clean water. Fish were checked three times a day to identify individuals showing typical clinical signs of VNN (abnormal swimming patterns, lethargy). As soon as a symptomatic larva was detected, it was sacrificed and stored in 96% ethanol for subsequent genomic DNA extraction. The experiment ended 14 d post-challenge and non-symptomatic larvae were sacrificed and stored in 96% ethanol. Overall, 1229 symptomatic and non-symptomatic larvae were collected.

The experimental infection protocol was evaluated by the Istituto Zooprofilattico Sperimentale delle Venezie (IZSVE) Animal Welfare Body and Ethics Committee (Opinion CE.IZSVE.3/2016 of 24/10/216) and approved by the Italian Ministry of Health (Law decree 101/2017-PR of 02/02/2017).

Genomic DNA extraction, genotyping, parentage assignment and survival analysis

Genomic DNA (gDNA) was extracted from the whole larva using a modified salting-out protocol described by Pardo et al. [28]. gDNA concentration was quantified using the Qubit dsDNA BR Assay Kit (Invitrogen – Life Technologies) and gDNA integrity was assessed by visualization on a 1% TAE agarose gel stained with SYBR® Safe DNA Gel Stain (Invitrogen – Thermo Fisher Scientific). gDNA was successfully extracted from 1044 of the 1229 larvae. gDNA from 54 parental tissue samples (fins) was extracted with the Invisorb Spin Tissue Mini Kit (Invitek Molecular Berlin).

Genotyping of the experimental fish and their parents was performed at IdentiGEN (Ireland) using the Med-Fish SNP array [19] which contains approximately 30K SNPs for the gilthead sea bream. Nine hundred and seventy-four fish and all their parents (54) were successfully genotyped. SNP quality control and genotype calling from the intensity files were performed using the Axiom™ Analysis Suite software v. 5.1.1.1 (Thermo Fisher Scientific) with default parameter values for diploid species (call rate > 97%, dish quality control (QC) > 0.82; QC is a metric provided by ThermoFisher generated by measuring signals at multiple sites on the genome known to be invariable across individuals), resulting in 27,706 available SNPs per fish. SNPs with a minor allele frequency (MAF) lower than 1%, a missing genotype in more than 15% of the individuals or that deviated from the expected Hardy–Weinberg equilibrium frequencies ($p < 0.001$) were discarded. Missing genotypes were imputed using the FImpute software [29], resulting in a dataset of 26,591 SNP genotypes per fish.

The APIS/R [30] software package was used to assign parentage. It uses the observed distributions of average Mendelian transmission probabilities to set assignment thresholds, and enables the estimation of the proportion of offspring with missing parental genotypes. For all APIS runs, parameters were set at default values, with the exception of the error rate, which was set to 5%. Parentage assignment to a unique parental pair was achieved for all the experimental fish. In total, 160 full-sibs families were identified as being generated by 23 dams and 29 sires, with the number of offspring per family ranging from 1 to 55. The number of offspring per sire ranged from 1 to 125, while the number of offspring per dam ranged from 1 to 339. After removing the family with a size of 1, 972 individuals from 22 dams and 28 sires were retained for the subsequent analyses.

The Kaplan–Meier product-limit survival curve of the challenged fish was estimated using survival/R [31].

Additive genetic variance and pedigree-based heritability

Estimation of the variance components, parameters, and prediction of breeding values were based on the solution of a threshold (probit) animal model. In such a model, the observed binary phenotypes for VNN symptomatology ($y=0$ for the asymptomatic animals and $y=1$ for the symptomatic ones) were assumed to depend on the value of an unobservable and normally distributed latent variable λ (liability) such that:

$$P(y = 0|x) = P(\lambda \leq \lambda_t|x),$$

$$\text{and } P(y = 1|x) = P(\lambda > \lambda_t|x),$$

where λ_t is a critical cut-off value or threshold and x is a set of explanatory variables. Using the cumulative standard normal density function as a link function, such probabilities can be obtained as follows:

$$P(\lambda \leq \lambda_t|x) = \Phi(\lambda_t - \lambda|x), P(\lambda > \lambda_t|x) = 1 - \Phi(\lambda_t - \lambda|x),$$

where $\Phi(\cdot)$ denotes the cumulative standard normal density function (probit function).

The model used to interpret the variation in the liability was:

$$\lambda = \mu + \mathbf{W}\mathbf{u} + \mathbf{e}, \tag{1}$$

where λ is a vector of liabilities, μ is a constant vector corresponding to model intercept, \mathbf{u} is a vector of unknown random animal additive genetic effects, \mathbf{W} is an incidence matrix relating λ to \mathbf{u} , \mathbf{e} is a vector of random residuals. Additive genetic effects and the residuals were assumed to be distributed as $N(0, \mathbf{A}\sigma_a^2)$ and $N(0, \mathbf{I}\sigma_e^2)$, respectively, where $N()$ indicates a normal probability density function, \mathbf{A} is the numerator relationship matrix, σ_a^2 is the additive genetic variance, \mathbf{I} is an identity matrix of appropriate order and σ_e^2 is the residual variance.

Marginal posterior distributions of the genetic variance in VNN symptomatology were estimated using a Bayesian approach employing Monte-Carlo Markov chain (MCMC) and the Gibbs sampler, as implemented in the software TM [32]. A single Gibbs chain of 1,000,000 samples was generated, the first 1000 were discarded as burn-in and from the remaining 999,000 samples, we saved one from every 100. The convergence of each Markov chain was checked using the Geweke convergence diagnostic [33].

Heritability of VNN symptomatology was estimated as:

$$h_p^2 = \frac{\sigma_a^2}{\sigma_p^2}, \tag{2}$$

where σ_a^2 is the additive genetic component of the variance and σ_p^2 is the phenotypic variance (i.e., $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$).

The median of the estimated marginal posterior distribution was used a point estimate of the heritability. The 95% highest posterior density intervals (HPD95%) and the probability for h_p^2 to be greater than 0.2 (arbitrary critical threshold) were obtained from the estimated posterior density of h_p^2 using the *boa/R* software [34].

Prediction of breeding values

Estimated breeding values for VNN symptomatology were obtained using Model (1), using all available information (i.e., the observed phenotype of the individual and those of its full- and half-sibs). Variance components used to solve the model were those obtained in the previous analysis. A Bayesian approach employing MCMC and Gibbs sampling methods was implemented using the software TM [32].

Genomic variance and heritability

Genomic variance and genomic heritability for VNN symptomatology were estimated with Bayesian procedures as implemented in BGLR/R [35] and following the methodology described by de Los Campos et al. [36]. Briefly, given $\sigma_a^2 = \sigma_g^2 + \sigma_g^2$, where σ_a^2 is the additive

genetic variance of the trait, σ_g^2 represents the amount of additive genetic variance captured by a linear regression on available marker genotypes (i.e., the genomic variance) and σ_g^2 is the “missing” additive genetic variance, the genomic heritability is defined as the proportion of variance in phenotypes explained by the regression on marker genotypes:

$$h_g^2 = h_p^2 \frac{\sigma_g^2}{\sigma_a^2} = \frac{\sigma_g^2}{\sigma_p^2}, \tag{3}$$

where h_p^2 is the pedigree-based heritability of the trait and $\frac{\sigma_g^2}{\sigma_a^2}$ is the proportion of additive genetic variance explained by a linear regression on marker genotypes. Hence, the upper limit of the genomic heritability is the pedigree-based heritability, that we assumed to be properly estimated using Model (1) and Eq. (2).

Three regression models were implemented: Bayes B, Bayes C [37] and Bayesian Ridge Regression [38]. A major difference between the models is represented by the prior density used for the marker genotype effects; further details can be found in Pérez and de Los Campos [35]. All models included an intercept and the effects of the animal genotypes at 26,591 SNPs, coded as the number of copies of the minor allele (0, 1 or 2). Features of the estimated marginal posterior density of genomic heritability were obtained through the same procedure as that described for the pedigree-based heritability.

Genome-wide association study

A GWAS was performed to test the association between the VNN symptomatology phenotype and SNP genotypes (26,591 SNPs) using *gaston/R* [39] following a mixed logistic regression model:

$$\boldsymbol{\eta} = \boldsymbol{\mu} + \mathbf{W}\mathbf{u} + \boldsymbol{z}\boldsymbol{\alpha}, \quad (4)$$

where $\boldsymbol{\eta}$ is a vector of logits as $\log\left[\frac{\pi_i}{1-\pi_i}\right]$, π_i is the probability that the phenotype is 1 (symptomatic) for animal i , $\boldsymbol{\mu}$ is the model intercept, \mathbf{u} is an unknown vector of random animal additive genetic effects assumed to be $N(0, \mathbf{G}\sigma_a^2)$ where $N()$ indicates a normal probability density function, \mathbf{G} is the genomic relationship matrix and σ_a^2 is the estimated additive genetic variance, \mathbf{W} is an incidence matrix relating the vector of logits $\boldsymbol{\eta}$ to \mathbf{u} , $\boldsymbol{\alpha}$ is the unknown allele substitution effect of a single SNP, \boldsymbol{z} is a vector of marker genotypes, coded as the number of copies of the minor allele, relating $\boldsymbol{\eta}$ to $\boldsymbol{\alpha}$. The genome-wide significance threshold was set to $p\text{-value} = 0.05/p$, where p is the number of SNPs considered in the analysis. Hence, to address the problem of multiple testing, a Bonferroni correction was applied to the significance threshold. Manhattan plots of GWAS were obtained with the R package *ggplot2* [40].

Genomic predictions

Genotypes at 26,591 SNPs were used as predictors of the phenotype or pseudo-phenotypes (EBV) for VNN symptomatology. Three Bayesian regression models were fitted to the data using *BGLR/R* [35]: Bayes B, Bayes C and Bayesian Ridge Regression. Bayesian regression models were used because they simultaneously fit a larger number of markers than the number of observations available for the analysis (the well-known “small n , large p ” problem), by combining phenotypic information with priors for marker effects and variance components. Markers are fitted as random effects, accounting for most of the genetic variance. The general form of such models was:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (5)$$

where \mathbf{y} is a vector of pseudo-phenotypes (EBV) or, in the case of binary phenotype (VNN symptomatology), corresponds to a vector of liabilities, $\boldsymbol{\mu}$ is the vector of the constant term or model intercept, \mathbf{X} is a $n \times p$ matrix of SNP genotypes (n is the number of animals, p is the number of markers), coded as the number of copies of the minor allele (0, 1 or 2), $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients (SNP effects) to be estimated, and \mathbf{e} is a vector of random residuals. The residuals were assumed to be $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where $N()$ indicates a normal probability density function, \mathbf{I} is an identity matrix of appropriate order and σ_e^2 is the residual variance.

Each Bayesian analysis was carried out generating a single Gibbs chain of 800,000 samples, with a burn-in of 100,000 iterations and a thinning interval of 50 samples.

In the analysis of the phenotype for VNN symptomatology, the probit function was used as a function to link the probability of each category (0 = asymptomatic, 1 = symptomatic) to the linear predictor, whereas a linear model was used for pseudo-phenotypes (EBV); the residual variance was set to 1.

Model performance for the prediction of EBV

Prediction performance of the Bayesian models was assessed in three CV strategies: (a) a fourfold random CV, (b) a fourfold CV based on genomic clustering (K-means clustering), and (c) a leave-one-family-out (LOFO) validation, focused on the parents of the challenged animals. These strategies were used to mimic three genomic selection scenarios, where the genomic prediction equations are obtained by training models with information from a reference population that could be more closely (random CV) or more distantly (K-means clustering and LOFO) related to the animals to be predicted.

In the first strategy (a), the dataset consisting of EBV that were obtained according to the methodology described above and Model (1), was randomly split into four equally-sized data segments. Seventy-five percent of the data (training set) were used to obtain the solutions for allele substitution effects. The pseudo-phenotypes of the remaining 25% of the data (test set) were masked and predicted from the genotypes using solutions of the Bayesian models obtained in the analysis of the training set. The analysis was repeated four times; predictions for each test set were saved and, at the end, aggregated with those of the other test sets. Prediction accuracy was measured as the Pearson product-moment correlation between EBV and genomic-predicted values of EBV. To assess prediction bias, the regression coefficient (slope, β) of the EBV on the genomic-predicted EBV was also calculated, with slopes equal to 1 indicating absence of bias.

The accuracy in the estimation of the additive genetic value, which depends on the accuracy of EBV and on the accuracy of genomic-predicted EBV, was calculated as in [41]:

$$r = \sqrt{(\text{corr}(\text{EBV}, \text{genomic predicted EBV}))^2 \cdot \text{reliability}(\text{EBV})}, \quad (6)$$

where $(\text{corr}(\text{EBV}, \text{genomic predicted EBV}))^2$ is the squared Pearson product-moment correlation between EBV and genomic-predicted values of EBV; reliabilities of EBV were calculated, for each individual, as $r_{\text{EBV}}^2 = 1 - (\frac{\text{PEV}}{\sigma_a^2})$, where PEV is the prediction error variance, and σ_a^2 is the additive genetic variance.

In the second strategy (b), the K-means method clustered the animals according to distance of genomic relationships, reducing the relatedness between training and testing populations while maximizing the relatedness within each cluster. K-means clustering was performed using the function `kmeans` (stats/R) [42], which implements the Hartigan and Wong algorithm [43].

Clusters were used one at a time as a test set. As in (a), pseudo-phenotypes (EBV obtained according to the methodology described above and Model (1)) of the test set were masked and predicted from the genotypes using solutions of the Bayesian models obtained in the analysis of the training set consisting of the other three clusters. The predictions were saved and, at the end of the procedure, aggregated with those of the other clusters to compute, as a measure of prediction accuracy, the Pearson product-moment correlation between EBV and genomic-predicted EBV; prediction bias was assessed as the regression coefficient (slope, β) of the EBV on the genomic-predicted EBV.

In the third procedure (c), the EBV obtained according to the methodology described above and Model (1), were retrieved for the parents of the experimentally-infected fish. Solutions for allele substitution effects were obtained by training the models on the EBV of all animals except the offspring of the sire (or the dam) used as a test set (just one parent in the test set). The EBV of the parent in the test set was then predicted from its genotypes by making use of the solutions of Bayesian models obtained in the analysis of the training set. This was

replicated as many times as the number of parents. The prediction and EBV for each parent were saved and at the end of the LOFO, aggregated with those of the other parents to compute the Pearson product-moment correlation between the EBV and the genomic-predicted EBV; prediction bias was assessed as the regression coefficient (slope, β) of the EBV on the genomic-predicted EBV.

Model performance for the classification of the phenotype for VNN symptomatology

Different classifiers of the phenotype for VNN symptomatology were tested, including the EBV, the genomic predictions of the phenotype and of the EBV provided by Bayesian models. Classification performance was assessed in a set of five independently generated fourfold CV. For each CV, three metrics were used to evaluate the performance for the classification of the phenotype for VNN symptomatology: the area under the receiving operating characteristics (ROC) curve (AUC), the classification accuracy (ACC) computed as (true positives + true negatives)/total number of samples, and the Matthews correlation coefficient (MCC). All the metrics were computed using the package `ROCR/R` [44].

Results

Nervous necrosis virus challenge test

The Kaplan–Meier product-limit survival curve for the challenged fish is reported in Fig. 1. Mortality reached a peak at day 7 post-challenge. In total, 566 symptomatic larvae were identified between days 6 and 13

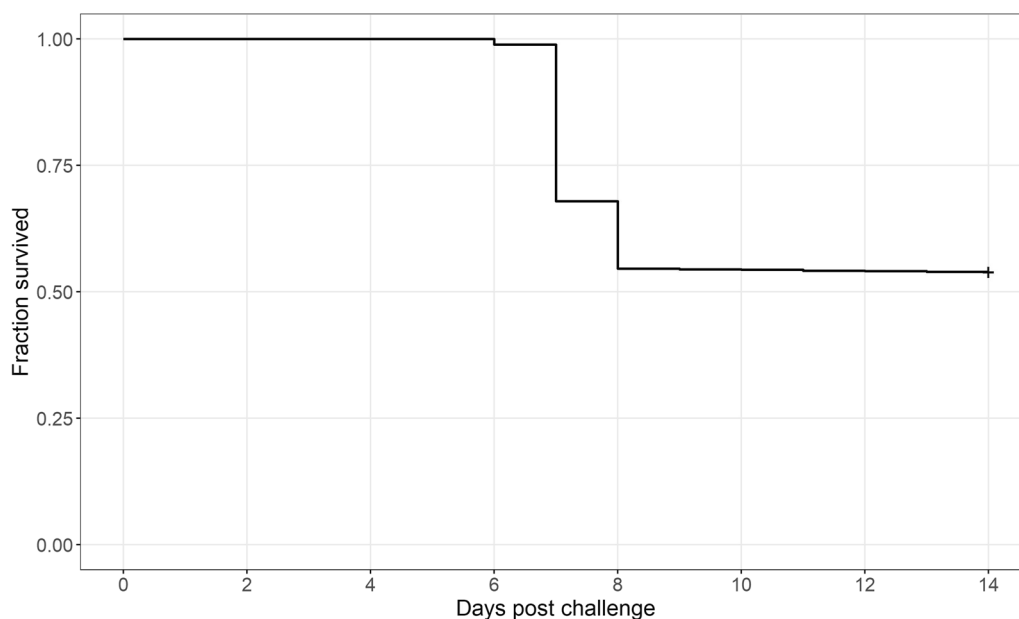


Fig. 1 Kaplan–Meier product-limit curve for survival of fish during the challenge test with NNV

post-challenge. At the end of the experiment, 663 animals were classified as survivors and were sacrificed (overall survival rate: 53.9%).

Additive genetic variance and pedigree-based heritability

The median of the marginal posterior density for the additive genetic (σ_a^2) and phenotypic variance (σ_p^2) for VNN symptomatology was equal to 0.263 and 1.263, respectively. Viral nervous necrosis symptomatology had a moderate heritability ($h_p^2 = 0.208$ on the underlying scale), with a probability of 57.2% for the estimate to be higher than the critical value of 0.2. The estimated 95% highest posterior density interval for heritability was narrow, with values between 0.077 and 0.366.

Genomic variance and genomic heritability

Point estimates and features of the marginal posterior densities for variance components and genomic heritability for VNN symptomatology are in Table 1. The estimates provided by the three Bayesian models were similar; Bayes C provided the highest heritability estimate

($h_g^2 = 0.195$), while the estimates provided by Bayes B and Bayes Ridge Regression were 1.03 and 0.7 percentage points lower, respectively. The estimated genomic heritability averaged over models ($h_g^2 = 0.188$) was similar to the pedigree-based heritability, and the probability for the estimate to be higher than 0.2 ranged from 38.6% (Bayes B) to 50.2% (Bayes C). The width of the HPD95% intervals for the estimated genomic heritability was approximately half that for h_p^2 , which indicates that the uncertainty in the estimation process of the genomic heritability was lower.

Genome-wide association study

In the GWAS, the genome-wide significance threshold was set to 5.73 (after Bonferroni correction and $-\log_{10}$ transformation). The GWAS did not identify any SNP exceeding the genome-wide significance level (Fig. 2); SNPs located on linkage groups 22 and 23 are those that are closest to the significance level.

Table 1 Estimates of the genomic variance (σ_g^2) and heritability (h_g^2) for VNN symptomatology, 95% highest posterior density interval (HPD95%) for h_g^2 and probability for h_g^2 to be greater than 0.2 obtained with three Bayesian regression models

Model	σ_g^2	h_g^2	HPD95%	P ($h_g^2 \geq 0.2$)
Bayes B	0.222	0.182	0.106, 0.262	0.386
Bayes C	0.242	0.195	0.118, 0.283	0.502
Bayes Ridge Regression	0.232	0.188	0.120, 0.259	0.434

Average genomic relatedness within and between training/testing sets

In random CV, the dataset was split into four equally-sized subsets (Table 2). The genomic relatedness both within group and between one group and the others was comparable across groups. K-means clustering produced four groups composed of an unbalanced number of animals (Table 2). The genomic relatedness within each group was maximized (values ranging from 0.04 to 0.26), whereas the relatedness between one group and

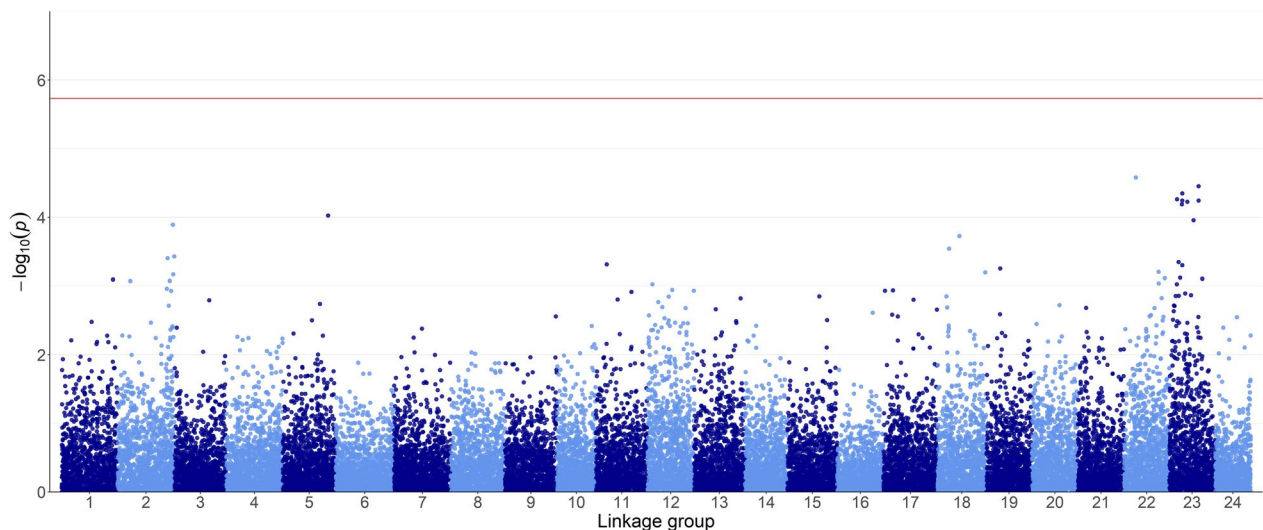


Fig. 2 Genome-wide association plots between genotypes at 26,591 SNPs and VNN symptomatology. The red line indicates the genome-wide significance threshold $p\text{-value} = 0.05/p$ (where p is the total number of SNPs) after Bonferroni correction and $-\log_{10}$ transformation (5.73)

Table 2 Number of animals (N) in each group, average genomic relatedness (standard deviation SD) within group ($g_{ij,w}$) and average genomic relatedness (SD) between one group and the others ($g_{ij,b}$), for each cross-validation strategy

Cross-validation strategy	Group	N	$g_{ij,w}$	$g_{ij,b}$
Random	1	243	- 0.001 (0.113)	- 0.001 (0.114)
	2	243	- 0.001 (0.116)	- 0.001 (0.115)
	3	243	- 0.001 (0.114)	- 0.001 (0.115)
	4	243	- 0.001 (0.118)	- 0.001 (0.116)
K-means clustering	1	336	0.125 (0.099)	- 0.064 (0.070)
	2	177	0.239 (0.114)	- 0.054 (0.073)
	3	326	0.039 (0.100)	- 0.024 (0.069)
	4	133	0.258 (0.098)	- 0.043 (0.077)
Leave-one-family-out	Parents	54	0.015 (0.088)	- 0.002 (0.109)

For the leave-one-family-out validation, $g_{ij,w}$ refers to the group of parents, whereas $g_{ij,b}$ refers to the parents and the challenged fish

Table 3 Accuracy of prediction of EBV for VNN symptomatology and β coefficient provided by the Bayesian models in three cross-validation strategies (random, K-means clustering and leave-one-family-out focused on the parents of the challenged fish)

Cross-validation strategy	Model	Accuracy	β
Random	Bayes B	0.903	1.010
	Bayes C	0.903	1.010
	Bayes Ridge Regression	0.903	1.010
K-means clustering	Bayes B	0.525	1.935
	Bayes C	0.530	1.956
	Bayes Ridge Regression	0.529	1.952
Leave-one-family-out	Bayes B	0.119	0.587
	Bayes C	0.126	0.620
	Bayes Ridge Regression	0.127	0.625

EBV breeding value estimated using the phenotypic information of the animal and of its full- and half-sibs

Accuracy Pearson product-moment correlation between EBV and genomic-predicted EBV

β slope of the regression of the observed on the predicted EBV

the others was minimized (values ranging from - 0.06 to - 0.02; Table 2). The average genomic relatedness within the group of parents was 0.015, whereas between challenged fish and the population of parents was -0.002.

Model performance for genomic prediction of the estimated breeding values

In the random CV procedure, genomic predictions of the EBV for VNN symptomatology were highly accurate and unbiased, with a Pearson product-moment correlation equal to 0.90 and β coefficients (slope of the regression of the observed on the predicted EBV) equal to 1.01 (Table 3). The comparison between Bayesian genomic models showed that their performance for prediction was similar. The accuracy of the prediction of the true additive genetic value, which depends on the accuracy of EBV and on the accuracy of genomic-predicted EBV, ranged from 0.44 to 0.51 and was, on average, 0.46.

The prediction accuracy decreased to 0.53 when the genomic relatedness between training and testing sets was minimized (K-means clustering) and dropped remarkably (0.12) when the most penalizing validation (LOFO), which is able to break down the family linkage disequilibrium due to the experimental design, was applied (Table 3).

The three Bayesian models exhibited similar prediction accuracies and similar β coefficients, both in the K-means clustering validation and in the LOFO validation (Table 3).

Model performance for the genomic classification of VNN symptomatology

When used as classifiers of VNN symptomatology, breeding values estimated from the observed phenotype of the individual and those of its full- and half-sibs (EBV) exhibited an AUC of 0.905, an ACC of 0.864 and a MCC of 0.730. The classification of the observed symptomatology based on the genomic-predicted phenotype resulted

Table 4 Average metrics (SD) for classification of VNN symptomatology based on genomic predictions of phenotype or of breeding values as classifiers

Classifier	Method	Metric		
		AUC	ACC	MCC
Genomic-predicted phenotype	Bayes B	0.581 (0.017)	0.569 (0.016)	0.175 (0.024)
	Bayes C	0.597 (0.013)	0.581 (0.013)	0.176 (0.023)
	Bayes Ridge Regression	0.601 (0.013)	0.583 (0.009)	0.178 (0.022)
Genomic-predicted EBV	Bayes B	0.657 (0.002)	0.617 (0.006)	0.237 (0.008)
	Bayes C	0.657 (0.002)	0.617 (0.006)	0.237 (0.008)
	Bayes Ridge Regression	0.657 (0.002)	0.616 (0.005)	0.236 (0.008)

EBV breeding values estimated using the phenotypic information of the animal and of its full- and half-sibs

AUC area under the ROC curve, ACC accuracy as (true positives + true negatives)/number of samples, MCC Matthews correlation coefficient

in moderate performances for all methods, showing AUC between 0.581 and 0.601, ACC between 0.569 and 0.583, and MCC between 0.175 and 0.178 (Table 4). The genomic-predicted EBV used as a classifier provided an added value, increasing the metrics to 0.657 (AUC), 0.617 (ACC) and 0.237 (MCC). The results were consistent across Bayesian models, with slightly greater metrics for Ridge Regression.

Discussion

The recently discovered susceptibility of gilthead sea bream to a NNV reassortant strain has posed a new and unexpected threat to the sea bream Mediterranean aquaculture [9, 10]. Selective breeding for enhanced genetic resistance of sea bream against VNN aims at protecting larval stages, whose immune system is still in development [12], and may exploit evidence on the same pathology obtained for European sea bass, Asian sea bass and Atlantic cod [13–15] or on other diseases for other species [17]. Incorporation of genomic information in selection procedures is of great interest for complex traits such as disease resistance and is encouraged by the development of high-density marker arrays [19]. To date, in sea bream, the genetic basis of VNN resistance has not been investigated yet. Moreover, studies based on experimental infections in early developmental stages are scarce, particularly those with samples of size suitable to estimate variance components and genetic parameters.

In this study, the experimental trial was conducted on very young fish (21 d post-hatching) and of small size compared to previous studies (150–200 d post-hatching, 10–20 g) [45–48]. In spite of the young age and small size of the fish, clinical signs were clearly recognizable and guaranteed a reliable recording of the symptomatology. As reported in case studies focused on natural outbreaks in sea bream hatcheries, the clinical signs preceding the death in larvae aged 17–35 d post-hatching were, as commonly described for VNN, lethargy, apathy, anorexia, abnormal swimming, loss of balance and swim bladder hyperinflation [8–10].

Due to the small size of the fish, the experimental infection was carried out by immersion which, when compared to infection by injection, leads to outcomes which depend on the same defence mechanisms as those activated in natural disease outbreaks. The limited size of the experimental tank guaranteed an equal level of exposure to the virus for all animals, which is important when assessment of across-family differences in resistance is the key aspect of the study.

The mortality rate in our study (46.1%) was comparable to that reported in sea bream of similar age experiencing natural infections (40%) [10]. Previous studies on natural outbreaks reported significantly higher mortality rates

(>80%) [8, 9], which may result from different environmental conditions (water temperature and density) and from inherent genetic differences across populations in terms of VNN resistance as detected, for example, in sea bass [46].

Frequently, investigations on the genetic variation in disease resistance are focused only on mortality assessed in challenge tests, where fish are classified as dead or survived after a controlled viral, bacterial or parasitic infection. In future studies, it could be worth disentangling the genetic variation of the different mechanisms implied in survival, as already performed in pacific white shrimp (*Penaeus vannamei*) for susceptibility and endurance to taura syndrome [49] and in Atlantic cod (*Gadus morhua*) for susceptibility and endurance to VNN and vibriosis [45]. Although the traits under consideration were characterized by different pedigree-based heritability estimates ($h_p^2 = 0.33$ – 0.91 for susceptibility and $h_p^2 = 0.02$ – 0.12 for endurance), those studies provided encouraging results that could motivate further investigations on different traits which have a role in survival, rather than focusing only on mortality.

The heritability estimated for VNN symptomatology ($h_p^2 = 0.21$ on the underlying scale) was in the range of estimates obtained for VNN resistance in European sea bass (0.14–0.38) [13, 46–48, 50] and Asian sea bass (*Lates calcarifer*; 0.25) [14], but lower than estimates obtained for VNN resistance in Atlantic cod (0.43–0.91) [15, 45, 51, 52]. It must be noted that all the cited experiments were performed on fish of different size and age compared to our experimental animals and, with the exception of the study of Palaiokostas et al. [47], they also used a different method of infection (injection), which may lead to differences in the estimated heritability.

The genomic heritability estimate ($h_g^2 = 0.19$) was consistent with the pedigree-based heritability estimate, which indicates that SNP effects could pick up all the variation due to additive genetic effects. In addition, the HPD95% interval for the genomic heritability was approximately half as large as that of the pedigree-based heritability (0.115–0.268 vs. 0.077–0.366), which indicates that h_g^2 was estimated with less uncertainty and that genomic inference is more accurate than the pedigree-based one. The genomic heritability estimate is anyway lower than that obtained by Palaiokostas et al. [47] and Vela-Avitúa et al. [48] in sea bass using genomic information (0.40–0.43), but, since in those studies the experimental animals had undergone several full cycles of selection, the linkage disequilibrium derived from selective breeding could be a reason for such an overrating of the additive genetic variance.

The weak statistical power of single-marker GWAS [53] led, in our case, to unsatisfactory results in terms of detection of genomic regions significantly associated with VNN resistance in gilthead sea bream. One genomic

region located in LG23 was suggested to be involved in sea bream VNN resistance, as it is the region closest to the significance threshold. Likewise, in European sea bass a linkage group explaining 9.2% of the variance was associated with VNN resistance [48, 50]. In our study, the presence of three minor genomic regions explaining 2.6 to 2.8% of the genetic variance reinforces once again the hypothesis that disease resistance traits are complex and polygenic. Overall, the total variance explained by the top four windows was 11.73%.

At first glance, the obtained genomic prediction values seem to be encouraging in view of the practical implementation of genomic selection in sea bream to enhance the resistance to the NNV reassortant strain, but the CV strategy may affect estimates of prediction accuracy; in particular, random k-fold CV performed on close relatives from a single generation may lead to overly optimistic results, as already demonstrated by different authors both in livestock [21, 22] and aquaculture species [23–25]. The availability of data from only one generation is due to the complexity of the experiments required to record the disease resistance traits, which prevents routine phenotyping, and to the fact that in aquaculture species, unlike terrestrial species, breeding programs for disease resistance (and to complex traits in general) have only recently started and are not widely applied yet [54].

In our study, the accuracy of genomic prediction of EBV provided by random fourfold CV was high ($r=0.90$), but, since this strategy may not be informative in the long-term due to the disruption of family-based linkage disequilibrium, two other validation strategies (K-means clustering and LOFO) were tested to mimic more realistic scenarios where the genomic ties between animals in the training and testing sets due to close relationships, as those occurring in full- and half-sib families, were weakened. As expected, both K-means clustering and LOFO accuracies decreased in comparison with that estimated in random fourfold CV. Although in the K-means clustering strategy the accuracy remained moderate, suggesting that genomic selection could be feasible in a mid-term view, in the LOFO validation strategy the inflection was marked ($r=0.12$). Such differences across validation strategies may also be ascribed to the use of the EBV as pseudo-phenotypes. Due to the mating design and the moderate heritability of VNN symptomatology, EBV include a family component, which, in our case, represents the largest proportion of the EBV variance. The genomic prediction accuracies obtained in the CV procedures, where the animals included in the training and testing sets were randomly chosen, are expected to be high because many sibs are in both validation and training sets and family means are predicted with family information. When family data are not considered in the validation procedure, as in LOFO

and, to a lower extent, in K-means clustering, the accuracy of the genomic prediction is expected to be low. Even with individual phenotypes, the difference between validation strategies would have been observed, but it would have been more limited.

Anyway, in Atlantic salmon, Frasin et al. [25] evidenced how a systematic reduction of the relatedness between the training and testing sets resulted in decreased genomic prediction accuracies, compared to that of the random CV. When the genomic relationship between the training and testing populations was set to an extreme value of 0.3, the accuracy for sea lice resistance and body weight decreased up to -44% and -51%, respectively. In our study, the scenario was even more extreme, as the parent contribution to the challenged sample was rather unbalanced, with three dams out of 22 producing almost 67% of the individuals, and six sires out of 28 producing nearly 48% of the individuals. In the K-means and LOFO validation procedures, this resulted in weak relationships between the parent used as a test set and the animals in the training set. de Los Campos et al. [55] demonstrated how, when the data involve unrelated individuals, linkage disequilibrium spans over shorter regions and within-family disequilibrium cannot be exploited. This implies an imperfect linkage disequilibrium between markers and causal loci, which, in turn, results in very low prediction accuracy of genomic best linear unbiased prediction (GBLUP). Estimated breeding values for VNN symptomatology was the best classifier of the observed trait (AUC=0.91), which was expected since the individual phenotypic information was included in the data used to estimate the breeding value. The classification of the observed VNN symptomatology based on the genomic prediction of the phenotype exhibited an AUC equal to 0.59, whereas genomic prediction of EBV provided a 12% increase of the classification performance (AUC=0.66). The AUC metric for classification based on the genomic predictions of the phenotype was in the range of values reported for other species challenged with viral diseases (0.52 to 0.70 for VNN in sea bass, [47, 50]; 0.74 for koi herpesvirus disease in common carp, [56]).

The fish used in this study were produced from a breeding stock that has never undergone a selective breeding program directed to the enhancement of resistance to VNN; in this context, several steps are specific to the planning of an efficient and long-term view breeding program: (1) the creation of a breeding nucleus which develops the elite broodstock, (2) the design of breeding strategies to produce both full-sib and half-sib families, (3) the development of a high-depth pedigree database including all the information of breeding candidates and tested fish, and (4) the exploitation of genomic information, which can be combined to the aforementioned information to accurately select the breeding candidates.

Conclusions

The estimated heritability for VNN mortality indicates that selective breeding programmes are feasible for increasing resistance to VNN of fish larvae/juveniles, which would overcome the problem of vaccination. Traditional selective breeding approaches for VNN resistance are expected to have a limited efficiency, mainly due to the low accuracy of the estimated breeding values of the breeding candidates which, for sanitary reasons, cannot be included in VNN challenge tests. The results of the GWAS suggest a complex polygenic nature for VNN resistance with many genomic regions each with a small effect on the phenotypic variation of the trait. In view of the implementation of genome-enabled selection in sea bream to enhance resistance to the NNV reassortant strain, the accuracies of the genomic predictions were high when assessed in CV. However, the application of validation procedures, which can cancel the favourable effects of family-based linkage disequilibrium that are determined by the experimental design, decreased dramatically the estimated accuracies. This raises concerns on the use of factorial mating design paired with CV as a tool to assess genomic prediction accuracies and emphasizes the need of the periodical update of the training set with new data to keep prediction accuracies at a satisfactory level in the case of populations where, like the one investigated in this study, genomic relationships across families are weak. Further studies should focus on disentangling the genetic variation in the different mechanisms that have a role in VNN survival, such as susceptibility and endurance, rather than recording only mortality of experimentally-infected animals.

Acknowledgements

The authors acknowledge Panittica Italia Società Agricola S.R.L. for providing the animals used in this experiment. The authors would like to thank the Editor and the Reviewers for their useful comments that helped to improve the manuscript.

Author contributions

SF performed the statistical analyses, estimation of genetic and genomic variance components, prediction of breeding values, genomic predictions and validations, participated in data interpretation and prepared the manuscript (drafting and editing); PC conceived and supported the analyses, contributed to the interpretation of results and critically revised the manuscript; RF, GDR and FP performed the challenge test and collected the samples; RF and GDR performed the gDNA extraction; MB performed the bioinformatic analyses; MC and HC produced the experimental fish; AT supervised the challenge test; LB conceived the overall design, undertook the project management, contributed to the interpretation of results and critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement n. 727610 (PerformFISH).

Availability of data and materials

Data are available from the authors upon reasonable request.

Declarations

Ethics approval and consent to participate

The experimental infection protocol was evaluated by the Istituto Zooprofilattico Sperimentale delle Venezie (IZSVE) Animal Welfare Body and Ethics Committee (Opinion CE.IZSVE.3/2016 of 24/10/216) and approved by the Italian Ministry of Health (Law decree 101/2017-PR of 02/02/2017).

Consent for publication

Not applicable.

Competing interests

MC and HC are employees of Panittica Italia Società Agricola S.R.L., a hatchery producing European sea bass and sea bream. The other authors declare that they have no competing interests.

Author details

¹Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Università, 16, 35020 Legnaro, PD, Italy. ²Division of Comparative Biomedical Sciences, OIE Reference Centre for Viral Encephalopathy and Retinopathy, Istituto Zooprofilattico Sperimentale delle Venezie (IZSVE), Padova, Italy. ³Panittica Italia Società Agricola S.R.L., Strada del Procaccio, 72016 Torre Canne di Fasano, Italy.

Received: 8 August 2022 Accepted: 21 March 2023

Published online: 03 April 2023

References

- Bovo G, Nishizawa T, Maltese C, Borghesan F, Mutinelli F, Montesi F, et al. Viral encephalopathy and retinopathy of farmed marine fish species in Italy. *Virus Res.* 1999;63:143–6.
- Bitchava K, Chassalevris T, Lampou E, Athanassopoulou F, Economou V, Dovas Cl. Occurrence and molecular characterization of betanodaviruses in fish and invertebrates of the Greek territorial waters. *J Fish Dis.* 2019;42:1773–83.
- Mori Kl, Nakai T, Muroga K, Arimoto M, Mushiaki K, Furusawa I. Properties of a new virus belonging to nodaviridae found in larval striped jack (*Pseudocaranx dentex*) with nervous necrosis. *Virology.* 1992;187:368–71.
- Munday BL, Nakai T. Nodaviruses as pathogens in larval and juvenile marine finfish. *World J Microbiol Biotechnol.* 1997;13:375.
- Yong CY, Yeap SK, Omar AR, Tan WS. Advances in the study of nodavirus. *PeerJ.* 2017;5: e3841.
- Castric J, Thiery R, Jeffroy J, de Kinkelin P, Raymond JC. Sea bream *Sparus aurata*, an asymptomatic contagious fish host for nodavirus. *Dis Aquat Org.* 2001;47:33–8.
- Valero Y, García-Alcázar A, Ángeles Esteban M, Cuesta A, Chaves-Pozo E. Antimicrobial response is increased in the testis of European sea bass, but not in gilthead seabream, upon nodavirus infection. *Fish Shellfish Immunol.* 2015;44:203–13.
- Toffan A, Pascoli F, Pretto T, Panzarin V, Abbadi M, Buratin A, et al. Viral nervous necrosis in gilthead sea bream (*Sparus aurata*) caused by reassortant betanodavirus RGNNV/SJNNV: an emerging threat for Mediterranean aquaculture. *Sci Rep.* 2017;7:46755.
- Volpe E, Gustinelli A, Caffara M, Errani F, Quaglio F, Fioravanti ML, et al. Viral nervous necrosis outbreaks caused by the RGNNV/SJNNV reassortant betanodavirus in gilthead sea bream (*Sparus aurata*) and European sea bass (*Dicentrarchus labrax*). *Aquaculture.* 2020;523: 735155.
- Savoca S, Abbadi M, Toffan A, Salogni C, Iaria C, Capparucci F, et al. Betanodavirus infection associated with larval enteropathy as a cause of mortality in cultured gilthead sea bream (*Sparus aurata*, Linnaeus, 1758). *Aquaculture.* 2021;541: 736844.
- NaveenKumar S, Hassan MA, Mahmoud MA, Al-Ansari A, Al-Shward WK. Betanodavirus infection in reared marine fishes along the Arabian Gulf. *Aquacult Int.* 2017;25:1543–4.
- Jósefsson S, Tatner MF. Histogenesis of the lymphoid organs in sea bream (*Sparus aurata* L.). *Fish Shellfish Immunol.* 1993;3:35–49.
- Faggion S, Bertotto D, Babbucci M, Dalla Rovere G, Franch R, Bovolenta M, et al. Resistance to viral nervous necrosis in European sea bass

- (*Dicentrarchus labrax* L.): heritability and relationships with body weight, cortisol concentration, and antibody titer. *Genet Sel Evol.* 2021;53:32.
14. Wang L, Liu P, Huang S, Ye B, Chua E, Wan Z, et al. Genome-wide association study identifies loci associated with resistance to viral nervous necrosis disease in Asian seabass. *Mar Biotechnol (NY)*. 2017;19:255–65.
 15. Ødegård J, Sommer AI, Præbel AK. Heritability of resistance to viral nervous necrosis in Atlantic cod (*Gadus morhua* L.). *Aquaculture*. 2010;300:59–64.
 16. Gjedrem T. Disease resistant fish and shellfish are within reach: a review. *J Mar Sci Eng.* 2015;3:146–53.
 17. Ødegård J, Baranski M, Gjerde B, Gjedrem T. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquac Res.* 2011;42:103–14.
 18. Yáñez JM, Houston RD, Newman S. Genetics and genomics of disease resistance in salmonid species. *Front Genet.* 2014;5:415.
 19. Peñaloza C, Manousaki T, Franch R, Tsakogiannis A, Sonesson AK, Aslam ML, et al. Development and testing of a combined species SNP array for the European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*). *Genomics.* 2021;113:2096–107.
 20. Zenger KR, Khatkar MS, Jones DB, Khaliliani N, Jerry DR, Raadsma HW. Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front Genet.* 2019;9:693.
 21. Clark SA, Hickey JM, Daetwyler HD, van der Werf JH. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet Sel Evol.* 2012;44:4.
 22. Pérez-Cabal MA, Vazquez A, Gianola D, Rosa G, Weigel K. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. *Front Genet.* 2012;3:27.
 23. Tsai HY, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB, et al. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet Sel Evol.* 2016;48:47.
 24. Faggion S, Bertotto D, Bonfatti V, Freguglia M, Bargelloni L, Carnier P. Genomic predictions of phenotypes and pseudo-phenotypes for viral nervous necrosis resistance, cortisol concentration, antibody titer and body weight in European sea bass. *Animals (Basel)*. 2022;12:367.
 25. Frasin C, Yáñez JM, Robledo D, Houston RD. The impact of genetic relationship between training and validation populations on genomic prediction accuracy in Atlantic salmon. *Aquac Rep.* 2022;23: 101033.
 26. Pszczola M, Strabel T, Mulder HA, Calus M. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci.* 2012;95:389–400.
 27. Silva RMO, Fragomeni BO, Lourenco DAL, Magalhães AFB, Irano N, Carvalho R, et al. Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. *J Anim Sci.* 2016;94:3613–23.
 28. Pardo BG, Machordom A, Foresti F, Porto-Foresti F, Azevedo MF, Bañon R, et al. Phylogenetic analysis of flatfish (order *Pleuronectiformes*) based on mitochondrial 16S rDNA sequences. *Sci Mar.* 2005;69:531–43.
 29. Sargolzaei M, Chesnais JP, Shenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* 2014;15:478.
 30. Griot R, Allal F, Brard-Fudulea S, Morvezen R, Haffray P, Phocas F, et al. APIS: An auto-adaptive parentage inference software that tolerates missing parents. *Mol Ecol Resour.* 2020;20:579–90.
 31. Therneau T. A package for survival analysis in R. R package version 3.2–13. 2021. <https://CRAN.R-project.org/package=survival>. Accessed Mar 2021.
 32. Legarra A, Varona L, López de Maturana E. TM Threshold model. 2008. http://genoweb.toulouse.inra.fr/~alegarratm_folder/manualtm.pdf. Accessed Apr 2021.
 33. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In: Bernardo JM, Berger JO, Dawid AP, Smith AF, editors. *Bayesian statistics*. Oxford: Oxford University Press; 1992. p. 169–93.
 34. Smith BJ. boa: an R package for MCMC output convergence assessment and posterior inference. *J Stat Softw.* 2007;21:1–37.
 35. Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* 2014;198:483–95.
 36. Campos G de los, Sorensen D, Gianola D. Genomic heritability: what is it? *PLoS Genet.* 2015;11:e1005048.
 37. Habier D, Fernando R, Kizilkaya K, Garrick D. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics.* 2011;12:186.
 38. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157:1819–29.
 39. Perdry H, Dandine-Roulland C, Bandyopadhyay D, Kettner L. Package 'gaston': genetic data handling (QC, GRM, LD, PCA) and linear mixed models. Version 1.5.3 <https://CRAN.R-project.org/package=gaston>. Accessed 12 Oct 2022.
 40. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. <https://ggplot2.tidyverse.org>. Accessed Jun 2021.
 41. Legarra A, Reverter T. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Sel Evol.* 2018;50:53.
 42. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2012.
 43. Hartigan JA, Wong MA. A k-means clustering algorithm. *J R Stat Soc C Appl Stat.* 1979;28:100–8.
 44. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr, visualizing classifier performance in R. *Bioinformatics.* 2005;21:3940–1.
 45. Bangera R, Ødegård J, Nielsen H, Gjølven H, Mortensen A. Genetic analysis of vibriosis and viral nervous necrosis resistance in Atlantic cod (*Gadus morhua* L.) using a cure model. *J Anim Sci.* 2013;91:3574–82.
 46. Doan QK, Vandeputte M, Chatain B, Haffray P, Vergnet A, Breuil G, et al. Genetic variation of resistance to Viral Nervous Necrosis and genetic correlations with production traits in wild populations of the European sea bass (*Dicentrarchus labrax*). *Aquaculture.* 2017;478:1–8.
 47. Palaiokostas C, Cariou S, Bestin A, Bruant J-S, Haffray P, Morin T, et al. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet Sel Evol.* 2018;50:30.
 48. Vela-Avitúa S, Thorland I, Bakopoulos V, Papanna K, Dimitroglou A, Kotaras E, et al. Genetic basis for resistance against viral nervous necrosis: GWAS and potential of genomic prediction explored in farmed European sea bass (*Dicentrarchus labrax*). *Front Genet.* 2022;13: 804584.
 49. Ødegård J, Gitterle T, Madsen P, Meuwissen T, Yazdi M, Gjerde B, et al. Quantitative genetics of taura syndrome resistance in Pacific white shrimp (*Penaeus vannamei*): a cure model approach. *Genet Sel Evol.* 2011;43:14.
 50. Griot R, Allal F, Phocas F, Brard-Fudulea S, Morvezen R, Bestin A, et al. Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. *Aquaculture.* 2021;530: 735930.
 51. Bangera R, Ødegård J, Præbel A, Mortensen A, Nielsen H. Genetic correlations between growth rate and resistance to vibriosis and viral nervous necrosis in Atlantic cod (*Gadus morhua* L.). *Aquaculture.* 2011;317:67–73.
 52. Bangera R, Baranski M, Lien S. A genome-wide association study for resistance to viral nervous necrosis in Atlantic cod using a 12K single nucleotide polymorphism array. In: *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production: 17–22 August 2014; Vancouver*. https://www.asas.org/docs/default-source/wcgalp-proceedings-oral/270_paper_9537_manuscript_847_0.pdf?sfvrsn=2. Accessed 26 Nov 2021.
 53. Wang H, Misztal I, Aguilar I, Legarra A, Muir W. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)*. 2012;94:73–83.
 54. Palaiokostas C. Predicting for disease resistance in aquaculture species using machine learning models. *Aquac Rep.* 2021;20: 100660.
 55. de Los CG, 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.
 56. Palaiokostas C, Vesely T, Kocour M, Prchal M, Pokorova D, Piackova V, et al. Optimizing genomic prediction of host resistance to koi herpesvirus disease in carp. *Front Genet.* 2019;10:543.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.