Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing

Christos Palaiokostas† *, Serena Ferarreso‡ *, Rafaela Franch‡, Ross D. Houston† and Luca Bargelloni‡,1

†The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland, United Kingdom,
‡Department of Comparative Biomedicine and Food Science, University of Padova, viale dell’Università 16, 35020 Legnaro, Italy. *Equal contribution

ABSTRACT Gilthead sea bream (*Sparus aurata*) is a species of paramount importance to the Mediterranean aquaculture industry, with an annual production exceeding 140,000 metric tonnes. Pasteurellosis due to the Gram-negative bacterium *Photobacterium damselae* subsp. piscicida (Phdp) causes significant mortality, especially during larval and juvenile stages, and poses a serious threat to bream production. Selective breeding for improved resistance to pasteurellosis is a promising avenue for disease control, and the use of genetic markers to predict breeding values can improve the accuracy of selection, and allow accurate calculation of estimated breeding values (EBV) of non-challenged animals. In the current study, a population of 825 sea bream juveniles, originating from a factorial cross between 67 broodfish (32 sires; 35 dams), were challenged by 30 min immersion with 1 × 10^5 CFU virulent Phdp. Mortalities and survivors were recorded and sampled for genotyping by sequencing. The 2b-RAD sequencing approach was used to generate genome-wide SNP genotypes for all samples. A high-density linkage map containing 12,085 SNPs grouped into 24 linkage groups (consistent with the karyotype) was constructed. The heritability of surviving days (censored data) was 0.22 (95% highest density interval: 0.11-0.36) and 0.28 (95% highest density interval: 0.17-0.4) using the pedigree and the genomic relationship matrix respectively. A genome wide association study did not reveal individual SNPs significantly associated with resistance at a genome wide significance level. Genomic prediction approaches were tested to investigate the potential of the SNPs obtained by 2bRAD for estimating breeding values for resistance. The accuracy of the genomic prediction models (r = 0.38 - 0.46) outperformed the traditional BLUP approach based on pedigree records (r = 0.30). Overall results suggest that major quantitative trait loci (QTL) affecting resistance to pasteurellosis were not present in this population, but highlight the effectiveness of 2b-RAD genotyping by sequencing for genomic selection in a mass spawning fish species.

KEYWORDS 2b-RAD, Aquaculture, Breeding, High-Throughput sequencing

INTRODUCTION Infectious diseases are a major threat to the profitability, sustainability and welfare status of farmed fish production (Yáñez et al. 2014). Gilthead sea bream (*Sparus aurata*) is one of the most important farmed fish in Mediterranean countries, with an annual production of approximately 146,000 metric tonnes (FEAP http://www.feap.info/default.asp?SHORTCUT=582). Pasteurellosis due to *Photobacterium damselae* subsp. piscicida is one of the primary disease problems faced by the sea bream aquaculture industry. High levels of mortality (90 – 100 %) are frequently observed, especially in periods where water temperature rise above 18°C, with larvae and juveniles being most susceptible (Noya et al. 1995;
MATERIALS AND METHODS

Sample collection and preparation

The experimental population used in the present experiment was part of a larger group of juvenile sea bream that were subjected to an experimental challenge with Phdp to estimate heritability of disease resistance as reported (Antonello et al. 2009). Fish were provided by the fish farm Valle Ca’ Zuliani (Monfalcone, Italy). All broodstock fish were originally sampled from wild populations. Fertilized eggs were collected on the same day (year 2006) from natural mass spawning events occurring in four different broodstock tanks, therefore all fish had approximately the same age. Each broodstock tank contained 50-60 fish with a sex ratio 3:1 females:males. Approximately 10,000 eggs were collected, pooled, and kept in a separate tank without any size sorting until 110 days old. All fish were then transferred to the Istituto Zooprofilattico Sperimentale delle Venezie (Legnaro, Italy) for the experimental challenge. Fish were divided into two aerated tanks (A and B) each with 800 l of re-circulating seawater (salinity 35 ppt). Water temperature was maintained at 19°C. After 1 week of acclimation, fish were experimentally infected with a highly virulent strain of Phdp (strain 249/fitto99) as described in Antonello et al. (2009). Mortality was monitored daily for 19 days (Table S1). Mortality levels were nearly identical for both tanks, and only fish from tank A were included in the current study. Fish used in the challenge originated from 32 broodfish (32 sires; 35 dams). As already described (Antonello et al. 2009; Massault et al. 2010), parentage analysis was carried out using a panel of nine microsatellite loci (Table S2).

2b-RAD library preparation and sequencing

A total of 892 2b-RAD libraries (67 parents and 825 juveniles) were constructed by following the protocol reported by Wang et al. 2012 with some modifications (Pecoraro et al. 2016). Template DNA for each individual (500 ng) was digested in 6 µl reaction volume using 1 U AluI at 37°C for 1 h followed by enzyme heat inactivation at 65°C for 20 min. The ligation reaction was performed by combining 5 µl of digested DNA with 20 µl ul of a ligation master mix containing 0.4 µM each of two library-specific adapters (with fully degenerate cohesive ends 5'-NN-3') 0.2 mM ATP (NEB) and 1000 U T4 DNA ligase (CABRU, Arcore, Italy). Ligation was carried out at 16°C for 3 h with subsequent heat inactivation for 10 min at 65°C.

Sample-specific barcodes were designed through a Barcode generator program (http://comailab.genomcenter.ucdavis.edu/index.php/Barcode_generator). 50 µl PCR reactions were prepared containing 12 µl of ligated DNA product, 0.2 µl M of each primer, 0.3 mM dNTP, 5X Phusion HF buffer and 2 U Phusion high-fidelity DNA polymerase (NEB). Each library was PCR amplified using the following conditions: 13 cycles of 95°C for 5 s, 60°C for 20 s and 72°C for 3 s.

Adaptor and primer sequences were those reported in Wang et al. (2012). PCR products were purified using the SPRSelect purification kit (Beckaman Coulter, Pasadena, CA, USA) and quantified through a Qubit 2.0 Fluorometer (Invitrogen). The quality of all amplicon libraries was checked at 1.8% agarose gel. Additionally, the quality of 10% of randomly-selected libraries, was also assessed by running them on an Agilent 2100 Bioanalyzer.

Individual libraries were pooled into equimolar amounts by adopting two different multiplexing strategies for parents (24 libraries / pool) and offspring (48 libraries / pool). The quality of each pool was verified on Agilent 2100 Bioanalyzer. Finally, pooled libraries were sequenced on an Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA) using 50 base single-end sequencing
### Table 1 Gilthead sea bream linkage map

<table>
<thead>
<tr>
<th>LG</th>
<th>No. markers</th>
<th>Length (cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>607</td>
<td>202</td>
</tr>
<tr>
<td>2</td>
<td>602</td>
<td>192</td>
</tr>
<tr>
<td>3</td>
<td>558</td>
<td>183</td>
</tr>
<tr>
<td>4</td>
<td>560</td>
<td>206</td>
</tr>
<tr>
<td>5</td>
<td>554</td>
<td>182</td>
</tr>
<tr>
<td>6</td>
<td>536</td>
<td>167</td>
</tr>
<tr>
<td>7</td>
<td>530</td>
<td>159</td>
</tr>
<tr>
<td>8</td>
<td>522</td>
<td>171</td>
</tr>
<tr>
<td>9</td>
<td>521</td>
<td>138</td>
</tr>
<tr>
<td>10</td>
<td>518</td>
<td>197</td>
</tr>
<tr>
<td>11</td>
<td>508</td>
<td>147</td>
</tr>
<tr>
<td>12</td>
<td>499</td>
<td>142</td>
</tr>
<tr>
<td>13</td>
<td>500</td>
<td>200</td>
</tr>
<tr>
<td>14</td>
<td>489</td>
<td>156</td>
</tr>
<tr>
<td>15</td>
<td>485</td>
<td>162</td>
</tr>
<tr>
<td>16</td>
<td>484</td>
<td>121</td>
</tr>
<tr>
<td>17</td>
<td>480</td>
<td>163</td>
</tr>
<tr>
<td>18</td>
<td>470</td>
<td>164</td>
</tr>
<tr>
<td>19</td>
<td>471</td>
<td>154</td>
</tr>
<tr>
<td>20</td>
<td>461</td>
<td>157</td>
</tr>
<tr>
<td>21</td>
<td>455</td>
<td>134</td>
</tr>
<tr>
<td>22</td>
<td>454</td>
<td>130</td>
</tr>
<tr>
<td>23</td>
<td>452</td>
<td>157</td>
</tr>
<tr>
<td>24</td>
<td>366</td>
<td>115</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12,089</strong></td>
<td><strong>3,899</strong></td>
</tr>
</tbody>
</table>

(v3 chemistry).

**Genotyping RAD alleles**

Quality and adapters trimming of sequenced reads were performed by running a customized script (Pauletto et al. 2016; Pecoraro et al. 2016), obtaining 34-bp long fragments. SNP calling was performed using STACKS v1.23 (Catchen et al. 2011).

For each family/cross, individual genotypes were constructed using components of the STACKS pipeline as follows: i) for each individual, ustacks program was employed for building loci from all QC passed reads using the following parameters: $-m10 - M2 - N3$ for parents and $-m5 - M2 - N3$ for offspring, ii) a catalog of loci unique for all families/crosses was constructed by using all parents’ reads on cstacks program, then iii) each set of parents / offspring per cross was matched separately against such a catalogue (sstacks program) followed by genotype assignment by setting the following parameters on genotypes program: $-c - \text{min\_hom\_seqs}\_7 - -\text{max\_het\_seqs}\_0.05$.

Unique tags created by STACKS were mapped against a draft assembly of Sparus aurata genome (L. Bargelloni, personal communication). Mapping analysis was carried out by means of CLC Genomic Workbench 7.5 with stringent criteria (length fraction = 0.9; similarity fraction = 0.9; nonspecific match handling= ignore).

**Linkage map construction**

Linkage map construction was performed using Lep-Map v2 (Rastas et al. 2013). Quality control was performed for each full sibling family by excluding SNPs with minor allele frequency below 0.05 and those deviating from expected Mendelian segregation ($P < 0.001$). Linkage groups were formed using a minimum LOD threshold value of 8 in the ‘SeparateChromosomes’ module, allowing a maximum distance between consecutive SNPs of 50 cM. Marker order within each linkage group was performed using the ‘OrderMarkers’ module, where the likelihood of marker order is computed by using a hidden Markov model (Rastas et al. 2013). Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function.
15% missing genotypes were removed in order to minimize impact
surviving at the end of the experiment treated as missing values sampled
from corresponding truncated normal distribution with the resulting
values being higher or equal to 20), \( y \) vector of the intercept, \( Z \) incidence matrix relating phenotypes with the random
animal effects, \( u \) is the vector of animal effects \( N(0, \sigma^2_u) \) with either \( A \) corresponding to the pedigree-based relationship matrix or \( G \) the corresponding genomic relationship matrix, and \( \sigma^2_e \) the
additive genetic variance. Finally, \( e \) is a vector of residuals. The
\( G \) matrix was estimated according to VanRaden (2008) using the
kin function of the R/synbreed package (Wimmer et al. 2012) (File
S1). The additive genetic variance was estimated by applying
Markov Chain Monte Carlo (MCMC) using a prior following the
inverse-scaled \( \chi^2 \) distribution (df=5), using 10M iterations out of
which the first 10% were discarded, and values were stored every 1,000
iterations thereafter. Convergence of the resulting posterior
distribution was assessed both visually ( inspecting the resulting
MCMC plots) and analytically with the Geweke diagnostic using
R/boa v1.1.7 (Smith 2007). Heritability for the number of surviving
days was estimated using the following formula:
\[
h^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}
\]
where \( \sigma^2_g \) the estimated additive genetic variance and \( \sigma^2_e \) the
residual variance.

**Genome-wide association analysis (GWAS)**

To test the association between individual SNPs and resistance
to pasteurellosis (measured as surviving days), a GWAS was per-
formed using R/rrBLUP (Endelman 2011). The mixed model ap-
plied (Yu et al. 2006) had the following format:
\[
y = X u + Z \alpha + \epsilon
\]
where \( y \) is the vector of observed phenotypes (surviving days or
overall survival), \( \alpha \) is the vector of unknown marker effects, \( u \) is the
vector of random animal effects \( N(0, \sigma^2_u) \) and \( \epsilon \) is the vector
of residuals. The matrix \( G \) represents the genomic relationship
matrix as described above, and \( \sigma^2_g \) additive genetic variance
estimated using REML. \( X \) and \( Z \) are incidence matrices relating \( y \)
to \( \alpha \) and \( u \), respectively. According to the above model additive
SNP effects are treated as fixed effects, with the inclusion of the
random animal effect to decrease spurious associations due to
(genomic) relationships between the animals (Yu et al. 2006). The
genome-wide significance threshold for the estimated additive
SNP effects was calculated using a Bonferroni correction (0.05 / N),
where N represents the number of QC-filtered SNPs across the
entire genome.

**Genomic Prediction**

A genomic prediction approach was conducted to quantify the ac-
curacy of the breeding values estimated using the SNP markers to
predict the phenotypic trait values (surviving days). SNPs with >
15% missing genotypes were removed in order to minimize impact
of imputed genotypes, since used software cannot handle missing
samples. Missing values of the remaining SNPs were imputed
using R/synbreed (Wimmer et al. 2012). Genomic breeding values
were estimated using rrBLUP, BayesA, BayesB (Meuwissen et al.
2001) and BayesC (Habier et al. 2011) models using the R/BGLR
(Pérez and de Los Campos 2014) software. The above models differ
in regard to the prior distribution of the marker effects. Briefly,
rrBLUP by using the Gaussian distribution, induces homogenous
shrinkage across markers, while in BayesA the usage of a scaled-
t distribution induces marker size effect shrinkage allowing for
variable marker effect sizes. Models BayesB and BayesC also per-
form variable selection, with the difference between the two being
the usage of a scaled-t or a Gaussian prior density respectively
(Meuwissen et al. 2001; de los Campos et al. 2013). To compare the
accuracy of genomic EBVs to the pedigree-based EBVs, pedigree-
based BLUP (PBLUP) (Henderson 1975) was applied to calculate
breeding values using the same software. The general form of the
fitted models was the following:
\[
y = \eta + \epsilon,
\]
where \( \eta \) the linear predictor \( \eta \) in the case of rrBLUP, BayesA, BayesB
and BayesC had the following general form:
\[
\eta = 1u + X_1 \beta_1,
\]
where \( u \) the animal random effect vector \( N(0, \sigma^2_u) \) with the
matrix \( A \) representing the pedigree estimated relationship ma-
trix. The parameters of the above models were estimated through
MCMC (110,000 iterations; burn-in: 10,000; thin: 100).

Assessment of the accuracy of breeding value predictions was
done according to the following procedure. The data set was
randomly split into a training set (n = 578 animals) and a valida-
tion set (n = 200). The above was repeated 100 times with the obtained
prediction accuracies being adjusted for the trait heritability for
each tested model. The genomic estimated breeding values (GBEV)
for each replicate of the validation data set were estimated as:
\[
GBEV = Xu,
\]
where \( X \) the incidence matrix relating GEBV with SNP geno-
types and \( u \) the vector of estimated SNP effects from the corre-
spending training data set. The accuracy of the estimated GEBV
was approximated as:
\[
\tau = \frac{GBEV, y}{h},
\]
where \( y \) the vector of recorded phenotypes and \( h \) the square
root of the heritability. In all tested scenarios the heritability esti-
mated using the genomic relationship matrix was used. Reported
accuracies for each tested model refer to the mean accuracy of the
above-mentioned 100 replicates of validation data sets. In order to
test the predictive ability of varying SNP densities, the above
procedure was followed using (i) SNPs spaced > 1 cM apart (2,614
SNPs) or (ii) SNPs spaced > 5 cM apart (705 SNPs) on the linkage
map.

**Data Availability**

Raw reads were deposited in EBI repository under project ID
SRP081498. Table S1 contains the phenotypic data. Table S2 con-
tains the pedigree. Table S3 and S4 contain summary of obtained
reads for parental and offspring samples respectively. Table S5
contains the location of the linkage map SNPs. Table S6 contains
the genotypic data. File S1 contains customed R script used for
genomic prediction.
RESULTS

Disease challenge
The challenged population consisted of 75 full sibling families with a mean family size of 10, originating from a factorial cross between 67 broodfish (32 sires; 35 dams). The largest full-sib family consisted of 114 animals, while the smallest had only 2 animals (3 full-sib families). The overall survival at the end of the pasteurellosis challenge was 4.7%. Observed mortality levels showed three distinct peaks on day 7 (10.4 % loss), day 11 (14 % loss) and day 15 (5.7 % loss) followed by a steady reduction in daily mortality rate (Figure 1).

Genotyping RAD alleles
The mean number of raw reads was 8.76 million (M) and 4.51 M while the number of reads passing QC was 7.74 M (88 %) and 3.69 M (81 %) for parents (Table S3) and offspring (Table S4), respectively. The STACKS catalogue consisted of 202,598 unique 2b-RAD loci, of which 73,876 contained at least one SNP in the parents (EBI repository SRP081498). To confirm the identity of loci created by STACKS, the 202,598 tags of the catalog were mapped against a Sparus aurata draft reference genome assembly (unpublished data). A high percentage of 2b-RAD loci were successfully mapped, with 93.5 % of tags showing a unique match to the reference genome. In order to maximise the number of informative SNPs and minimise the amount of missing or erroneous data, RAD-tags that were retrieved in at least 75% of the samples, and that carried only one or two SNPs were retained. 48 animals with missing data > 30% were excluded from subsequent analysis. A total of 21,974 putative SNPs were finally used for construction of the genetic map from 777 disease challenged offspring (genotypic missing data < 30%) and their corresponding 67 parents.

Linkage Map
The linkage map consisted of 12,085 SNPs that were grouped into 24 linkage groups, in accordance with gilthead sea bream karyotype, with a total map length of 3,899 cM (Table 1, Figure 2; Table S5). The remaining SNPs (9,889) either failed to pass QC filters, or were not placed on the resulting linkage groups during mapping, and these were discarded. The female and male maps were comparable with total lengths of 3,822 cM and 4,010 cM respectively. The number of SNPs per chromosome ranged between 366 to 607 (mean = 503; s.d = 53), while linkage group length ranged between 115 and 202 cM (mean = 162; s.d = 26). The correlation between number of SNPs and corresponding chromosome map length was 0.74 (n = 24 linkage groups).

Heritability estimation and Genome wide association analysis (GWAS)
The heritability of surviving days (censored data) was 0.22 (95% highest density interval: 0.11-0.36) and 0.28 (95% highest density interval: 0.17-0.4) using the pedigree and the genomic relationship matrix respectively of the 777 disease challenged offspring. No SNPs surpassed the Bonferroni-corrected genome-wide significance threshold ($P = 4.1x10^{-6}; \alpha = 0.05$; Table S6). The SNPs with the lowest p-values ($P < 10^{-3}$) were located in linkage groups 1 - 3, 10, 17, 20 and 21 (Figure 3).

Genomic Prediction
Genomic prediction was tested as a means of obtaining breeding values, and compared to prediction using a pedigree-based approach (File S1). The prediction was conducted using genotype information from 11,239 SNP markers (passed QC filters set for genomic prediction) for the 777 disease challenged animals that were randomly split in training (n=578) and validation (n=200) datasets. The application of all the genomic prediction models resulted in higher accuracies than those achieved using pedigree-based BLUP (Table 2). Prediction accuracy with PBLUP was 0.3, and versus 0.38 to 0.46 for the genomic prediction models, with highest accuracy being observed using the BayesA method. Prediction accuracy dropped when more sparse SNP marker datasets were used with the last scenario (utilizing SNPs > 5cM apart) giving similar accuracies with the ones obtained by using the pedigree-based BLUP. In the scenario of utilizing only SNPs located more than 1 cM apart on the linkage map (2,614 SNPs) accuracies ranged between 0.3 and 0.36, with the highest accuracy obtained using rrBLUP. For the dataset utilizing only SNPs located more than 5 cM apart (705 SNPs) accuracies ranged between 0.29 - 0.31 with highest accuracies obtained using rrBLUP and BayesA.

DISCUSSION
Gilthead sea bream (Sparus aurata) is a farmed species of paramount importance for Mediterranean aquaculture. While vaccines can offer some protection against pasteurellosis, the low immune competence observed in larval and juvenile stages renders this protection only temporary (Antonello et al. 2009). Breeding for
improved genetic resistance offers an additional and complementary tool to combat losses due to this disease. While traditional family-based selective breeding is applied in sea bream, it cannot utilise within-family genetic variation in the trait. Applying genomic information into selective breeding schemes raises the possibility of selecting directly for favourable alleles at major QTL (marker-assisted selection) or incorporating all markers in the prediction of breeding values (genomic selection). As such, genomics-enabled breeding can expedite the rate of genetic gain, and can potentially reduce the need for yearly trait recording. However, to enable these benefits, substantial genomic resources are typically required (e.g. a high density SNP genotyping platform), which sea bream is currently lacking. This is likely to change in the near future as the reference genome sequence and associated genomic tools/data become available. In the meantime, RAD sequencing and similar techniques can readily be applied to generate genome-wide SNP marker datasets even in the absence of such genomic resources (Baird et al. 2008).

High-density SNP linkage maps have been constructed for several aquaculture species, and are useful for both QTL positioning and reference genome assembly (Gonen et al. 2014; Palaiokostas et al. 2013b,a, 2015b). The most recent linkage map of sea bream consists mainly of microsatellites (Tsigenopoulos et al. 2014), lacking the necessary resolution for successful implementation of GWAS and genomic prediction. In the current study, we present the first high-density linkage map for this species, consisting of 12,085 SNPs on 24 linkage groups, which is consistent with the karyotype. The genetic map presented here spans 3,899 cM, while the previous map of (Tsigenopoulos et al. 2014) has a total length of 1769.7 cM, which may reflect the larger number of markers used in this study. This trend of increase in map distance with increased marker density was observed with previous sea bass (Dicentrarchus labrax) linkage maps (Chistiakov et al. 2005, 2008; Palaiokostas et al. 2015b).

The estimated heritability of resistance was moderate (0.22 and 0.28 for the different models) compared to those previously reported for disease resistance traits in various aquaculture species (Odegård et al. 2011). Nonetheless, successful implementation even in the case of low heritability traits in breeding programs is still possible, as demonstrated in livestock (Heringstad et al. 2003). Also, since heritability of mortality traits is frequency dependent, with maximal values reported at intermediate mortality levels (Bishop and Woolliams 2014), the low survival rate in the current study may have resulted in an underestimate, and analysis of additional challenge and field data is merited.

The GWAS results pointed to a polygenic or oligogenic genetic architecture for resistance to pasteurellosis, with no genome-wide significant QTL identified, with the lowest P values indicative of putative suggestive QTL on linkage groups 1 - 3, 10, 17, 20 and 21. Unfortunately, the lack of an integrated linkage map for sea bream does not allow direct comparison of the previously described resistance QTL (Massault et al. 2010) with the current study. Furthermore, no major QTL were identified in the above study, with the largest QTL for surviving days explaining approximately 4 % of the phenotypic variance. The fact that genomic prediction using all tested models gave similar prediction accuracies, a phenomenon often observed in the study of polygenic traits (Meuwissen et al. 2001; Kizilkaya et al. 2010) seem to support the hypothesis of polygenic resistance. Nevertheless, since the selected priors will influence the output of the Bayesian models (Gianola 2013), interpretation of genetic architecture based on these results should be treated with caution. Additionally, the moderate sample size in the current study is likely to limit the statistical power to detect small to medium size effect QTL. Further, the low mortality level precludes effective estimation of genetic parameters for survival per se, and the genetic correlation between survival time (days to death) and overall survival is unknown.

The results from the genomic prediction approach were encouraging for practical implementation of selective breeding for genetic resistance in sea bream, with the genomic prediction models outperforming traditional BLUP. The advantage of the genomic-based models was retained also when only SNPs at 1 cM location intervals were used, which may be useful for reducing genotyping costs and improving cost-effectiveness. However, due to the limited number

**Table 2 Genomic prediction accuracies for surviving days**

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy*</th>
<th>Accuracy SNPs 1cM apart</th>
<th>Accuracy SNPs 5cM apart</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBLUP</td>
<td>0.3 (se)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rrBLUP</td>
<td>0.44 (se ± 0.04)</td>
<td>0.36 (se ± 0.03)</td>
<td>0.31 (se ± 0.04)</td>
</tr>
<tr>
<td>BayesA</td>
<td>0.46 (se ± 0.03)</td>
<td>0.35 (se ± 0.03)</td>
<td>0.31 (se ± 0.04)</td>
</tr>
<tr>
<td>BayesB</td>
<td>0.38 (se ± 0.03)</td>
<td>0.30 (se ± 0.04)</td>
<td>0.29 (se ± 0.03)</td>
</tr>
<tr>
<td>BayesC</td>
<td>0.44 (se ± 0.04)</td>
<td>0.35 (se ± 0.04)</td>
<td>0.29 (se ± 0.03)</td>
</tr>
</tbody>
</table>

*Analysis included 12,085 SNPs

**Figure 3** A Manhattan plot highlighting the association between individual SNPs and surviving days. (B). A QQ plot showing the relationship between the observed and expected –log(P) values for each individual SNP and surviving days.
of families in the current study, the training and validation sets contain closely related animals, which will increase the accuracy of prediction. Additional testing of genomic prediction at varying marker densities on separate and preferably larger populations would be required to ascertain the appropriate density for commercial application of genomic selection. The benefit of genomic prediction over a pedigree-based approach is likely to be due to the ability to capture within-family genetic variation. In mass-spawning species such as sea bream, in which family size and structure is difficult to control, this approach is likely to be particularly advantageous. Overall, the current study demonstrates that SNP markers generated via 2b-RAD are effective at capturing the genetic variation in a complex trait in a sea bream breeding population. This approach is likely to be useful in other species with less-developed genomic tools, and provides further evidence that incorporation of genomic selection is likely to result in significant improvement in selection accuracy and genetic gain compared to traditional family selection in aquaculture breeding.

Conclusion

2b-RAD sequencing was applied to investigate genetic resistance to pasteurellosis in gilthead sea bream. The SNP data generated were applied to create the first high-density linkage map for sea bream. Only suggestive QTLs were detected, implying that resistance to pasteurellosis has an oligogenic-polygenic architecture for the studied population. Genomic prediction using the 2b-RAD genotype data was effective, with substantial improvement in prediction accuracy over the pedigree-based model. This highlights the utility of genotyping by sequencing for genomic prediction of disease resistance in aquaculture species, and its potential to apply genomic selection in commercial breeding programs.

ACKNOWLEDGEMENTS

This work is part of the project “Genomic technologies for the 21st century aquaculture industry: an integrated approach to improve fish growth and health (GenTechAqua)”, granted to LB by the Fondazione Cassa di Risparmio Padova e Rovigo in the framework of the call “Progetti di Eccellenza 2009-2010”. CP and RH are supported by funding from the European Union’s Seventh Framework Programme (FP7 2007-2013) under grant agreement no. 613611 (FISHBOOST). RH is supported by the Bioscience and Biotechnology Research Council (BBSRC) funds BB/J004235/1 and BB/J004324/1.

AUTHOR CONTRIBUTIONS

SF and RF carried out DNA extractions, 2bRAD library preparation and sequence data processing. LB conceived the study, contributed to designing the experiments, analysing the data. CP and RH designed the quantitative genetic analyses. CP conducted the genetic analysis. All authors contributed to drafting the ms. bibliographyexample-bibliography

LITERATURE CITED


Endelman, J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP.


Houston, R. D., J. B. Taggart, T. Cezard, M. Bekaert, N. R. Lowe, A. Downing, R. Talbot, S. C. Bishop, A. L. Archibald, J. E. Bron,


Noya, M., B. Magaríños, and J. Lamas, 1995 Interactions between peritoneal exudate cells (PECs) of gilthead sea bream (Sparus aurata) and Pasteurella piscicida. A morphological study. Aquaculture 131: 11–21.


