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

Ronan Griot, François Allal, Florence Phocas, Sophie Brard-Fudulea, Romain Morvezen, Anastasia Bestin, Pierrick Haffray, Yoannah François, Thierry Morin, Charles Poncet, 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

Griot R.[a,b,c], Allal F.[c], Phocas F.[b], Brard-Fudulea S.[d], Morvezen R.[a], Bestin A.[a], Haffray P.[a], François Y.[a] Morin T.[f], Poncet C.[e], Vergnet A.[c], Cariou S.[g], Brunier J.[h], Bruant J.S.[g], Peyrou B.[h], Gagnaire P.A.[i], Vandeputte M.[b,c]

[a] SYSAAF, Station LPGP/INRAE, Campus de Beaulieu, 35042 Rennes, France

[b] Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

[c] MARBEC, Univ. Montpellier, Ifremer, CNRS, IRD, 34250 Palavas-les-Flots, France

[d] SYSAAF, UMR BOA, Centre INRAE Val de Loire, 37380 Nouzilly, France

[e] INRAE-UCA, UMR GDEC, Clermont-Ferrand, France

[f] ANSES, Ploufragan-Plouzané-Niort Laboratory, Viral Fish Diseases Unit, National Reference Laboratory for Regulated Fish Diseases, Technopôle Brest-Iroise, 29280 Plouzané, France

[g] Ferme Marine Du Douhet, 17840 La Brée Les Bains, France

[h] Ecloserie Marine de Gravelines-Ichtus, 59273 Gravelines, France

[i] ISEM, Université de Montpellier, CNRS-IRD-EPHE-UM, France

Corresponding author: Marc Vandeputte, e-mail: **marc.vandeputte@inrae.fr**

Abstract

Viral Nervous Necrosis (VNN) is a major threat for the European sea bass (*Dicentrarchus labrax*) aquaculture industry. The improvement of disease resistance through selective breeding is a promising option to reduce outbreaks. With the development of high-throughput genotyping technologies, identification of genomic regions involved in the resistance could improve the efficiency of selective breeding. The aim of this study was to identify quantitative trait loci (QTL) involved in VNN resistance and to quantify their effect.

Four experimental backcross families comprising 378, 454, 291 and 211 individuals and two commercial populations A and B comprising 1027 and 1042 individuals obtained from partial factorial crosses (59♂ x 20♀ for pop A; 39♂ x 14♀ for pop B) were submitted to a redspotted grouper nervous necrosis virus (RGNNV) challenge by bath. A high-density single nucleotide polymorphism (SNP) chip panel was designed to develop the ThermoFisher Axiom™ 57k SNP DlabChip, which was used for genotyping all individuals and building a high quality linkage map. In the backcross families, composite interval mapping was performed on 30917, 23592, 30656 and 31490 markers, respectively. In the commercial populations, 40263 markers in pop A and 41166 markers in pop B were used to perform genome-wide association studies (GWAS) using a GBLUP and a BayesCπ approach.

One QTL was identified on chromosome LG12 in three of the four experimental backcross families, and one additional QTL on LG8 was detected in only one family. In commercial populations, QTL mapping revealed a total of seven QTLs, among which the previously mentioned QTL on LG12 was detected in both. This QTL, which was mapped to an interval of 3.45 cM, explained 9.21% of the total genetic variance in pop A, while other identified QTLs individually explained less than 1% of the total genetic variance.

The identification of QTL regions involved in VNN resistance in European sea bass, with one having a strong effect, should have a great impact on the aquaculture industry. Future work could focus on the fine mapping of the causal mutation present on LG12 using whole genome sequencing.

Keywords

Disease resistance, VNN, Fish, SNP array, QTL, GWAS, Linkage map

1. Introduction

European sea bass (*Dicentrarchus labrax*) is a major species for Mediterranean aquaculture, with a production of more than 150,000 tons in 2016 (FEAP, 2017). A major threat for the sea bass aquaculture industry is the occurrence of disease outbreaks, and especially of Viral Nervous Necrosis (VNN) disease mostly in warmer growing Mediterranean areas (Breuil et al., 1991; Le Breton et al., 1997; Vendramin et al., 2016). The VNN disease is caused by the Nervous Necrosis Virus (NNV), a RNA virus belonging to the Betanodavirus genus (Mori et al., 1992). NNV is widespread worldwide and can infect more than 70 marine and freshwater species (Doan et al., 2017b). In sea bass production, mortalities up to 90% have been recorded especially at the very susceptible larval and juvenile stages (Le Breton et al., 1997). Several vaccine strategies have been tested, leading to an improvement of survival (Doan et al., 2017b). However, three points limit the use of vaccines in VNN outbreak management: i) vaccines are generally expensive, ii) the number of fish to vaccinate is high which increases the cost issue (Ulmer et al., 2006), and iii) VNN is more infectious in early life stages, while vaccination of small fish is challenging (Gomez-Casado et al., 2011; Sommerset et al., 2005).

One promising solution to reduce the effects of VNN outbreaks in aquaculture farms is selective breeding for improved resistance to VNN. Selective breeding for disease resistance in aquaculture species has demonstrated its interest with a genetic gain in survival of 12.5% per generation on average, across different host-pathogen pairs (Gjedrem and Robinson, 2014). For selective breeding to be successful, enough genetic variation for disease resistance within the species is however required. Moderate to high heritability has already been reported for resistance to VNN in European sea bass (from 0.26 to 0.43), confirming the interest of selective breeding for this trait (Doan et al., 2017a; Palaiokostas et al., 2018).

Genomic markers can be used to detect Quantitative Trait Loci (QTL), which are regions of the genome involved in the phenotypic variation of the trait (Georges et al., 1993). Their use allows a better understanding of the genetic architecture of complex traits as well as the location of the potential candidate genes involved in the resistance phenotype. With the access to high-throughput genotyping technologies at a reasonable cost for non-model species, some disease resistance QTLs have been found in a limited number on aquaculture species (see review by Ødegård et al., (2011)). In Atlantic salmon, major QTLs were discovered for IPNV resistance, explaining up to 83% of the genetic variance (Houston et al., 2008; Moen et al., 2009). These results can then be used to accurately choose broodstock among

the candidates for selection or commercial diffusion by Marker Assisted Selection (MAS) (Houston et al., 2008; Moen et al., 2009). In Asian sea bass (*Lates calcarifer*), QTL mapping revealed nine QTLs involved in VNN resistance, explaining between 1.6% and 2.7% of the phenotypic variance (Wang et al., 2017). In Atlantic cod (*Gadus morhua*), five QTLs were detected, including three QTL with major effects, explaining between 14.2% and 19.7% of the phenotypic variance (Baranski et al., 2010). For VNN resistance in sea bass, a recent study reported three first QTLs, each of them explaining a small part of the genetic variance (1.5% to 4%) (Palaiokostas et al., 2018).

In this study, we investigated the genetic architecture of VNN resistance in sea bass. We performed Genome Wide Association Studies (GWAS) based on two commercial cohorts and four experimental backcross families, all genotyped on the newly developed ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip. Two approaches were considered to detect associations between SNPs and resistance to VNN, interval mapping and a multi-marker Bayesian variable selection model to estimate the sharing of genetic variance explained by the QTLs, as well as their credibility intervals.

2. Material and Methods

2.1. Ethics approval

All infection challenges were carried out in accordance with the European guidelines (Directive 2010–63-EU) and the corresponding French legislation. Animal experiment procedures were approved by the ethics committee on animal experimentation COMETH ANSES/ENVA/UPC No. 16 and were authorized by the French Ministry of Higher Education, Research and Innovation under numbers 2017022816255366, 29/01/13-5 and 10/03/15-1.

2.2. Fish material

The animals challenged came from two commercial cohorts from the breeding programs of two different companies and four experimental backcross families. The global study workflow was presented in Table 1. The commercial cohorts were produced by artificial mating and further mentioned as pop A and pop B. Pop A (1680 individuals) was produced from a mating of 59 sires by 20 dams in four partial factorial subsets (15x5, 14x5, 15x5 and 15x5). Pop B (1737 individuals) was generated from 39 sires and 14 dams mated in six factorial subsets (6x3, 6x1, 6x3, 7x2, 7x3, 7x2). Four experimental backcross full-sib families totalizing 2500 individuals were also produced by *in vitro* fertilization at Ifremer (Palavas-les-Flots, France). In a previous study (Doan et al., 2017a), 60 sires from four different geographic origins (north-eastern, south-eastern and western Mediterranean Sea as well as Atlantic) were mated with nine west Mediterranean dams in a full factorial design (G0). Part of the offspring (G1) was challenged for VNN resistance and the estimated breeding values were estimated for the entire population, including the parents and the sibs of the challenged fish that were the candidates for selection. From these non-challenged candidates, four males were selected from families in which one parent was highly resistant and the other highly susceptible to VNN, and those were from three different geographic origins (north-eastern, south-eastern and western Mediterranean Sea). Those males were mated with four western Mediterranean females that we expected to be more susceptible due to their western-Mediterranean origin, in order to produce four full-sibs backcross families (G2). The G2 families were then named according to their grand-sire origin, NEM10, NEM12, SEM8 and WEM18 (NEM = north-east Mediterranean Sea, SEM = south-east Mediterranean Sea and WEM = western Mediterranean Sea).

[Insert Table 1]

2.3. VNN challenge

All fish were challenged to VNN at the SYSAAF-ANSES Fortior Genetics platform (ANSES, Plouzané, France). For all experiments, fish were maintained in filtered seawater at a temperature of 27°C ± 2°C in an open circuit. Pop A, pop B and backcross families infectious challenges were performed separately but in the same way. All the fish sent to the infection challenge were individually tagged. A total of 1680 from pop A (25g mean weight), 1737 from pop B (20g mean weight) and 2500 fish from the backcross families (8g mean weight) were received and acclimated for a minimum period of three weeks. Then, the whole batch of each population was split into a pre-test batch (120, 150 and 500 individuals for pop A, pop B and backcrosses, respectively), a challenge batch (1350, 1212 and 1719 individuals for pop A, pop B and backcrosses, respectively) and a negative control batch (120, 275 and 150 fish from the corresponding populations). For the pre-test and challenge batches contamination, the fish were immersed for 2 to 3 hours in a static bath of aerated seawater containing $1x10^5$ $TCID_{50}$/ml of the W80 strain of RGNNV (redspotted grouper nervous necrosis virus), that was previously produced on a striped snakehead SSN1 cell line (Thiéry et al., 2004). Then, the batch was split into two tanks, water circulation was restarted and the mortality was recorded during 16 days. The mortality rate was 44% and 51% in the pop A pre-test, 51% and 53% in the pop B pre-test and 22.6% and 23.7% for backcross families. For the infectious challenges, a protocol similar to the pre-test was applied on the challenge batches. The negative controls were immersed for three hours into static seawater containing sterile cell culture medium. The mortality was recorded each day during the challenge period that was 27 days for pop A, 42 days for pop B and 33 days for the backcross families. Bacteriologic and virologic analyses were performed before the pre-test and challenges as well as during the mortality peak to check the sanitary status of the fish. Virologic analyses were done by injecting a homogenized mixture of eye and brain sampled on random dead fish on SSN1 cells. Then, if cytopathic effects were observed, a virus identification was performed by immunofluorescence using anti-NNV antibodies. For bacteriological analyses, spleen and kidney are randomly sampled and used for bacterial cultures followed by MALDI-TOF identification.

2.4. Design of the Sea Bass 57k SNP DLabCHIP array

The design of the high-density SNP array was based on the selection of high-quality variants from a database of ~2.6 million SNPs identified through whole-genome resequencing of 8 parents-offspring trios generated by experimental crossing of wild sea bass in Duranton et al. (2018). From this database, we excluded rare variants using a minor allele count threshold of 4 over 16 diploid parental genomes, and only retained SNPs located more than 35bp away from another known variant in order to ensure a high probe specificity. The final set of variants were chosen to cover the whole genome (including ungrouped scaffolds), but with a variable density depending on the estimated local nucleotide diversity (π) reported by Tine et al. (2014). This strategy aimed at increasing the density of SNPs within chromosome regions displaying a higher recombination rate, in order to homogenize recombination distances between two consecutive markers on the map. A list of 57,907 selected makers was submitted to ThermoFisher to develop the Axiom™ Sea Bass 57k SNP DlabChip array.

2.5. Genotyping and Parentage assignment

All individuals were genotyped with the ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip at the genotyping platform Gentyane (INRAE, Clermont-Ferrand, France). A total of 1152 individuals were genotyped in each commercial cohort and 1536 fish in total were also genotyped in the experimental backcross families. In pop A and pop B, the genotyped individuals were randomly selected from the

challenged ones among the dead and the surviving individuals. In the backcross families, they were selected to have the same average mortality per family as the whole challenged family. SNP calling was done using ThermoFisher software AxiomAnalysisSuite™. Preliminary quality controls were applied with threshold values of 95% for SNP call rate and 90% for sample call rate. Parentage assignment was done using 1000 randomly sampled markers with the R package APIS (Griot et al., 2019) with a positive assignment error rate set to 1%.

2.6. Creation of the genetic map

A genetic map was constructed with LepMap3 (Rastas, 2017). Backcross families were merged with another Ifremer sea bass dataset composed of the ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip genotypes of 880 individuals from 94 sires mated with 39 dams in three partial factorial designs. The resulting data set included 2232 individuals genotyped at 51179 markers.

We ran the recommended procedure of LepMap3 with custom settings: 1% segregation distortion for Filtering2, a LOD score of 50 as well as a subset of 25% of the markers for SeparateChromosomes2, and LOD score of 30 for JoinSingles2All.

2.7. Genotyping and sample quality control

For each commercial cohort, the SNPs retained for further analysis had a minor allele frequency (MAF) above 5% and a p-value for the Hardy-Weinberg equilibrium test above a threshold of 10^{-8}. From the remaining markers, we subset the ones mapped on the genetic map. After quality controls, 1089 individuals genotyped for 40623 markers were kept for pop A and 1110 individuals genotyped for 41166 markers for pop B. Prior to GWAS analysis, which is known to be highly sensitive to population structures (Hayes, 2013; Pritchard et al., 2000), we performed a principal component analysis (PCA) based on the SNP genotypes, on both commercial cohorts, using PLINK 1.9 (Purcell and Chang, 2015). A strong within population structure was observed for pop B, and consequently we discarded individuals that belonged to minor groups which were distant from the main one (Suppl. Fig. 1), leading to only 476 individuals retained for the genetic analysis of pop B. All individuals were kept in pop A.

In backcross families, we retained only SNPs with a MAF above 5% that were mapped on the genetic map. We obtained 30917, 23592, 30656 and 31490 informative markers for NEM10, NEM12, SEM8 and WEM18 respectively, with a sample size of 378 in NEM10, 454 in NEM12, 291 in SEM8 and 211 in WEM18 family.

Finally, within each cohort, the few missing genotype data were imputed using Fimpute software (Sargolzaei et al., 2014) to obtain complete genotypes for GWAS analysis.

2.8. Heritability estimation

For each data set, we estimated the heritability of VNN resistance under a threshold model using THRGIBBSF90 (Tsuruta and Misztal, 2006) and a linear model using AIREMLF90 (Misztal et al., 2002), both from the blupf90 program suite. Only individuals with a phenotype, a genotype and a pedigree were used in heritability estimates, thus the sample size was 1027 in pop A and 476 in pop B. Heritability was also estimated within each of the backcross families, as reliable estimation of heritability within a single family using genomic information was shown to be accurate by Ødegård and Meuwissen (2012).

The following model was computed in each cohort using both threshold and linear models:

$$y = 1b + Zu + e$$

With y the vector of the phenotypes measured as binary dead/survival trait, 1 the incidence (unity) vector of the intercept, b the estimate of the intercept effect, u the vector of breeding values and Z the corresponding incidence matrix. It is assumed that u follows a multivariate normal distribution N(0, $G\sigma^2_g$) with G the genomic relationship matrix proposed by VanRaden (2008) and σ^2_g is the additive genetic variance. e is the vector of the random residual errors that follows a normal distribution N(0, $I\sigma^2_e$) with σ^2_e the residual variance and *I* the identity matrix.

With the threshold model, the variance components (σ^2_g and σ^2_e) were estimated using a Gibbs sampler with 500,000 iterations, 100,000 of burn-in and one sample was kept every 20 iterations for posterior analysis. The posterior distributions were analyzed with the R package `boa` (Smith, 2007). With the linear model, the same components were estimated using a restricted maximum likelihood algorithm, considering the observed binary phenotype as a continuous variable.

The heritability for survival was estimated as:

$$h^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}$$

Heritability on the observed scale (h^2_o) was estimated using the variance components from the linear model, while the heritability on the underlying liability scale (h^2_u) was computed using the variance components from the threshold model.

2.9. QTL mapping in experimental crosses

In the experimental backcross families, we used a composite interval mapping approach from the R package `qtl` (Broman et al., 2003). The LOD score under the hypothesis of the absence of a QTL on the chromosome was computed for each interval between two consecutive markers of each chromosome using a Haley-Knott regression (Haley and Knott, 1992). Genome-wide significance LOD thresholds were estimated by permutation tests considering 1000 permutations (Churchill and Doerge, 1994).

2.10. QTL mapping in commercial cohorts

We performed GWAS under multi-marker linear regression models using GBLUP and Bayesian approaches.

2.10.1. GBLUP-based GWAS

We used the blupf90 suite of programs to perform GWAS by GBLUP analysis for VNN resistance in both commercial data sets. The breeding values were estimated with BLUPF90 using the following linear model:

$$y = 1b + Zu + e$$

With the same notation as in section 2.8. The p-values were computed using POSTGSF90 (Aguilar et al., 2019). The –log10 of the p-values were compared to the chromosome-wide significance threshold and to the genome-wide significance threshold at 5% after Bonferroni correction for the average number of markers per chromosome and the total number of markers, respectively.

2.10.2. Bayesian-based GWAS

Because GBLUP is known to shrink SNP effects towards 0, we also used a Bayesian variable selection model with a BayesCπ approach (Habier et al., 2011) to refine QTL positions as well as to estimate the proportions of genetic variance explained by the QTLs. In this model, a proportion π of the markers is assumed to have a non-zero effect. The marker effects are estimated with a mixture of a proportion π of markers with effects following a normal distribution N(0, σ^2_a) and a proportion 1 – π of markers with a

zero effect. σ^2_a is the part of total genetic variance explained by the SNP markers. In addition to the SNP effect described as above, we added a random polygenic genetic effect to account for the genetic variation that could not be captured by the markers (Legarra et al., 2008; Solberg et al., 2009). This model is defined as follows:

$$y = 1b + Zu + Wg + e$$

With y the vector of the phenotypes measured as binary dead/survival trait, 1 the incidence (unity) vector of the intercept, b the estimate of the intercept effect, u the vector of the polygenic effects, Z the corresponding incidence matrix, g the vector of the SNP random effects, W the corresponding incidence matrix and e the vector of the random residual errors. The total genetic variance is decomposed into one part explained by the polygenic effect and one part explained by the marker effects. As initial priors in pop A, the part of genetic variance explained by the polygenic effect represented 95% of the total genetic variance and the markers genetic variance represented 5% of the total genetic variance.

$$\sigma_g^2 = 0.95\sigma_u^2 + 0.05\sigma_a^2$$

Where σ_g^2 is the total genetic variance, σ_u^2 is the genetic variance explained by the polygenic effect and σ_a^2 is the genetic variance explained by SNP markers. In pop B, the small sample size led to a share of polygenic and SNP variance that remained quasi-identical to the priors, whatever they were, thus the total genetic variance was split between polygenic and SNP variance in the priors with the same proportions (57 and 43 %, respectively) obtained in pop A after 200,000 iterations of Gibbs sampling.

σ^2_u and σ^2_e were sampled from a scaled inverse chi-square distribution. The initial values for residual variance and total genetic variance priors were the variance component estimates from the linear model described in section 2.8.

In pop A, the degrees of freedom of both parameter distributions were set to 5. In pop B, as the number of individuals was small and to keep the sampling values of σ^2_e within a reasonable range, the degrees of freedom of σ^2_u were set at 5 and the ones of σ^2_e were set at 10,000 to put a strong degree on belief on its prior and ensure the convergence.

The BESSiE software was used to compute this model (Boerner and Tier, 2016). A total of 200,000 iterations of Gibbs sampling were performed with a burn-in of 10,000 iterations. One Gibbs sample was kept every 20 iterations for further analysis. For every iteration, the proportion 1 - π was sampled in a beta distribution $B(\alpha, \beta)$. α was set as the total number of makers in each cohort and β was set at 40.

The degree of confidence in the association between the phenotypes and each SNP was computed using the Bayes Factor (Kass and Raftery, 1995) calculated as:

$$BF_i = \frac{P_i/1 - P_i}{\pi/1 - \pi}$$

where P_i is the probability of the SNP i to have a non-zero effect and π is the proportion of markers with a non-zero effect. BF was transformed into $logBF = 2\log(BF)$ to produce values within the same range as -log(p-values). Strong evidence for the existence of a QTL was considered when at least one SNP had a logBF greater than 8 according to Michenet et al. (2016).

As the causative mutation may not be the marker with the highest logBF (hereafter named the peak SNP), the definition of QTL region was done following the method described in Michenet et al. (2016). All markers close to the peak SNP and having a logBF greater than 3 were considered to be in the QTL region. For each chromosome, the algorithm started at the peak SNP. Then, in a sliding window of 0.5 cM starting from the peak SNP, every SNP with a logBF greater than the noise detection threshold of 3 was included in the QTL region. A sliding window was applied until no SNP had a logBF greater than 3 in the current window. The border of the QTL region was then defined by the last SNP included in the window. This procedure was applied on both sides of the peak SNP, which leads to a QTL region defined as a credibility interval for the causative mutation.

The genetic variance explained by one SNP was calculated as:

$$\sigma^2_{SNP} = 2p(1-p)a^2$$

With p the MAF of the SNP and a the effect of the SNP. To obtain the proportion of the genetic variance explained by one SNP, σ^2_{SNP} was divided by σ^2_g. The proportion of the genetic variance explained by the QTL region was the sum of the proportions explained by each SNP located in the QTL region as previously defined.

3 Results

3.1. VNN challenge

The three challenges were conducted up to 42 days. VNN presence was confirmed by virologic analyses on several subset of fish dead during the infection kinetic, in absence of significant bacterial coinfection. Survival rates ranging from 37.8% to 78.7% were recorded (Figure 1). The peak of mortality was around 10 days after infection.

[Insert Figure 1]

3.2. Performance of the Axiom DlabCHIP SNP array assessed with commercial populations

Among submitted SNPs, 56,730 markers were spotted on the SNP array after ThermoFisher internal selection procedure, among which 537 markers were duplicated. Genotyping commercial populations with the new 57K SNP array revealed a large number of polymorphic SNPs with high clustering resolution (PolyHighResolution, MonoHighResolution and NoMinorHom categories) respectively 50,186, 675 and 2,400 (93.1%) in pop A and 51,686, 252 and 2,045 (94.4%) in pop B. For passing samples, the average genotype call rate was >99.6% in pop A and >99.8% in pop B. The minor allele frequency (MAF) distribution for PolyHighResolution SNPs was similar in both populations (Supplementary Table 1) with an average frequency of 0.28 for pop A and 0.29 for pop B.

3.3. Heritability estimates

The estimates of genomic heritability using linear and threshold models are summarised in Table 2. Estimates of the genomic heritability of the death/survival binary trait on the observed scale (h^2_o) were 0.23 (± 0.05) in pop A and 0.08 (± 0.09) in pop B. In backcross families, h^2_o estimates were 0.38 (± 0.09) in the NEM10 family, 0.59 (± 0.07) in the NEM12 family, 0.50 (± 0.09) in the SEM8 family and 0.23 (± 0.15) in the WEM18 family. The heritabilities estimated with a threshold model on the underlying liability scale (h^2_u) were higher than h^2_o estimates, as expected. Estimates of h^2_u obtained by transforming h^2_o with the

formula by Dempster and Lerner (1950) were similar than the ones obtained with the threshold model in pop A and the WEM18 family, smaller in pop B and higher in the NEM10, NEM12 and SEM8 families (data not shown).

[Insert Table 2]

3.4. Genetic map reconstruction and QTL mapping in the backcross families

We obtained a new high-density genetic map containing 49638 markers that were homogeneously mapped on 24 linkage groups (LG) corresponding to the 24 known chromosomes in the species' karyotype. The total length of the genetic map was 1873.1 cM, corresponding to a density of 26.5 markers per cM.

In the composite interval mapping analysis done in the backcross families, LG12 had a very high LOD score compared to other chromosomes in the NEM10, NEM12 and SEM8 families (Figure 2). At 5% LOD threshold, 725, 322 and 840 markers were detected as potential QTLs on LG12 (out of a total number of 1273, 925 and 1282 on this chromosome for NEM10, NEM12 and SEM8 respectively). In the WEM18 family, no QTL was detected. Additionally, in the NEM12 family, 4 markers of LG8 (out of 925 in this chromosome) were detected as potential QTLs.

[Insert Figure 2]

3.5. QTL detection using GBLUP approach in the commercial cohorts

In the GWAS performed by GBLUP analysis, the p-values of 20 markers on LG12 exceeded the chromosome-wide significance threshold in pop A and, among them, 6 exceeded the genome-wide significance threshold (Figure 3). One additional marker located on LG8 exceeded the chromosome-wide significance threshold. In pop B, one marker on LG12 and one marker on LG15 exceeded the chromosome-wide significance threshold (Figure 3).

[Insert Figure 3]

3.6. QTL detection using a BayesCπ approach in the commercial cohorts

The estimation of the share of genetic variance explained by the polygenic effect represented 57% of the total genetic variance in pop A and 55% in pop B (Table 2). From the BayesCπ model, a total of 5 QTLs were detected with strong evidence (*logBF*>8) in pop A and were located on LG3, LG8, LG12, LG14 and LG19 (Figure 4). The QTL on LG12 was located between 31.71 and 35.16 cM and explained 9.21% of the total genetic variance. Other QTLs explained less than 1% of the total genetic variance each and their locations are summarized in Table 3. QTLs on LG3 and LG19 in pop A were single marker QTLs. In pop B, three QTLs located on LG12, LG15 and LG20 were detected. They all explained 1% of the total genetic variance or less. The QTL on LG12 was located in a smaller confidence interval than the one in pop A, between 33.26 and 33.91 cM.

[Insert Figure 4]

[Insert Table 3]

4. Discussion

In this study, we designed and used a high-density SNP array specifically developed for mapping studies in the European sea bass. The average physical distance (<12kb) and genetic distance (<0.5cM) between

consecutive markers make this array an excellent tool for implementing cost-effective screening of QTLs and genomic selection. The high call rate (>99.6%) and the high number of polymorphic SNPs in the studied populations (>88%) also validate the SNP selection strategy, that permitted to cover the whole genome with high-quality variants, while avoiding technically undesignable SNPs. We used the SNP array to construct a high-density linkage map for composite interval QTL mapping in the backcross families. The accuracy and level of resolution of this new genetic map exceeds that of previous map built in the European sea bass (Palaiokostas et al., 2015).

Using different types of populations (four large backcross families and two commercial admixed populations) and complementary analytical strategies (Composite Interval Mapping, as well as GBLUP and BayesCπ based GWAS), we were able to detect QTLs involved in VNN resistance, the major viral threat to European sea bass aquaculture (Vendramin et al, 2016). While the commercial cohorts were used to estimate the genetic parameters and to detect potential QTL for direct improvement in ongoing breeding programs, the backcross families constitute classical QTL mapping populations. We confirmed that VNN resistance in European sea bass has a moderate heritability in the commercial populations (0.24 to 0.38 using a threshold model), similar to the previously reported estimates (0.26 to 0.43; (Doan et al., 2017a; Palaiokostas et al., 2018)). In the backcross families, the heritability estimates were much higher (0.48 to 0.84). Even though the genomic relationship matrix enables the estimation of genetic parameters in full-sibs families, the estimates can be inaccurate when the family size is too small and when QTLs segregate within the families (Ødegård and Meuwissen, 2012). Except in pop A where the heritability estimated with a threshold model was similar to the one using a linear model and corrected with the Dempster and Lerner formula, all the heritabilities estimated using a threshold model were very different the ones estimated with a linear model and corrected with the Dempster and Lerner formula. In the backcross families, the heritability estimation was challenging due to the single-family structure as well as the small number of individuals. In pop B, it could be because of the small number of individuals, which causes high sampling variance in heritability estimates.

Regardless of the detection method, one QTL was detected on the LG12 in five of the six data sets. Its effect was strong, explaining up to 9.21% of the total genetic variance in the commercial population pop A. The combination of the different results reduced the confidence interval of the QTL to a likely position between 33.26 and 33.91 cM, equivalent to 3.7 Mb and containing 125 of our SNP markers, corresponding to the intersection of all the confidence intervals. From the results we obtained, we cannot yet reach a putative causal mutation and thus, the best markers to use in MAS to predict the phenotype. Another cross-population QTL was detected on LG8, shared by pop A and the NEM12 family. This QTL explained 1.1% of the total genetic variance in pop A. Other QTLs were population-specific as they were only detected in one population.

A previous study led to the detection of minor effect QTLs (Palaiokostas et al., 2018) located on chromosome 3, 20 and 25 explaining 4%, 1.5% and 2% of the total genetic variance respectively. After chromosome correspondence checking by aligning markers sequences on the European sea bass reference genome (GCA_000689215.1) using BLAST (Altschul et al., 1990), chromosome 3 and 20 from that previous study corresponded to the LG12 and LG6 in our study. The QTL explaining the largest part of genetic variance was located on the same chromosome (LG12) in both studies. In Asian seabass, only minor effect QTLs were found, explaining between 1.6% and 2.7% of the phenotypic variance (Wang et al., 2017). Those QTLs, located on chromosomes 1, 8, 14, 15, 16, 19, 20, 21 and 24 were linked with potential candidate genes, all involved in stress response. However, using syntenic blocks proposed by

Vij et al. (2016), we found that no homologous genomic regions were in common between QTLs detected in European seabass and Asian seabass. In Atlantic cod, five QTLs were detected on LG1, LG6, LG18, LG19 and LG20, explaining up to 19.7% of the phenotypic variance (Baranski et al., 2010). As VNN resistance in Atlantic cod has a high heritability (0.75), the QTLs explaining 14.2%, 18.2% and 19.7% of the phenotypic variance could be considered as strong effect QTLs.

In back-cross full-sib families, we used a Composite Interval mapping approach that is the most appropriate for such data. From those analyses, we obtained very high LOD values for the detection of the QTL on LG12 as well as very wide confidence interval (37.8 cM in NEM10, 9.74 cM in NEM12 and 42.3 cM in SEM8) mainly due to strong linkage-disequilibrium between markers. No QTL was found in the WEM18 even though it was the family with the highest survival rate. This absence of the QTL detection may be due to the fact that the WEM18 family did not segregate for resistance QTLs. The mating that produced those families was designed to produce families that segregate for resistance QTLs (section 2.2), by mating putative Rs (R : resistant ; s : susceptible) sires with putative ss dams. Thus, it is likely that each parent of this particular WEM18 family carried the resistance QTLs in only one version. As the global resistance of the family was high, it is rather likely that the WEM18 family could be from a RR sire x ss dam or from a RR sire x RR dam, thus not segregating for the QTL on LG12.

In commercial data sets, we used GBLUP and Bayesian GWAS. In the GBLUP analysis, only the QTL on LG12 exceeded the genome-wide significance threshold in pop A. In pop A and B, the QTL on LG12 exceeded the chromosome-wide significance threshold, as did one marker on LG19 in pop A. In GWAS by BayesCπ, more QTLs were detected due to the variable selection process that increases the power of detection (Tam et al., 2019). In addition, as we had very different number of phenotypes in commercial cohorts, BayesCπ allowed us to test different priors to fit the best model for each data set.

Thanks to the high number of individuals in pop A and thus, the high power of the QTL detection, we were able to refine the location of the QTL on LG12. Both GBLUP and BayesCπ approaches revealed a strong association between VNN resistance and markers located on LG12 as well as a moderate part of the total genetic variance explained by this QTL. Even though the QTL on LG12 explained 9.21% of the total genetic variance, several markers located on the chromosome and detected in pop A seemed to have a potential interest in MAS. Among them, one marker (LG12_8815613) detected had the highest effect on that chromosome and the third highest logBF on LG12. In addition, it had a low MAF (0.13), which is expected when resistance is not well spread in a population, and which also mechanically decreases the percentage of variance explained. Thus, this marker had interesting properties to be a candidate marker to MAS. Its effect on survival across data sets is shown in Figure 5. Interestingly, this marker had a strong effect on survival both in commercial populations (Figure 5) and in the backcross families, with an average of 43.3% survival for genotype AA, 72.8% survival for genotype AB and 78.3% survival for genotype BB. In pop B only, the BB genotyped survived less than the AB genotype, but only five individuals were BB in pop B, and thus the condition to quantify the effect of the BB genotype in this population are far from optimal. Although the consistent effect of this marker on six different cohorts highlights its potential interest, further validation in additional populations or cohorts from the same population would be of great value. The low observed frequency for allele B in the commercial populations could be due to a founder effect or maybe from the fact that these populations have been selected for several traits, among which growth rate for at least 6 generations. As there is a moderate but negative genetic correlation of VNN resistance with body weight (−0.35 ± 0.14; Doan et al., 2017a), this selection for growth might have had a negative effect on the frequency of a VNN resistance allele.

The QTL detection in pop B was more challenging. The strong population structure that we revealed with the PCA led to a strong decrease in the number of individuals that could be used in this data set and thus, to a decrease in the QTL detection power. We only managed to detect two QTLs at chromosome-wide level using GWAS by GBLUP, and three QTLs with Bayesian GWAS. Bayesian GWAS was very sensitive to the part of genetic variance explained by the markers given as priors. In this population, the proportions estimated after 200,000 iterations of Gibbs sampling was similar to the priors, contrary to pop A that estimated new values for those proportions. Thus, a larger number of individuals would be needed to accurately estimate those priors and improve the overall QTL detection. For example, in Korte and Farlow (2013), the authors showed that 800 individuals are necessary to detect a QTL explaining 5% of the phenotypic variance with a power of 0.8 when applying a false discovery rate of 5%. Here, we only had 476 usable individuals in the pop B cohort.

This study produced encouraging results with the discovery of one important QTL involved in VNN resistance in European sea bass. Even though the part of genetic variance explained by this QTL is limited (9.21%), further work on the location of the causal mutation and the discovery of diagnostic marker informing on the resistance genotype of individuals could greatly improve the global sea bass aquaculture. The best example of the application of MAS in aquaculture is the discovery of one major QTL involved in the Infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon and explaining 90% of the genetic variance (Houston et al., 2008; Moen et al., 2009). Since its use by the salmon aquaculture industry to improve selection based on pedigree information, IPNV outbreaks occurrence decreased by 75% (Hjeltnes, 2014). This success was due to the identification of one SNP located on the gene responsible of the resistance and the genotyping of selection candidates to improve the IPNV resistance in aquaculture farms. In the context of VNN resistance in European sea bass, the effect of the QTL on LG12, although significant, is smaller. Its use in MAS will have less benefits due to the smaller part of genetic variance explained. In that situation, genomic selection could be more appropriate to improve VNN resistance. As shown with the example of marker LG12_8815613 (Figure 5), it remains possible that a marker with very large effects on survival would only explain a moderate part of the genetic variance if the resistance allele is rare in the population studied. The presence of a major QTL explaining a large part of the genetic variance in VNN resistance is thus still a valid hypothesis and could be investigated in further work using genome sequence combined with fine mapping approach.

Acknowledgments

This work was partially financially supported by the GeneSea project (n° R FEA 4700 16 FA 100 0005) funded by the French Government and the European Union (EMFF, European Maritime and Fisheries Fund) at the "Appels à projets Innovants" managed by the France Agrimer Office. The doctoral scholarship of Ronan Griot was partially supported by the ANRT (doctoral scholarship n° 2017/0731) and EMG and FMD breeding companies through SYSAAF organisation. We are grateful to the INRAE genotyping platform Gentyane (INRAE, Clermont-Ferrand, France) for the production of genotype data and to the Fortior Genetics platform (ANSES, Plouzané, France) for the realization of the challenges. We especially thank Daniel Guémené for his implication in the creation of the Fortior Genetics platform as well as his implication in the conception of this project.

Declaration of Competing Interest

FA, FP, TM, CP, AV, PAG and MV declare that they have no conflict of interest. RG, SBF, RM, AB, YF and PH are employed by SYSAAF, that provides expertise to the management of aquaculture breeding programs in France. SC, JC, JSB and BP are employed by companies that run fish breeding programs.

Research data

Data on pop A and pop B are not available as they contain confidential information on commercial breeding programs. Data on backcross populations (Allal et al., 2020a) as well as ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip SNP information (Allal et al., 2020b) are publicly available on SeaNoe (https://www.seanoe.org/).

References

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., Misztal, I., 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. Genet. Sel. Evol. 51, 28. https://doi.org/10.1186/s12711-019-0469-3

Allal, F., Griot, R., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin, T., Poncet, C., Vergnet, A., Cariou, S., Brunier, J., Bruant, J.-S., Peyrou, B., Gagnaire, P.-A., Vandeputte, M., 2020a. Genotyping of 4 European sea bass families with the DlabCHIP SNPs genotyping array. SEANOE. https://doi.org/10.17882/75696

Allal, F., Griot, R., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin, T., Poncet, C., Vergnet, A., Cariou, S., Brunier, J., Bruant, J.-S., Peyrou, B., Lavisse, P., Gagnaire, P.-A., Vandeputte, M., 2020b. Axiom DlabCHIP, a 57K SNPs genotyping array for European sea bass (Dicentrarchus labrax): array design and genetic map. SEANOE. https://doi.org/10.17882/75680

Baranski, M., Kettunen Praebel, A., Sommer, A., Haanes, K., Wesmajervi, M., 2010. Major quantitative trait loci for viral nervous necrosis resistance in atlantic cod, in: Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: August 1-6, 2010 Leipzig, Germany. de.

Boerner, V., Tier, B., 2016. BESSiE: a software for linear model BLUP and Bayesian MCMC analysis of large-scale genomic data. Genet. Sel. Evol. 48, 63. https://doi.org/10.1186/s12711-016-0241-x

Breuil, G., Bonami, J.R., Pepin, J.F., Pichot, Y., 1991. Viral infection (picorna-like virus) associated with mass mortalities in hatchery-reared sea-bass (Dicentrarchus labrax) larvae and juveniles. Aquaculture 97, 109–116. https://doi.org/10.1016/0044-8486(91)90258-9

Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19, 889–890. https://doi.org/10.1093/bioinformatics/btg112

Churchill, G.A., Doerge, R.W., 1994. Empirical threshold values for quantitative trait mapping. Genetics 138, 963.

Dempster, E.R., Lerner, I.M., 1950. Heritability of Threshold Characters. Genetics 35, 212–236.

Doan, Q.K., Vandeputte, M., Chatain, B., Haffray, P., Vergnet, A., Breuil, G., Allal, F., 2017a. Genetic variation of resistance to Viral Nervous Necrosis and genetic correlations with production traits in wild

518 populations of the European sea bass (Dicentrarchus labrax). Aquaculture 478, 1–8.
519 https://doi.org/10.1016/j.aquaculture.2017.05.011

520 Doan, Q.K., Vandeputte, M., Chatain, B., Morin, T., Allal, F., 2017b. Viral encephalopathy and retinopathy
521 in aquaculture: a review. J. Fish Dis. 40, 717–742. https://doi.org/10.1111/jfd.12541

522 Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., Gagnaire, P.-A., 2018. The origin and
523 remolding of genomic islands of differentiation in the European sea bass. Nat. Commun. 9, 2518.
524 https://doi.org/10.1038/s41467-018-04963-6

525 FEAP, 2017. Annual Report 2017 [WWW Document]. FEAP – The Federation of European Aquaculture
526 Producers. URL https://issuu.com/feapsec/docs/feap_ar2017 (accessed 4.16.20).

527 Georges, M., Dietz, A.B., Mishra, A., Nielsen, D., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X.,
528 Leipold, H., Womack, J.E., 1993. Microsatellite mapping of the gene causing weaver disease in cattle will
529 allow the study of an associated quantitative trait locus. Proc. Natl. Acad. Sci. 90, 1058–1062.
530 https://doi.org/10.1073/pnas.90.3.1058

531 Gjedrem, T., Robinson, N., 2014. Advances by Selective Breeding for Aquatic Species: A Review. Agric. Sci.
532 05, 1152–1158. https://doi.org/10.4236/as.2014.512125

533 Gomez-Casado, E., Estepa, A., Coll, J.M., 2011. A comparative review on European-farmed finfish RNA
534 viruses and their vaccines. Vaccine 29, 2657–2671. https://doi.org/10.1016/j.vaccine.2011.01.097

535 Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F., Vandeputte, M., 2020. APIS: An
536 auto-adaptive parentage inference software that tolerates missing parents. Mol. Ecol. Resour. 20, 579–
537 590. https://doi.org/10.1111/1755-0998.13103

538 Habier, D., Fernando, R.L., Kizilkaya, K., Garrick, D.J., 2011. Extension of the bayesian alphabet for
539 genomic selection. BMC Bioinformatics 12, 186. https://doi.org/10.1186/1471-2105-12-186

540 Haley, C.S., Knott, S.A., 1992. A simple regression method for mapping quantitative trait loci in line
541 crosses using flanking markers. Heredity 69, 315–324. https://doi.org/10.1038/hdy.1992.131

542 Hayes, B., 2013. Overview of Statistical Methods for Genome-Wide Association Studies (GWAS), in:
543 Gondro, C., van der Werf, J., Hayes, B. (Eds.), Genome-Wide Association Studies and Genomic Prediction,
544 Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 149–169. https://doi.org/10.1007/978-1-
545 62703-447-0_6

546 Houston, R., Gheyas, A., Hamilton, A., Guy, D., Tinch, A., Taggart, J., McAndrew, B., Haley, C., Bishop, S.,
547 2008. Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis
548 (IPN) in Atlantic salmon (Salmo salar), in: Animal Genomics for Animal Health. Karger Publishers, pp.
549 199–204.

550 Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795.
551 https://doi.org/10.1080/01621459.1995.10476572

552 Korte, A., Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant
553 Methods 9, 29. https://doi.org/10.1186/1746-4811-9-29

554 Le Breton, A., Grisez, L., Sweetman, J., Ollevier, F., 1997. Viral nervous necrosis (VNN) associated with
555 mass mortalities in cage-reared sea bass, Dicentrarchus labrax (L.). J. Fish Diseases 20, 145–151.
556 https://doi.org/10.1046/j.1365-2761.1997.00284.x

557 Legarra, A., Robert-Granié, C., Manfredi, E., Elsen, J.-M., 2008. Performance of Genomic Selection in
558 Mice. Genetics 180, 611–618. https://doi.org/10.1534/genetics.108.088575

559 Michenet, A., Barbat, M., Saintilan, R., Venot, E., Phocas, F., 2016. Detection of quantitative trait loci for
560 maternal traits using high-density genotypes of Blonde d'Aquitaine beef cattle. BMC Genet. 17, 88.
561 https://doi.org/10.1186/s12863-016-0397-y

562 Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D., 2002. BLUPF90 and related programs
563 (BGF90). Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Communication 28, 743–
564 744.

565 Moen, T., Baranski, M., Sonesson, A.K., Kjøglum, S., 2009. Confirmation and fine-mapping of a major QTL
566 for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level
567 associations between markers and trait. BMC Genomics 10, 368. https://doi.org/10.1186/1471-2164-10-
568 368

569 Mori, K.-I., Nakai, T., Muroga, K., Arimoto, M., Mushiake, K., Furusawa, I., 1992. Properties of a new virus
570 belonging to nodaviridae found in larval striped jack (Pseudocaranx dentex) with nervous necrosis.
571 Virology 187, 368–371. https://doi.org/10.1016/0042-6822(92)90329-N

572 Ødegård, J., Baranski, M., Gjerde, B., Gjedrem, T., 2011. Methodology for genetic evaluation of disease
573 resistance in aquaculture species: challenges and future prospects: Genetic evaluation of disease
574 resistance. Aquac. Res. 42, 103–114. https://doi.org/10.1111/j.1365-2109.2010.02669.x

575 Ødegård, J., Meuwissen, T.H., 2012. Estimation of heritability from limited family data using genome-
576 wide identity-by-descent sharing. Genet. Sel. Evol. 44, 16. https://doi.org/10.1186/1297-9686-44-16

577 Palaiokostas, C., Bekaert, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., Chatain, B., Penman, D.J.,
578 Vandeputte, M., 2015. A new SNP-based vision of the genetics of sex determination in European sea bass
579 (Dicentrarchus labrax). Genet. Sel. Evol. 47, 68. https://doi.org/10.1186/s12711-015-0148-y

580 Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J.-S., Haffray, P., Morin, T., Cabon, J., Allal, F., Vandeputte,
581 M., Houston, R.D., 2018. Genome-wide association and genomic prediction of resistance to viral nervous
582 necrosis in European sea bass (Dicentrarchus labrax) using RAD sequencing. Genet. Sel. Evol. 50, 30.
583 https://doi.org/10.1186/s12711-018-0401-2

584 Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P., 2000. Association Mapping in Structured
585 Populations. Am. J. Hum. Genet. 67, 170–181. https://doi.org/10.1086/302959

586 Purcell, S., Chang, C., 2015. PLINK 1.9. URL https://www. cog-genomics. org/plink2.

587 Rastas, P., 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing
588 data. Bioinformatics 33, 3726–3732. https://doi.org/10.1093/bioinformatics/btx494

589 Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype imputation
590 using information from relatives. BMC Genomics 15, 478. https://doi.org/10.1186/1471-2164-15-478

591 Smith, B.J., 2007. boa: An R Package for MCMC Output Convergence Assessment and Posterior
592 Inference. J. Stat. Softw. 21, 1–37.

593 Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Ødegard, J., Meuwissen, T.H., 2009. Persistence of
594 accuracy of genome-wide breeding values over generations when including a polygenic effect. Genet.
595 Sel. Evol. 41, 53. https://doi.org/10.1186/1297-9686-41-53

596 Sommerset, I., Krossøy, B., Biering, E., Frost, P., 2005. Vaccines for fish in aquaculture. Expert Rev.
597 Vaccines 4, 89–101. https://doi.org/10.1586/14760584.4.1.89

598 Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-
599 wide association studies. Nat. Rev. Genet. 20, 467–484. https://doi.org/10.1038/s41576-019-0127-1

600 Thiéry, R., Cozien, J., de Boisséson, C., Kerbart-Boscher, S., Névarez, L., 2004. Genomic classification of
601 new betanodavirus isolates by phylogenetic analysis of the coat protein gene suggests a low host-fish
602 species specificity. J. Gen. Virol. 85, 3079–3087. https://doi.org/10.1099/vir.0.80264-0

603 Tsurata, S., Misztal, I., 2006. THRGIBBS1F90 for estimation of variance components with threshold-linear
604 models. Presented at the Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil.
605 Instituto Prociencia, Belo Horizonte, Brazil., pp. 27–31.

606 Ulmer, J.B., Valley, U., Rappuoli, R., 2006. Vaccine manufacturing: challenges and solutions. Nat.
607 Biotechnol. 24, 1377–1383. https://doi.org/10.1038/nbt1261

608 VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91, 4414–4423.
609 https://doi.org/10.3168/jds.2007-0980

610 Vendramin, N., Zrncic, S., Padros, F., Oraic, D., Le Breton, A., Zarza, C., Olesen, N.J., 2016. Fish health
611 inMediterranean Aquaculture, past mistakes and future challenges. Bulletin of The European Association
612 of Fish Pathologists 36, 38–45.

613 Vij, S., Kuhl, H., Kuznetsova, I.S., Komissarov, A., Yurchenko, A.A., Van Heusden, P., Singh, S.,
614 Thevasagayam, N.M., Prakki, S.R.S., Purushothaman, K., Saju, J.M., Jiang, J., Mbandi, S.K., Jonas, M., Hin
615 Yan Tong, A., Mwangi, S., Lau, D., Ngoh, S.Y., Liew, W.C., Shen, X., Hon, L.S., Drake, J.P., Boitano, M., Hall,
616 R., Chin, C.-S., Lachumanan, R., Korlach, J., Trifonov, V., Kabilov, M., Tupikin, A., Green, D., Moxon, S.,
617 Garvin, T., Sedlazeck, F.J., Vurture, G.W., Gopalapillai, G., Kumar Katneni, V., Noble, T.H., Scaria, V.,
618 Sivasubbu, S., Jerry, D.R., O'Brien, S.J., Schatz, M.C., Dalmay, T., Turner, S.W., Lok, S., Christoffels, A.,
619 Orbán, L., 2016. Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads
620 and Multi-layered Scaffolding. PLoS Genet. 12, e1005954. https://doi.org/10.1371/journal.pgen.1005954

621 Wang, L., Liu, P., Huang, S., Ye, B., Chua, E., Wan, Z.Y., Yue, G.H., 2017. Genome-Wide Association Study
622 Identifies Loci Associated with Resistance to Viral Nervous Necrosis Disease in Asian Seabass. Mar.
623 Biotechnol. 19, 255–265. https://doi.org/10.1007/s10126-017-9747-7

624

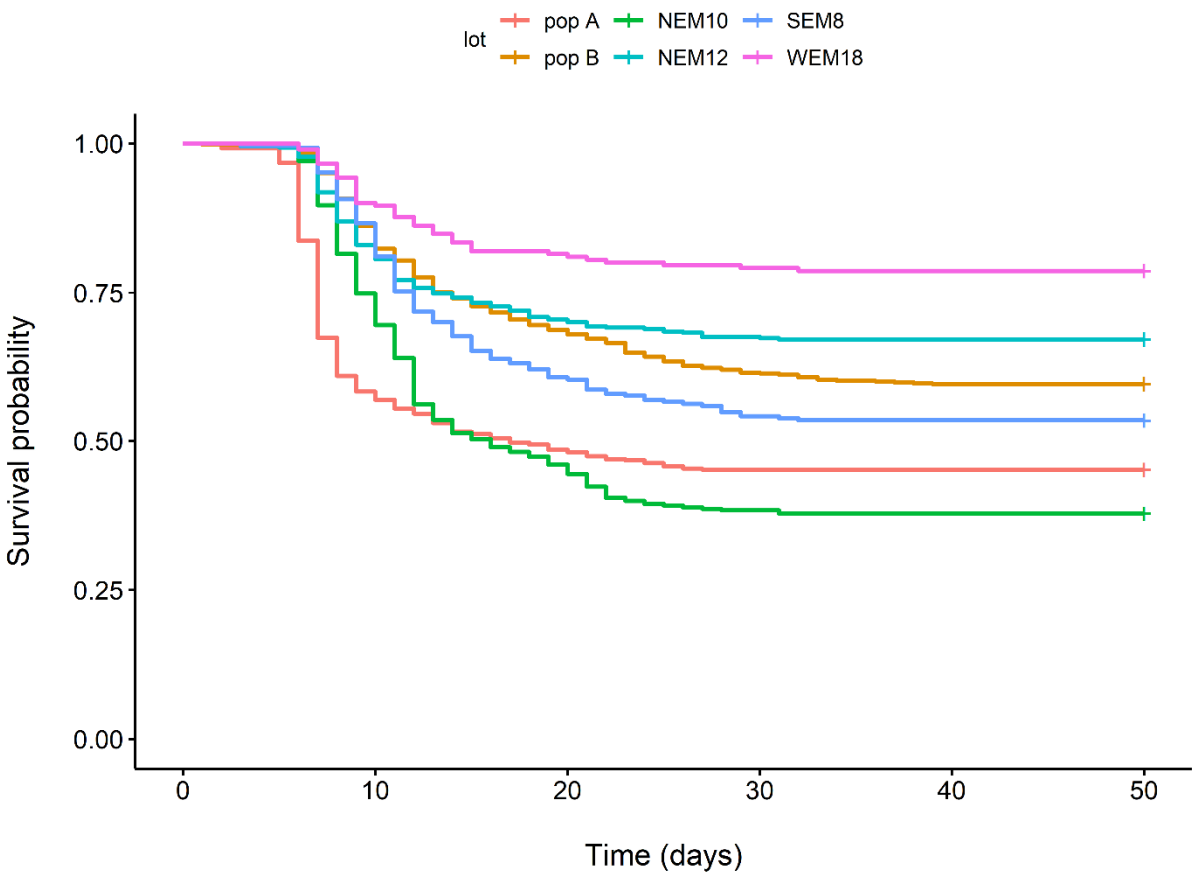

Figure 1 : Evolution of the Kaplan-Meier probability of survival of six different European sea bass populations during their respective VNN infection challenge. Pop A and pop B are commercial populations, and NEM10, NEM 12, SEM8 and WEM 18 are backcross families.

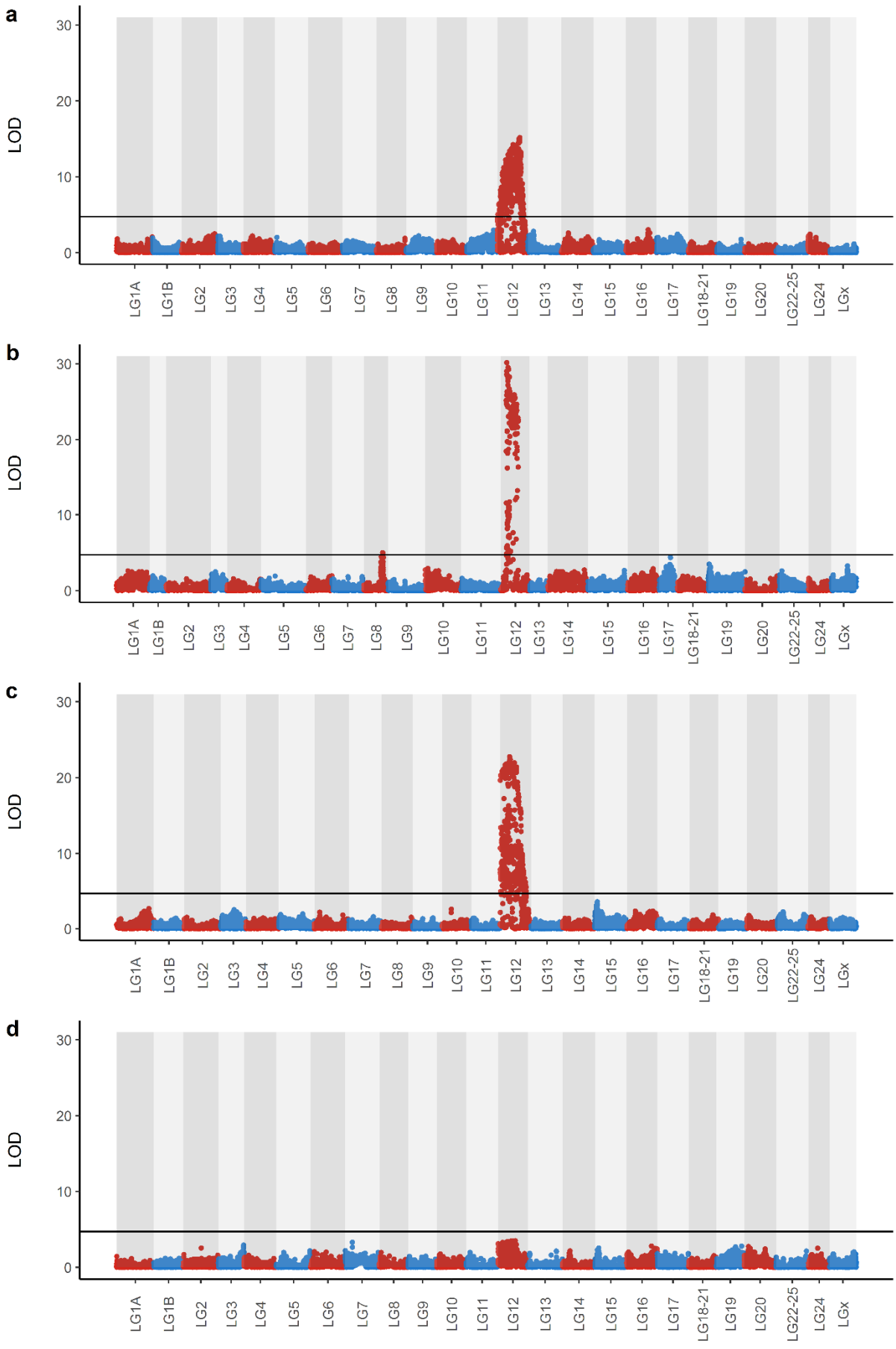
a
b
c
d
LOD
0
10
20
30
LG1A
LG1B
LG2
LG3
LG4
LG5
LG6
LG7
LG8
LG9
LG10
LG11
LG12
LG13
LG14
LG15
LG16
LG17
LG18-21
LG19
LG20
LG22-25
LG24
LGx

Figure 2: Manhattan plot of the LOD obtained by composite interval mapping for VNN resistance QTLs in European sea bass in the NEM 10 (a), NEM12 (b), SEM8 (c) and WEM18 (b) experimental backcrosses. The horizontal black line represents the 5% genome-wide significance threshold calculated over 1000 permutations.

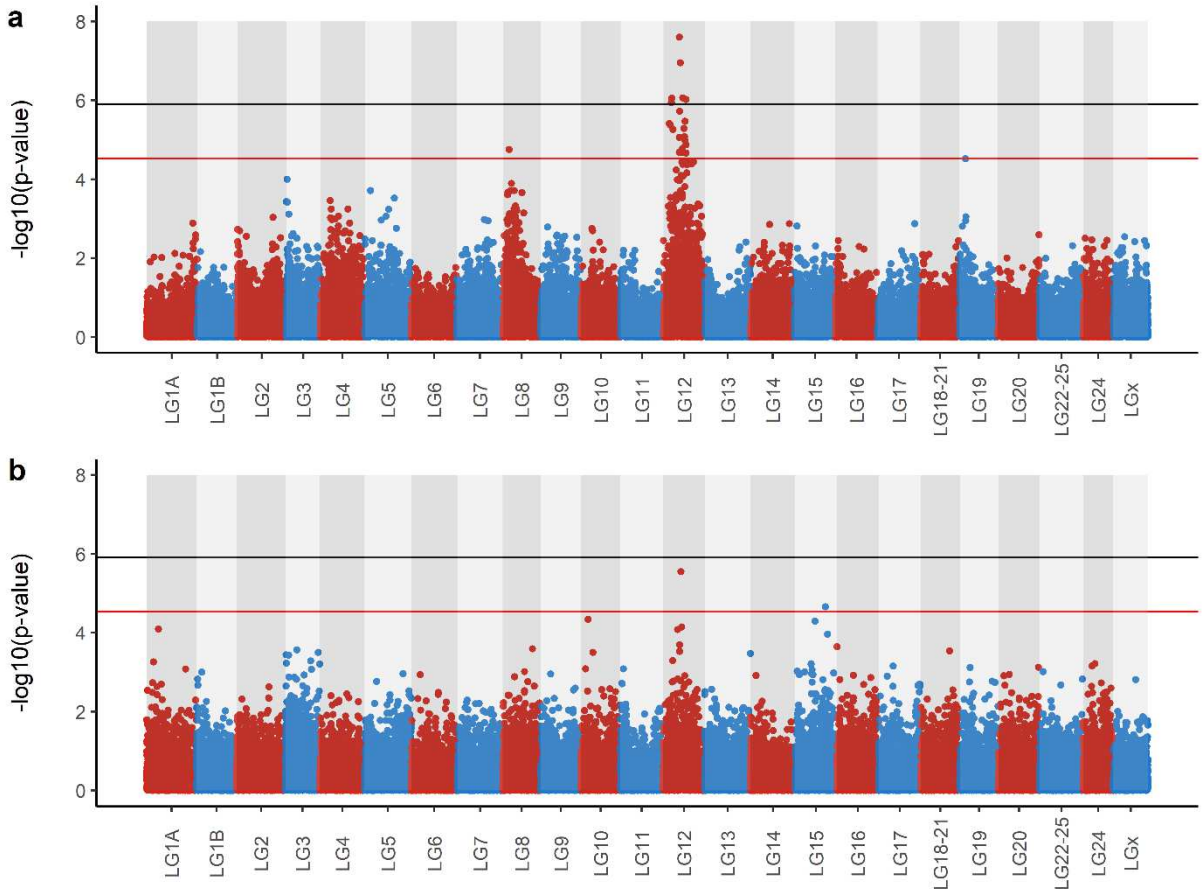

Figure 3: Manhattan plot of -log10(p-value) obtained from GWAS for VNN resistance QTLs in commercial European sea bass populations pop A (a) and pop B (b). The horizontal black line represents the genome-wide significance threshold and the red line the chromosome-wide significance threshold calculated with the Bonferroni correction.

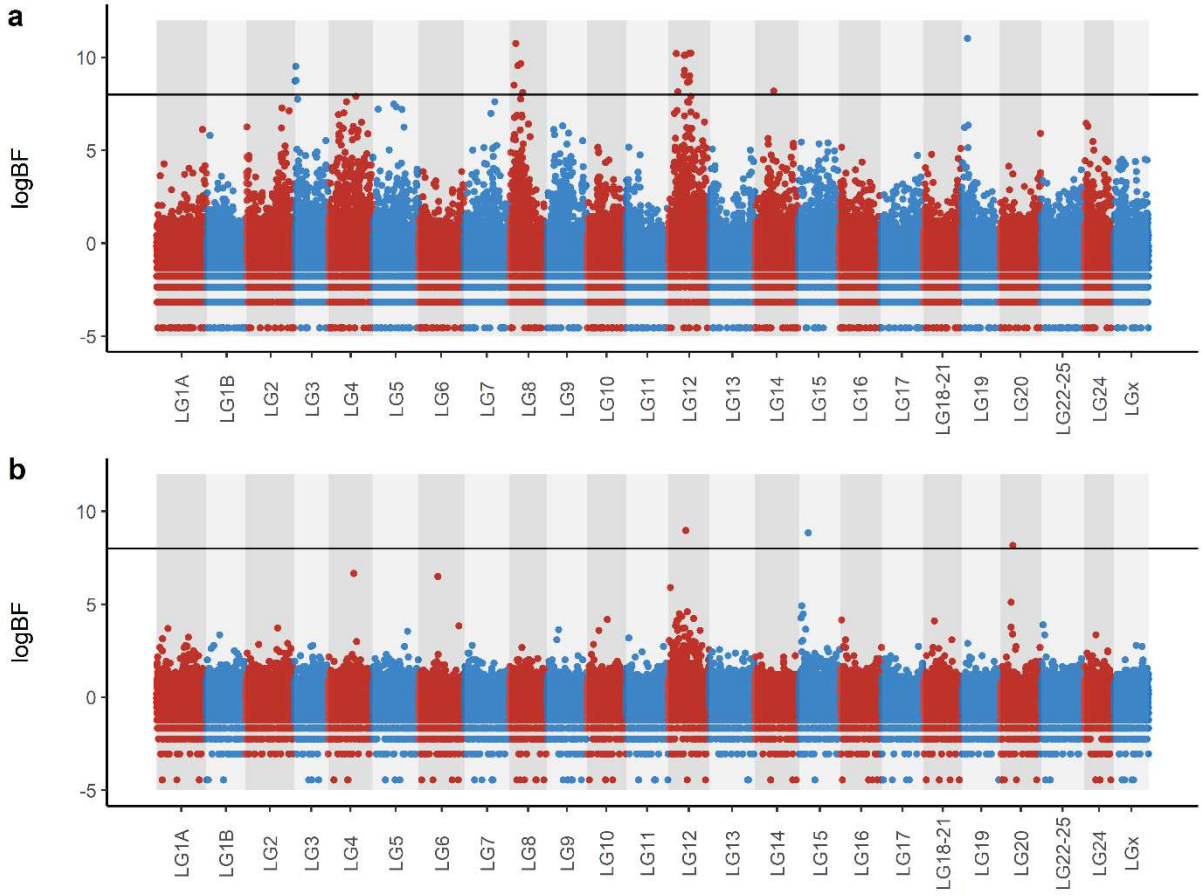

Figure 4: Genome-wide logBF plot for VNN resistance across the genome in the European sea bass populations pop A (a) and pop B (b) using a BayesCπ model. Horizontal black lines represent the logBF threshold of 8, corresponding to strong evidence for the presence of a QTL.

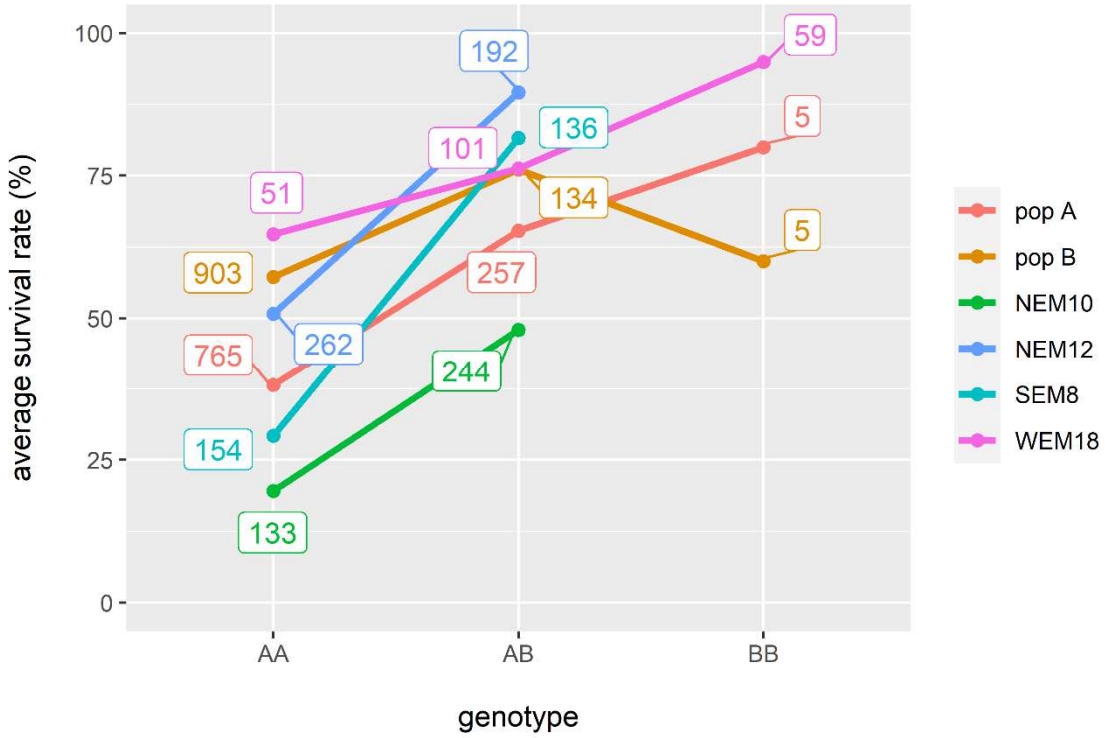

Figure 5: Average survival rate (%) to VNN for each genotype at marker LG12_8815613 in six cohorts of European sea bass. Each colour corresponds to one data set. The numbers in the boxes are the numbers of individuals per genotype.

Table 1: Study design and sampling strategy applied for each of the six sea bass populations challenged for VNN and the additional one used to create the genetic map.

	pop A	pop B	NEM10	NEM12	SEM8	WEM18	additionnal population
number of individuals produced	1680	1737		2500			880
number of parents (sires / dams)	59 / 20	39 / 14	1 / 1	1 / 1	1 / 1	1 / 1	94 / 39
number of individuals challenged	1350	1212		1719			
average survival rate	45.2	59.7	37.8	67.2	53.6	78.7	
number of individuals genotyped	1152	1152		1536			880
number of markers retained for the creation of the genetic map				51179			51179
number of individuals retained after quality control	1089	476	378	454	291	211	
number of markers retained after quality control	40623	41166	30917	23592	30656	31490	
type of analyses	GBLUP BayesCpi		composite interval mapping				building of the genetic map

Table 2: Variance components and genetics parameters for VNN resistance in European sea bass estimated with three models in six different populations. h^2_o is the heritability on the observed (binary) scale and h^2_u is the heritability estimate on the underlying liability scale.

population	linear model			threshold model			BayesCpi			
	Vg	Ve	h2	Vg	Ve*	h2	Vpoly	Vsnp	Ve	h2
pop A	0.057	0.189	0.231 (± 0.049)	0.627	1.086	0.377 (± 0.065)	0.029	0.039	0.182	0.272
pop B	0.019	0.223	0.078 (± 0.085)	0.404	1.016	0.244 (± 0.140)	0.010	0.012	0.223	0.090
NEM10	0.111	0.180	0.381 (± 0.089)	1.538	1.018	0.584 (± 0.088)				
NEM12	0.187	0.130	0.591 (± 0.067)	5.711	1.013	0.838 (± 0.044)				
SEM8	0.164	0.165	0.499 (± 0.094)	2.328	1.023	0.677 (± 0.078)				
WEM18	0.043	0.147	0.227 (± 0.147)	1.164	1.028	0.478 (± 0.170)				

* : In threshold models, σ^2_e is constrained to a value close to 1

Table 3: QTL detection by BayesCpi for resistance to VNN in two commercial populations of European sea bass : position of the peak SNPs, credibility intervals and proportions of genetic variance explained

population	chromosome	peak SNP position (cM)	start QTL position (cM)	end QTL position (cM)	share of total genetic variance explained (%)
pop A	LG3	5.89	5.89	5.89	0.45
	LG8	23.65	23.27	23.65	1.10
	LG12	34.47	31.71	35.16	9.21
	LG14	55.93	55.68	55.93	0.29
	LG19	22.24	22.24	22.24	1.04
pop B	LG12	33.64	33.26	33.91	1.09
	LG15	57.94	57.94	57.94	0.50
	LG20	41.58	41.51	41.58	0.39