



**HAL**  
open science

# 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.

## ► To cite this version:

Ronan Griot, François Allal, Florence Phocas, Sophie Brard-Fudulea, Romain Morvezen, 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, pp.735930. 10.1016/j.aquaculture.2020.735930 . hal-03544239

**HAL Id: hal-03544239**

**<https://hal.inrae.fr/hal-03544239>**

Submitted on 21 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Genome-wide association studies for resistance to viral nervous  
2 necrosis in three populations of European sea bass  
3 (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip

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

7 <sup>a</sup> SYSAAF, Station LPGP/INRAE, Campus de Beaulieu, 35042 Rennes, France

8 <sup>b</sup> Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

9 <sup>c</sup> MARBEC, Univ. Montpellier, Ifremer, CNRS, IRD, 34250 Palavas-les-Flots, France

10 <sup>d</sup> SYSAAF, UMR BOA, Centre INRAE Val de Loire, 37380 Nouzilly, France

11 <sup>e</sup> INRAE-UCA, UMR GDEC, Clermont-Ferrand, France

12 <sup>f</sup> ANSES, Ploufragan-Plouzané-Niort Laboratory, Viral Fish Diseases Unit, National Reference Laboratory  
13 for Regulated Fish Diseases, Technopôle Brest-Iroise, 29280 Plouzané, France

14 <sup>g</sup> Ferme Marine Du Douhet, 17840 La Brée Les Bains, France

15 <sup>h</sup> Ecloserie Marine de Gravelines-Ictus, 59273 Gravelines, France

16 <sup>i</sup> ISEM, Université de Montpellier, CNRS-IRD-EPHE-UM, France

17 Corresponding author: Marc Vandeputte, e-mail: [marc.vandeputte@inrae.fr](mailto:marc.vandeputte@inrae.fr)

18 Abstract

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

25 Four experimental backcross families comprising 378, 454, 291 and 211 individuals and two commercial  
26 populations A and B comprising 1027 and 1042 individuals obtained from partial factorial crosses (59♂ x  
27 20♀ for pop A; 39♂ x 14♀ for pop B) were submitted to a redspotted grouper nervous necrosis virus  
28 (RGNNV) challenge by bath. A high-density single nucleotide polymorphism (SNP) chip panel was  
29 designed to develop the ThermoFisher Axiom™ 57k SNP DlabChip, which was used for genotyping all  
30 individuals and building a high quality linkage map. In the backcross families, composite interval mapping  
31 was performed on 30917, 23592, 30656 and 31490 markers, respectively. In the commercial  
32 populations, 40263 markers in pop A and 41166 markers in pop B were used to perform genome-wide  
33 association studies (GWAS) using a GBLUP and a BayesCπ approach.

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

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

## 43 Keywords

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

## 45 1. Introduction

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

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

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

74 the candidates for selection or commercial diffusion by Marker Assisted Selection (MAS) (Houston et al.,  
75 2008; Moen et al., 2009). In Asian sea bass (*Lates calcarifer*), QTL mapping revealed nine QTLs involved in  
76 VNN resistance, explaining between 1.6% and 2.7% of the phenotypic variance (Wang et al., 2017). In  
77 Atlantic cod (*Gadus morhua*), five QTLs were detected, including three QTL with major effects, explaining  
78 between 14.2% and 19.7% of the phenotypic variance (Baranski et al., 2010). For VNN resistance in sea  
79 bass, a recent study reported three first QTLs, each of them explaining a small part of the genetic  
80 variance (1.5% to 4%) (Palaiokostas et al., 2018).

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

## 87 2. Material and Methods

### 88 2.1. Ethics approval

89 All infection challenges were carried out in accordance with the European guidelines (Directive 2010–63-  
90 EU) and the corresponding French legislation. Animal experiment procedures were approved by the  
91 ethics committee on animal experimentation COMETH ANSES/ENVA/UPC No. 16 and were authorized by  
92 the French Ministry of Higher Education, Research and Innovation under numbers 2017022816255366,  
93 29/01/13-5 and 10/03/15-1.

### 94 2.2. Fish material

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

114 [Insert Table 1]

### 115 2.3. VNN challenge

116 All fish were challenged to VNN at the SYSAAF-ANSES Fortior Genetics platform (ANSES, Plouzané,  
117 France). For all experiments, fish were maintained in filtered seawater at a temperature of  $27^{\circ}\text{C} \pm 2^{\circ}\text{C}$  in  
118 an open circuit. Pop A, pop B and backcross families infectious challenges were performed separately but  
119 in the same way. All the fish sent to the infection challenge were individually tagged. A total of 1680  
120 from pop A (25g mean weight), 1737 from pop B (20g mean weight) and 2500 fish from the backcross  
121 families (8g mean weight) were received and acclimated for a minimum period of three weeks. Then, the  
122 whole batch of each population was split into a pre-test batch (120, 150 and 500 individuals for pop A,  
123 pop B and backcrosses, respectively), a challenge batch (1350, 1212 and 1719 individuals for pop A, pop  
124 B and backcrosses, respectively) and a negative control batch (120, 275 and 150 fish from the  
125 corresponding populations). For the pre-test and challenge batches contamination, the fish were  
126 immersed for 2 to 3 hours in a static bath of aerated seawater containing  $1 \times 10^5$  TCID<sub>50</sub>/ml of the W80  
127 strain of RGNNV (redspotted grouper nervous necrosis virus), that was previously produced on a striped  
128 snakehead SSN1 cell line (Thiéry et al., 2004). Then, the batch was split into two tanks, water circulation  
129 was restarted and the mortality was recorded during 16 days. The mortality rate was 44% and 51% in the  
130 pop A pre-test, 51% and 53% in the pop B pre-test and 22.6% and 23.7% for backcross families. For the  
131 infectious challenges, a protocol similar to the pre-test was applied on the challenge batches. The  
132 negative controls were immersed for three hours into static seawater containing sterile cell culture  
133 medium. The mortality was recorded each day during the challenge period that was 27 days for pop A,  
134 42 days for pop B and 33 days for the backcross families. Bacteriologic and virologic analyses were  
135 performed before the pre-test and challenges as well as during the mortality peak to check the sanitary  
136 status of the fish. Virologic analyses were done by injecting a homogenized mixture of eye and brain  
137 sampled on random dead fish on SSN1 cells. Then, if cytopathic effects were observed, a virus  
138 identification was performed by immunofluorescence using anti-NNV antibodies. For bacteriological  
139 analyses, spleen and kidney are randomly sampled and used for bacterial cultures followed by MALDI-  
140 TOF identification.

### 141 2.4. Design of the Sea Bass 57k SNP DLabCHIP array

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

### 153 2.5. Genotyping and Parentage assignment

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

158 challenged ones among **the** dead and the surviving individuals. In the backcross families, they were  
159 selected to have the same average mortality per family as the whole challenged family. SNP calling was  
160 done using ThermoFisher software AxiomAnalysisSuite™. Preliminary quality controls were applied with  
161 threshold values of 95% for SNP call rate and 90% for sample call rate. Parentage assignment was done  
162 using 1000 randomly sampled markers with the R package APIS (Griot et al., 2019) with a positive  
163 assignment error rate set to 1%.  
164

## 165 2.6. Creation of the genetic map

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

170 We ran the recommended procedure of LepMap3 with custom settings: 1% segregation distortion for  
171 Filtering2, a LOD score of 50 as well as a subset of 25% of the markers for SeparateChromosomes2, and  
172 LOD score of 30 for JoinSingles2All.  
173

## 174 2.7. Genotyping and sample quality control

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

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

189 Finally, within each cohort, the few missing genotype data were imputed using Fimpute software  
190 (Sargolzaei et al., 2014) to obtain complete genotypes for GWAS analysis.  
191

## 192 2.8. Heritability estimation

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

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

$$200 \quad y = 1b + Zu + e$$

201

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

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

213 The heritability for survival was estimated as:

$$214 \quad h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

215 Heritability on the observed scale ( $h_o^2$ ) was estimated using the variance components from the linear  
216 model, while the heritability on the underlying liability scale ( $h_u^2$ ) was computed using the variance  
217 components from the threshold model.

218

## 219 2.9. QTL mapping in experimental crosses

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

## 225 2.10. QTL mapping in commercial cohorts

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

### 228 2.10.1. GBLUP-based GWAS

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

$$232 \quad y = 1b + Zu + e$$

233  
234 With the same notation as in section 2.8. The p-values were computed using `POSTGSF90` (Aguilar et al.,  
235 2019). The  $-\log_{10}$  of the p-values were compared to the chromosome-wide significance threshold and to  
236 the genome-wide significance threshold at 5% after Bonferroni correction for the average number of  
237 markers per chromosome and the total number of markers, respectively.

### 238 2.10.2. Bayesian-based GWAS

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

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

$$248 \quad y = 1b + Zu + Wg + e$$

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

$$256 \quad \sigma_g^2 = 0.95\sigma_u^2 + 0.05\sigma_a^2$$

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

262  $\sigma_u^2$  and  $\sigma_e^2$  were sampled from a scaled inverse chi-square distribution. The initial values for residual  
 263 variance and total genetic variance priors were the variance component estimates from the linear model  
 264 described in section 2.8.

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

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

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

$$275 \quad BF_i = \frac{P_i/1 - P_i}{\pi/1 - \pi}$$

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



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

289 The genetic variance explained by one SNP was calculated as:

$$290 \quad \sigma_{SNP}^2 = 2p(1 - p)a^2$$

291 With  $p$  the MAF of the SNP and  $a$  the effect of the SNP. To obtain the proportion of the genetic variance  
292 explained by one SNP,  $\sigma_{SNP}^2$  was divided by  $\sigma_g^2$ . The proportion of the genetic variance explained by the  
293 QTL region was the sum of the proportions explained by each SNP located in the QTL region as previously  
294 defined.

## 295 3 Results

### 296 3.1. VNN challenge

297 The three challenges were conducted up to 42 days. VNN presence was confirmed by virologic analyses  
298 on several subset of fish dead during the infection kinetic, in absence of significant bacterial coinfection.  
299 Survival rates ranging from 37.8% to 78.7% were recorded (Figure 1). The peak of mortality was around  
300 10 days after infection.

301 [Insert Figure 1]

### 302 3.2. Performance of the Axiom DlabCHIP SNP array assessed with commercial 303 populations

304 Among submitted SNPs, 56,730 markers were spotted on the SNP array after ThermoFisher internal  
305 selection procedure, among which 537 markers were duplicated. Genotyping commercial populations  
306 with the new 57K SNP array revealed a large number of polymorphic SNPs with high clustering resolution  
307 (PolyHighResolution, MonoHighResolution and NoMinorHom categories) respectively 50,186, 675 and  
308 2,400 (93.1%) in pop A and 51,686, 252 and 2,045 (94.4%) in pop B. For passing samples, the average  
309 genotype call rate was >99.6% in pop A and >99.8% in pop B. The minor allele frequency (MAF)  
310 distribution for PolyHighResolution SNPs was similar in both populations (Supplementary Table 1) with  
311 an average frequency of 0.28 for pop A and 0.29 for pop B.

### 312 3.3. Heritability estimates

313 The estimates of genomic heritability using linear and threshold models are summarised in Table 2.  
314 Estimates of the genomic heritability of the death/survival binary trait on the observed scale ( $h_o^2$ ) were  
315 0.23 ( $\pm 0.05$ ) in pop A and 0.08 ( $\pm 0.09$ ) in pop B. In backcross families,  $h_o^2$  estimates were 0.38 ( $\pm 0.09$ ) in  
316 the NEM10 family, 0.59 ( $\pm 0.07$ ) in the NEM12 family, 0.50 ( $\pm 0.09$ ) in the SEM8 family and 0.23 ( $\pm 0.15$ )  
317 in the WEM18 family. The heritabilities estimated with a threshold model on the underlying liability scale  
318 ( $h_u^2$ ) were higher than  $h_o^2$  estimates, as expected. Estimates of  $h_u^2$  obtained by transforming  $h_o^2$  with the

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

322 [Insert Table 2]

323 3.4. Genetic map reconstruction and QTL mapping in the backcross families

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

328 In the composite interval mapping analysis done in the backcross families, LG12 had a very high LOD  
329 score compared to other chromosomes in the NEM10, NEM12 and SEM8 families (Figure 2). At 5% LOD  
330 threshold, 725, 322 and 840 markers were detected as potential QTLs on LG12 (out of a total number of  
331 1273, 925 and 1282 on this chromosome for NEM10, NEM12 and SEM8 respectively). In the WEM18  
332 family, no QTL was detected. Additionally, in the NEM12 family, 4 markers of LG8 (out of 925 in this  
333 chromosome) were detected as potential QTLs.

334 [Insert Figure 2]

335 3.5. QTL detection using GBLUP approach in the commercial cohorts

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

341 [Insert Figure 3]

342 3.6. QTL detection using a BayesC $\pi$  approach in the commercial cohorts

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

352 [Insert Figure 4]

353 [Insert Table 3]

## 354 4. Discussion

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

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

364 Using different types of populations (four large backcross families and two commercial admixed  
365 populations) and complementary analytical strategies (Composite Interval Mapping, as well as GBLUP  
366 and BayesC $\pi$  based GWAS), we were able to detect QTLs involved in VNN resistance, the major viral  
367 threat to European sea bass aquaculture (Vendramin et al, 2016). While the commercial cohorts were  
368 used to estimate the genetic parameters and to detect potential QTL for direct improvement in ongoing  
369 breeding programs, the backcross families constitute classical QTL mapping populations. We confirmed  
370 that VNN resistance in European sea bass has a moderate heritability in the commercial populations  
371 (0.24 to 0.38 using a threshold model), similar to the previously reported estimates (0.26 to 0.43; (Doan  
372 et al., 2017a; Palaiokostas et al., 2018)). In the backcross families, the heritability estimates were much  
373 higher (0.48 to 0.84). Even though the genomic relationship matrix enables the estimation of genetic  
374 parameters in full-sibs families, the estimates can be inaccurate when the family size is too small and  
375 when QTLs segregate within the families (Ødegård and Meuwissen, 2012). Except in pop A where the  
376 heritability estimated with a threshold model was similar to the one using a linear model and corrected  
377 with the Dempster and Lerner formula, all the heritabilities estimated using a threshold model were very  
378 different the ones estimated with a linear model and corrected with the Dempster and Lerner formula.  
379 In the backcross families, the heritability estimation was challenging due to the single-family structure as  
380 well as the small number of individuals. In pop B, it could be because of the small number of individuals,  
381 which causes high sampling variance in heritability estimates.

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

391 A previous study led to the detection of minor effect QTLs (Palaiokostas et al., 2018) located on  
392 chromosome 3, 20 and 25 explaining 4%, 1.5% and 2% of the total genetic variance respectively. After  
393 chromosome correspondence checking by aligning markers sequences on the European sea bass  
394 reference genome (GCA\_000689215.1) using BLAST (Altschul et al., 1990), chromosome 3 and 20 from  
395 that previous study corresponded to the LG12 and LG6 in our study. The QTL explaining the largest part  
396 of genetic variance was located on the same chromosome (LG12) in both studies. In Asian seabass, only  
397 minor effect QTLs were found, explaining between 1.6% and 2.7% of the phenotypic variance (Wang et  
398 al., 2017). Those QTLs, located on chromosomes 1, 8, 14, 15, 16, 19, 20, 21 and 24 were linked with  
399 potential candidate genes, all involved in stress response. However, using syntenic blocks proposed by

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

405 In back-cross full-sib families, we used a Composite Interval mapping approach that is the most  
406 appropriate for such data. From those analyses, we obtained very high LOD values for the detection of  
407 the QTL on LG12 as well as very wide confidence interval (37.8 cM in NEM10, 9.74 cM in NEM12 and 42.3  
408 cM in SEM8) mainly due to strong linkage-disequilibrium between markers. No QTL was found in the  
409 WEM18 even though it was the family with the highest survival rate. This absence of the QTL detection  
410 may be due to the fact that the WEM18 family did not segregate for resistance QTLs. The mating that  
411 produced those families was designed to produce families that segregate for resistance QTLs (section  
412 2.2), by mating putative Rs (R : resistant ; s : susceptible) sires with putative ss dams. Thus, it is likely that  
413 each parent of this particular WEM18 family carried the resistance QTLs in only one version. As the  
414 global resistance of the family was high, it is rather likely that the WEM18 family could be from a RR sire  
415 x ss dam or from a RR sire x RR dam, thus not segregating for the QTL on LG12.

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

422 Thanks to the high number of individuals in pop A and thus, the high power of the QTL detection, we  
423 were able to refine the location of the QTL on LG12. Both GBLUP and BayesCπ approaches revealed a  
424 strong association between VNN resistance and markers located on LG12 as well as a moderate part of  
425 the total genetic variance explained by this QTL. Even though the QTL on LG12 explained 9.21% of the  
426 total genetic variance, several markers located on the chromosome and detected in pop A seemed to  
427 have a potential interest in MAS. Among them, one marker (LG12\_8815613) detected had the highest  
428 effect on that chromosome and the third highest logBF on LG12. In addition, it had a low MAF (0.13),  
429 which is expected when resistance is not well spread in a population, and which also mechanically  
430 decreases the percentage of variance explained. Thus, this marker had interesting properties to be a  
431 candidate marker to MAS. Its effect on survival across data sets is shown in Figure 5. Interestingly, this  
432 marker had a strong effect on survival both in commercial populations (Figure 5) and in the backcross  
433 families, with an average of 43.3% survival for genotype AA, 72.8% survival for genotype AB and 78.3%  
434 survival for genotype BB. In pop B only, the BB genotyped survived less than the AB genotype, but only  
435 five individuals were BB in pop B, and thus the condition to quantify the effect of the BB genotype in this  
436 population are far from optimal. Although the consistent effect of this marker on six different cohorts  
437 highlights its potential interest, further validation in additional populations or cohorts from the same  
438 population would be of great value. The low observed frequency for allele B in the commercial  
439 populations could be due to a founder effect or maybe from the fact that these populations have been  
440 selected for several traits, among which growth rate for at least 6 generations. As there is a moderate  
441 but negative genetic correlation of VNN resistance with body weight ( $-0.35 \pm 0.14$ ; Doan et al., 2017a),  
442 this selection for growth might have had a negative effect on the frequency of a VNN resistance allele.

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

454 This study produced encouraging results with the discovery of one important QTL involved in VNN  
455 resistance in European sea bass. Even though the part of genetic variance explained by this QTL is limited  
456 (9.21%), further work on the location of the causal mutation and the discovery of diagnostic marker  
457 informing on the resistance genotype of individuals could greatly improve the global sea bass  
458 aquaculture. The best example of the application of MAS in aquaculture is the discovery of one major  
459 QTL involved in the Infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon and  
460 explaining 90% of the genetic variance (Houston et al., 2008; Moen et al., 2009). Since its use by the  
461 salmon aquaculture industry to improve selection based on pedigree information, IPNV outbreaks  
462 occurrence decreased by 75% (Hjeltnes, 2014). This success was due to the identification of one SNP  
463 located on the gene responsible of the resistance and the genotyping of selection candidates to improve  
464 the IPNV resistance in aquaculture farms. In the context of VNN resistance in European sea bass, the  
465 effect of the QTL on LG12, although significant, is smaller. Its use in MAS will have less benefits due to  
466 the smaller part of genetic variance explained. In that situation, genomic selection could be more  
467 appropriate to improve VNN resistance. As shown with the example of marker LG12\_8815613 (Figure 5),  
468 it remains possible that a marker with very large effects on survival would only explain a moderate part  
469 of the genetic variance if the resistance allele is rare in the population studied. The presence of a major  
470 QTL explaining a large part of the genetic variance in VNN resistance is thus still a valid hypothesis and  
471 could be investigated in further work using genome sequence combined with fine mapping approach.

## 472 Acknowledgments

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

482 Declaration of Competing Interest

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

486 Research data

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

491 References

- 492 Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., Misztal, I., 2019. Frequentist p-values for  
493 large-scale-single step genome-wide association, with an application to birth weight in American Angus  
494 cattle. *Genet. Sel. Evol.* 51, 28. <https://doi.org/10.1186/s12711-019-0469-3>
- 495 Allal, F., Griot, R., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin,  
496 T., Poncet, C., Vergnet, A., Cariou, S., Brunier, J., Bruant, J.-S., Peyrou, B., Gagnaire, P.-A., Vandeputte, M.,  
497 2020a. Genotyping of 4 European sea bass families with the DlabCHIP SNPs genotyping array. SEANOE.  
498 <https://doi.org/10.17882/75696>
- 499 Allal, F., Griot, R., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin,  
500 T., Poncet, C., Vergnet, A., Cariou, S., Brunier, J., Bruant, J.-S., Peyrou, B., Lavisse, P., Gagnaire, P.-A.,  
501 Vandeputte, M., 2020b. Axiom DlabCHIP, a 57K SNPs genotyping array for European sea bass  
502 (*Dicentrarchus labrax*): array design and genetic map. SEANOE. <https://doi.org/10.17882/75680>
- 503 Baranski, M., Kettunen Praebel, A., Sommer, A., Haanes, K., Wesmajervi, M., 2010. Major quantitative  
504 trait loci for viral nervous necrosis resistance in atlantic cod, in: *Proceedings of the 9th World Congress*  
505 *on Genetics Applied to Livestock Production: August 1-6, 2010 Leipzig, Germany.* de.
- 506 Boerner, V., Tier, B., 2016. BESSiE: a software for linear model BLUP and Bayesian MCMC analysis of  
507 large-scale genomic data. *Genet. Sel. Evol.* 48, 63. <https://doi.org/10.1186/s12711-016-0241-x>
- 508 Breuil, G., Bonami, J.R., Pepin, J.F., Pichot, Y., 1991. Viral infection (picorna-like virus) associated with  
509 mass mortalities in hatchery-reared sea-bass (*Dicentrarchus labrax*) larvae and juveniles. *Aquaculture* 97,  
510 109–116. [https://doi.org/10.1016/0044-8486\(91\)90258-9](https://doi.org/10.1016/0044-8486(91)90258-9)
- 511 Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses.  
512 *Bioinformatics* 19, 889–890. <https://doi.org/10.1093/bioinformatics/btg112>
- 513 Churchill, G.A., Doerge, R.W., 1994. Empirical threshold values for quantitative trait mapping. *Genetics*  
514 138, 963.
- 515 Dempster, E.R., Lerner, I.M., 1950. Heritability of Threshold Characters. *Genetics* 35, 212–236.
- 516 Doan, Q.K., Vandeputte, M., Chatain, B., Haffray, P., Vergnet, A., Breuil, G., Allal, F., 2017a. Genetic  
517 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 Durantou, 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](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/10.1007/978-1-62703-447-0\\_6](https://doi.org/10.1007/978-1-62703-447-0_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/10.1186/1471-2164-10-368>

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](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., Zrcic, 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

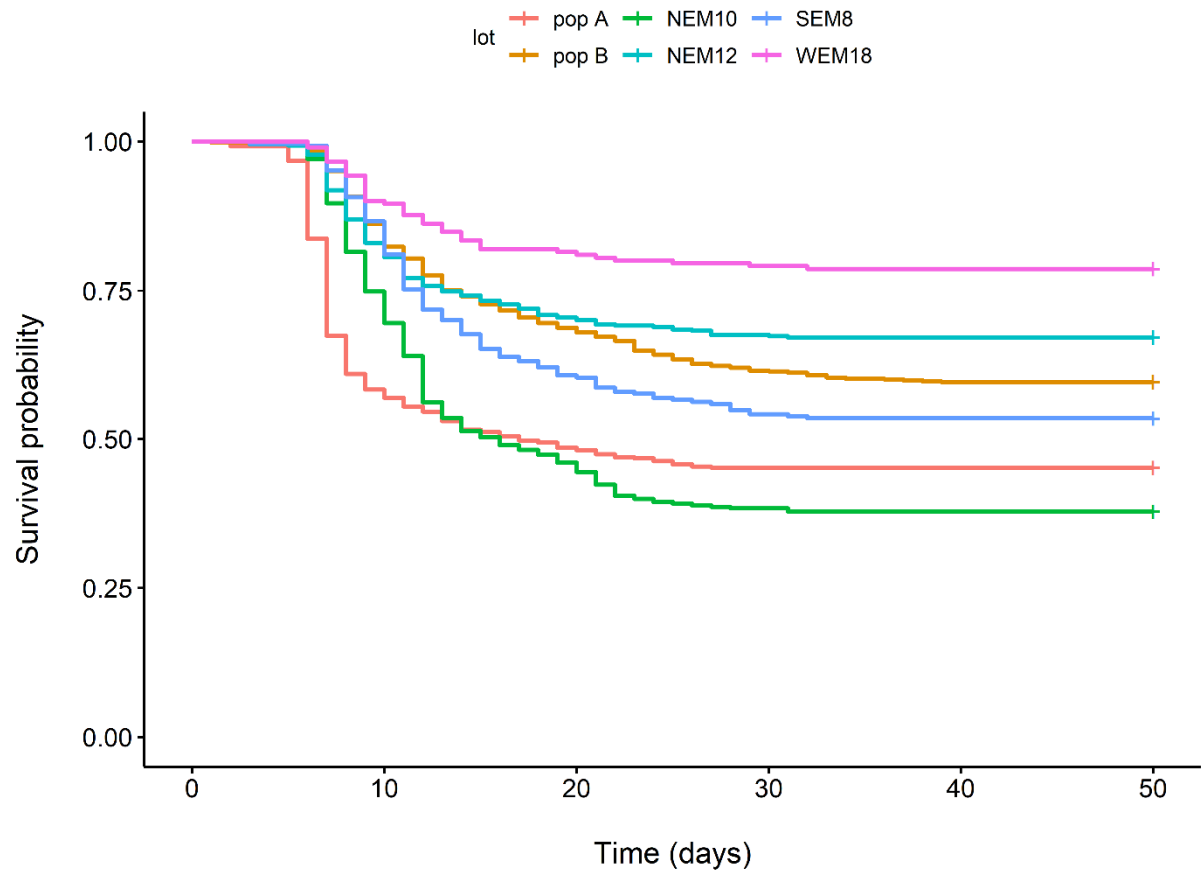


Figure 1 : Evolution of the Kaplan-Meier probability of survival of six different European sea bass populations during their respective VNN infection challenge. Pop A and pop B are commercial populations, and NEM10, NEM 12, SEM8 and WEM 18 are backcross families.

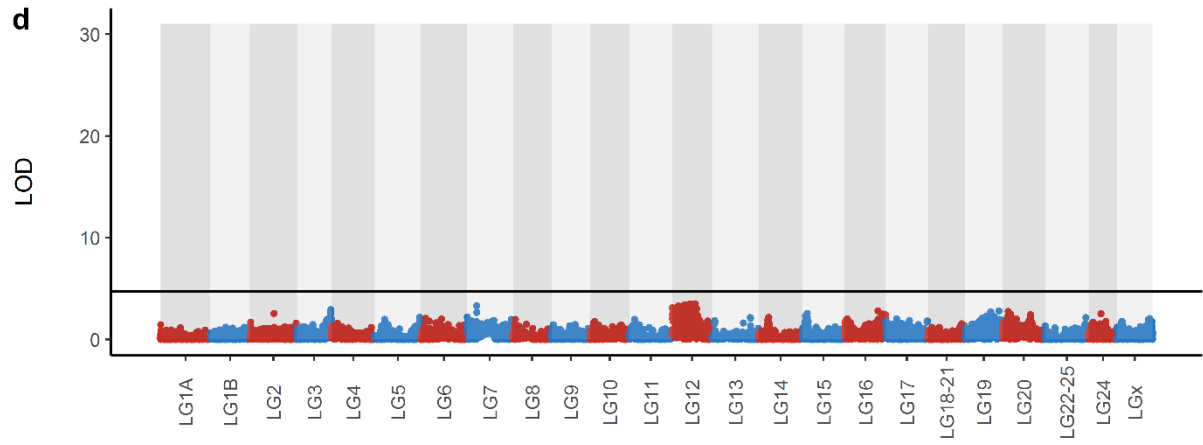
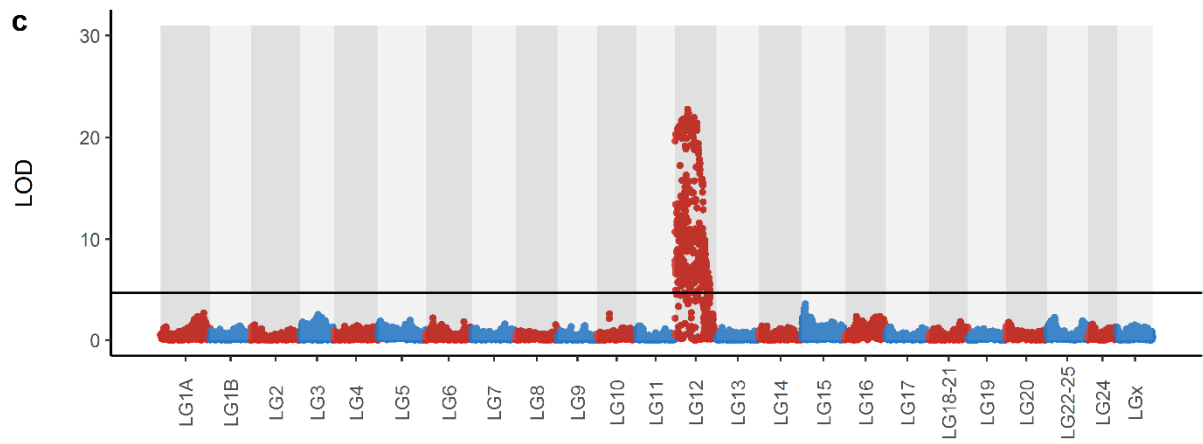
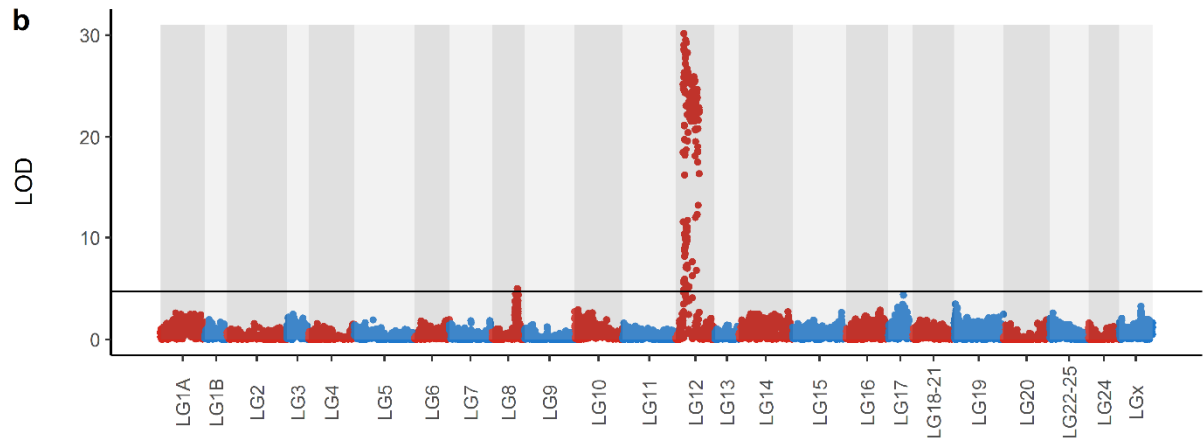
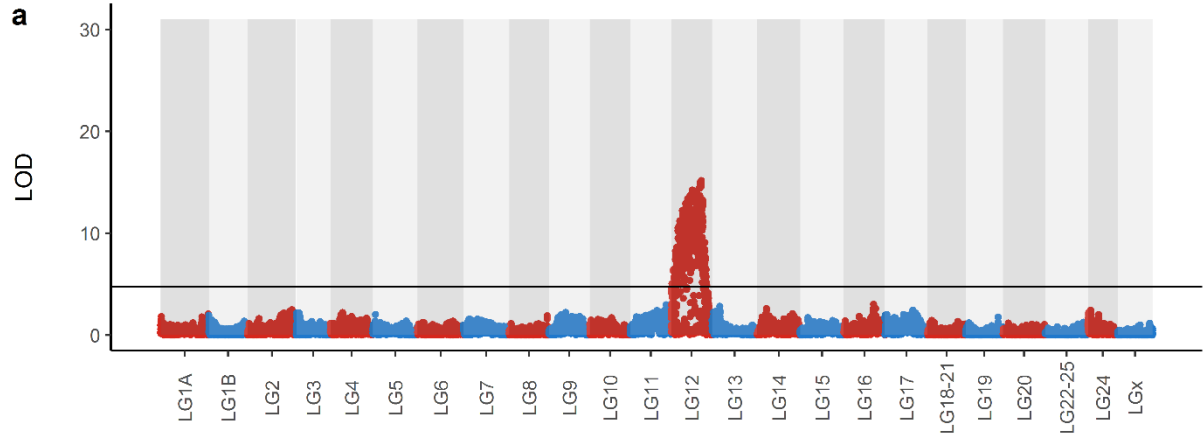


Figure 2: Manhattan plot of the LOD obtained by composite interval mapping for VNN resistance QTLs in European sea bass in the NEM 10 (a), NEM12 (b), SEM8 (c) and WEM18 (b) experimental backcrosses. The horizontal black line represents the 5% genome-wide significance threshold calculated over 1000 permutations.

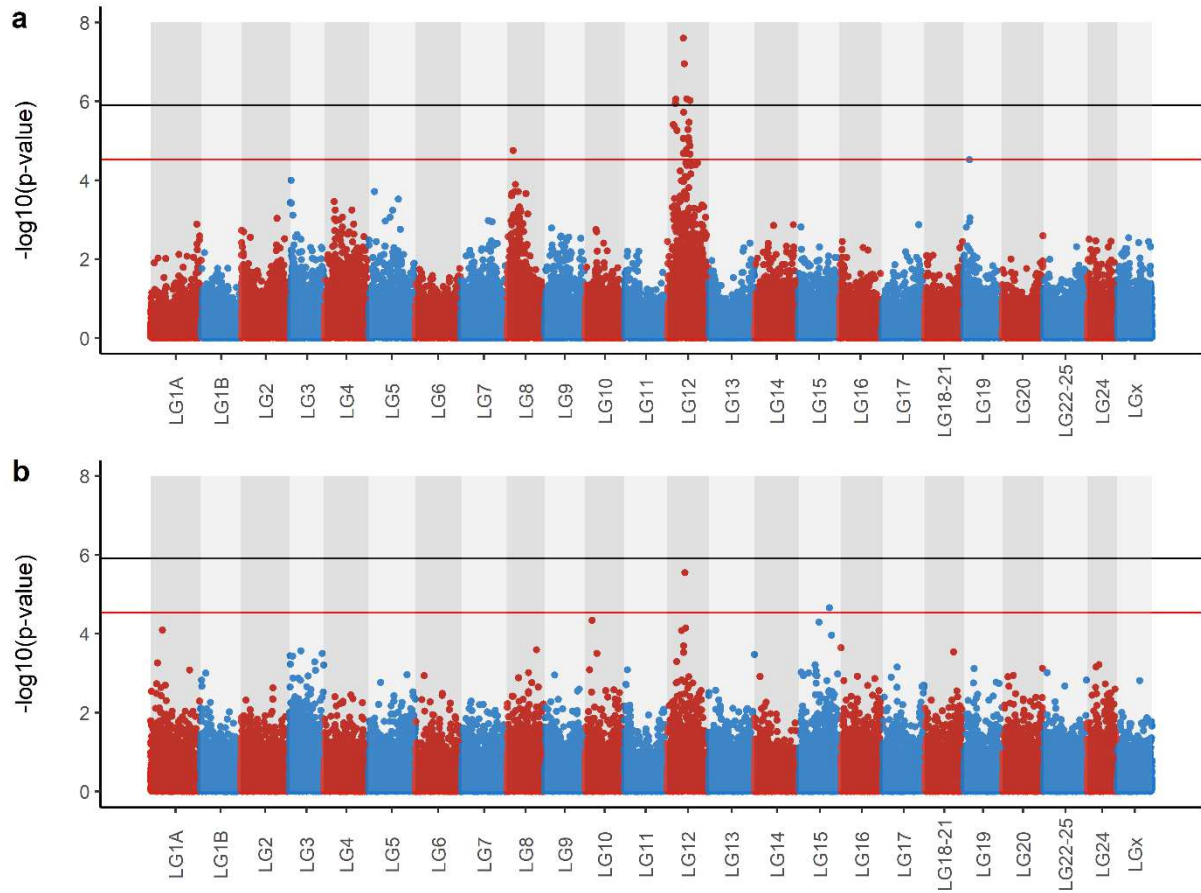


Figure 3: Manhattan plot of  $-\log_{10}(\text{p-value})$  obtained from GWAS for VNN resistance QTLs in commercial European sea bass populations pop A (a) and pop B (b). The horizontal black line represents the genome-wide significance threshold and the red line the chromosome-wide significance threshold calculated with the Bonferroni correction.

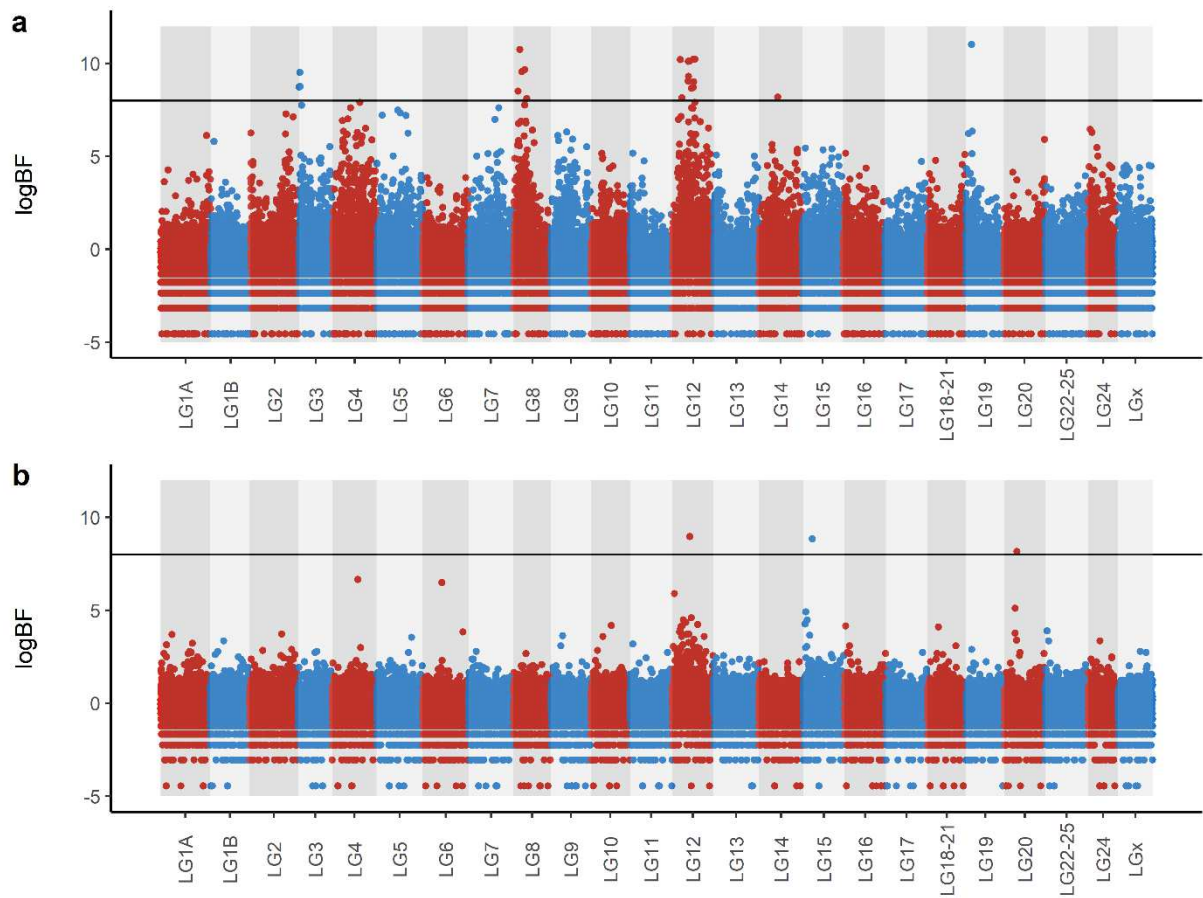


Figure 4: Genome-wide logBF plot for VNN resistance across the genome in the European sea bass populations pop A (a) and pop B (b) using a BayesC $\pi$  model. Horizontal black lines represent the logBF threshold of 8, corresponding to strong evidence for the presence of a QTL.

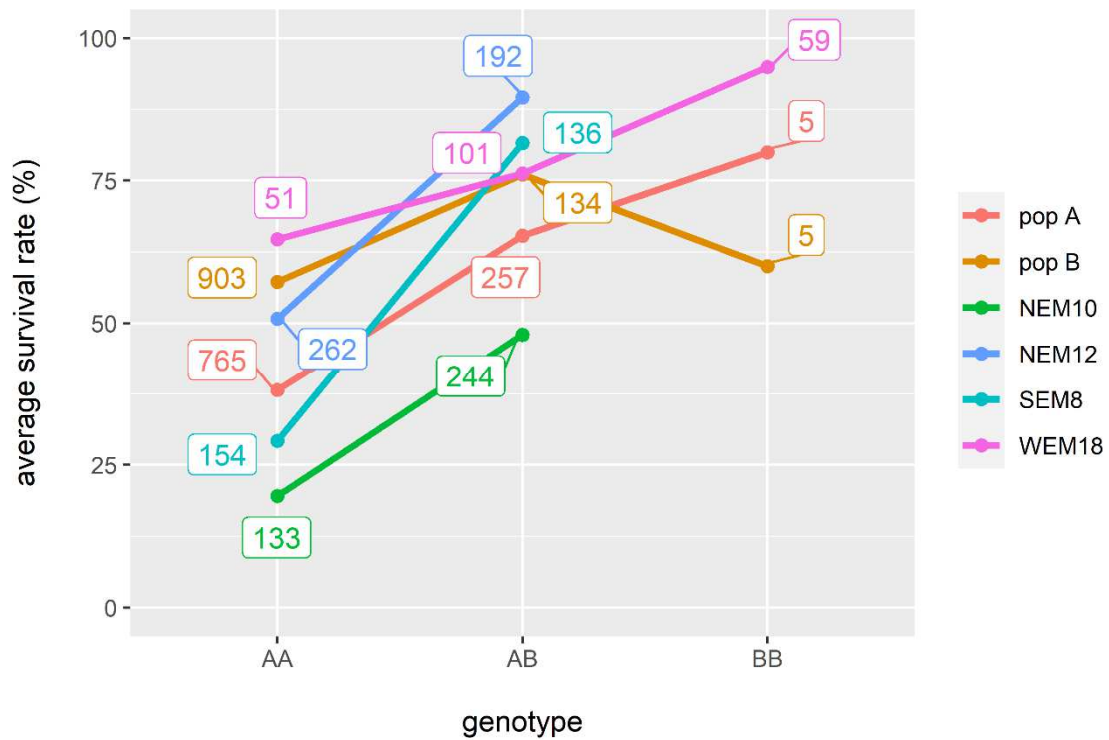


Figure 5: Average survival rate (%) to VNN for each genotype at marker LG12\_8815613 in six cohorts of European sea bass. Each colour corresponds to one data set. The numbers in the boxes are the numbers of individuals per genotype.

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,  $\sigma^2_e$  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