GSE **G**enetics **S**election **E**volution

RESEARCH ARTICLE   **Open Access**

CrossMark

A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*)

Christos Palaiokostas[1,2], Michaël Bekaert[1], John B. Taggart[1], Karim Gharbi[3], Brendan J. McAndrew[1], Béatrice Chatain[4], David J. Penman[1*] and Marc Vandeputte[4,5]

Abstract

Background: European sea bass (*Dicentrarchus labrax*) is one of the most important farmed species in Mediterranean aquaculture. The observed sexual growth and maturity dimorphism in favour of females adds value towards deciphering the sex determination system of this species. Current knowledge indicates the existence of a polygenic sex determining determination system that interacts with temperature. This was explored by restriction-site associated DNA (RAD) marker analysis in a test panel of 175 offspring that originated from a factorial cross between two dams and four sires from a single full-sib family.

Results: The first high-density single nucleotide polymorphism (SNP) based linkage map for sea bass was constructed, consisting of 6706 SNPs on 24 linkage groups. Indications for putative sex-determining QTL (quantitative trait loci) that were significant at the genome-wide threshold were detected on linkage groups 6, 11 and 18 to 21, although a genome-wide association study (GWAS) did not identify individual significant SNPs at a genome-wide threshold. A preliminary genomic prediction approach that tested the efficiency of SNP-based selection for female sea bass showed a slight advantage compared to traditional pedigree-based selection. However, when the same models were tested on the same animals for selection for greater length, a clear advantage of the SNP-based selection was observed.

Conclusions: Overall, the results of this study provide additional support to the polygenic sex determination hypothesis in sea bass. In addition, identification of sex-ratio QTL may provide new opportunities for sex-ratio control in sea bass.

Background

Both genetic and environmental factors are involved in the sex determination of various fish species, with sex in some species influenced by both factors [1, 2]. Both XX/XY male heterogametic and WZ/ZZ female heterogametic sex-determining systems exist in fish and, unlike in mammals, individuals with YY and WW genotypes are viable in most fish species tested [3, 4]. The genetic factors that underlie sex determination can range from a single gene to a few sex-determining quantitative trait loci (QTL) or even a complex combination of a large number of genes, as for polygenic traits. Understanding sex determination systems in fish has direct commercial applications, since several commercially important traits, such as growth and age at maturity, are sex-specific in a wide variety of aquaculture fish species.

Dicentrarchus labrax (European sea bass) is a highly valuable commercial aquaculture species in Europe, with more than 148 000 t produced in 2014 [5]. The production cycle of *D. labrax* is between 18 and 24 months, with a market size of about 350 to 400 g. Under aquaculture conditions, the percentage of males is usually very high (70 to 90 %), while current evidence suggests that, in wild populations, the sex-ratio is generally balanced [6]. Males tend to grow more slowly than females and lower body weights at harvest time (up to 40 % less) have been reported compared to those of identical cohorts of females [7, 8]. This dimorphism can be interpreted as an advantage to produce large females in a mass-spawning species, as observed in the Atlantic silverside *Menidia menidia* [9].

* Correspondence: d.j.penman@stir.ac.uk
[1]Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK
Full list of author information is available at the end of the article

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68

Page 2 of 10

The *D. labrax* sex determination system depends both on genetic factors and environmental effects, with temperature being considered as the most important environmental effect under hatchery conditions [10, 11]. Studies on induced gynogenesis in *D. labrax* have indicated that the mechanism of sex determination is not explained by a simple monofactorial system with either male or female homogamety [12]. The sex ratio of offspring from masculinised females is not female-biased, which rules out straightforward XX/XY or WZ/ZZ systems [13]. However, Francescon et al. [14] reported that the sex ratio of progeny from meiogynogenetic females is skewed in favour of females.

Environmental temperature influences sex ratio in *D. labrax*, and temperatures above 17 °C during early development (before 60 days post-fertilization) favour the development of males [10]. Higher temperatures (~21 °C), which are typically used during the larval and early juvenile stages in aquaculture hatcheries, are thought to masculinise fish that would have remained as females at lower temperatures. However, unlike what is observed in reptiles, there is no known temperature regime that produces 100 % males or 100 % females in *D. labrax*. In addition, Diaz et al. [15] reported that the growth rate during the stages prior to sex differentiation is linked to the sex ratio, so that faster-growing fish are more likely to be females.

No major sex-determining gene or genetic markers associated with sex have been identified in *D. labrax* as far as we are aware. Strong parental effects, as well as genotype-temperature interactions can modulate the sex ratio in *D. labrax*: the proportion of females resulting from individual crossings may range from 1 to 70 % [10, 16]. Vandeputte et al. [17] provided the first evidence for a polygenic sex-determining system in this species, based on an analysis of between-family variation in sex ratio.

Restriction-site associated DNA (RAD) sequencing is a reduced-representation sequencing platform that exploits high-throughput sequencing methodologies, while the use of barcodes allows multiplexing of samples [18]. This technique can be used for the rapid discovery of thousands of single nucleotide polymorphisms (SNPs) by sequencing parts of the genome at high depth, which offers the possibility to construct high-density linkage maps in a cost-efficient manner. Sex-determining regions have already been identified in *Danio rerio* (zebrafish), *Oreochromis niloticus* (Nile tilapia) and *Hippoglossus hippoglossus* (Atlantic halibut) using RAD-seq [19–22]. In this study, we used RAD sequencing to identify SNPs in F_2 crosses of *D. labrax* that originated initially from an F_0 cross between two families with divergent sex ratios. First, a high-density linkage map was constructed, then quantitative trait loci (QTL) mapping and a genome-wide

association study (GWAS) were performed. We also carried out a preliminary genomic prediction approach to test the potential of SNP-based selection for increasing either female ratio or body length.

Methods
Sample collection and preparation
The fish used in this study originated from an F_2 cross population of hatchery-reared *D. labrax*. Three F_0 males and three F_0 females, which were the offspring of wild West-Mediterranean *D. labrax*, were mated by artificial fertilization in a factorial cross to simultaneously produce eight families (one family was lost), which were reared in common garden conditions; the overall sex ratio in the F_1 offspring was 43.8 % females. F_1 males and females were chosen from one family, which was a cross between the F_0 female that produced the lowest proportion of females (26.6 %) and the F_0 male that produced the highest proportion of females (58.2 %). Two F_1 females (Dams 1 and 2) were each crossed to four F_1 males (Sires 1, 2, 3 and 4) to produce eight families (Table 1). Temperature was kept at 16 °C between 5 and 15 days post-fertilisation (dpf) and at 18 °C between 16 and 42 dpf, then it was increased to 23 to 25 °C between 48 and 98 dpf and finally decreased to an average of 21 °C until the end of the experiment. This protocol was expected to minimise any temperature effect on sex-ratio [11]. Offspring from each female (four paternal half-sib families for each of the two female parents) were reared separately until accurate sexing by visual inspection of the gonads was possible at 16 months of age at which time the fish were sacrificed, weighed, sexed and their body length was measured. A fin sample from each fish was collected and stored in ethanol at room temperature. In total, 175 F_2 fish (88 males and 87 females) plus the six F_1 parents and the F_0 female were used for RAD sequencing (no material from the F_0 male was available).

RAD library preparation and sequencing
DNA was extracted from fin samples using the REALPure genomic DNA extraction kit (Durviz S.L.) and treated with RNase. Each sample was quantified by spectrophotometry (Nanodrop), quality assessed by agarose gel electrophoresis, and diluted to a concentration of 50 ng/μL in 5 mmol/L Tris, pH 8.5.

The RAD library was prepared as originally described in Baird et al. [18] and comprehensively detailed in Etter et al. [23], with the minor modifications reported in Houston et al. [24]. Details on the RAD-specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are in Baxter et al. [25]. Briefly, each sample (0.72 μg parental DNA/0.24 μg offspring DNA) was digested at 37 °C for 40 min with

Palaiokostas et al. Genetics Selection Evolution (2015) 47:68

Page 3 of 10

Table 1 Family summary, descriptive statistics and testing of deviations from equal sex ratio

Dam	Sire	Family	Contribution	Offspring males	Offspring females	Average weight in g (±se)	Average length in mm (±s.e)	P-value
Dam 1	Sire 1	1	7.4 %	4	9	194.2 (58.7)	256.5 (20.9)	0.27
Dam 1	Sire 2	2	10.8 %	14	5	179.3 (34.1)	249.5 (15.3)	0.066
Dam 1	Sire 3	3	9.7 %	5	12	181.9 (65.2)	248.6 (31.4)	0.15
Dam 1	Sire 4	4	17.7 %	17	14	178.2 (48.2)	248.7 (19.9)	0.719
Dam 2	Sire 1	5	9.1 %	5	11	148.1 (37)	231.5 (16.1)	0.21
Dam 2	Sire 2	6	12.6 %	14	8	150.2 (54.9)	233.0 (26.8)	0.28
Dam 2	Sire 3	7	10.9 %	3	16	170.0 (67.5)	240.2 (28.3)	0.0059[**]
Dam 2	Sire 4	8	21.7 %	25	13	160.5 (52.8)	239.6 (24.7)	0.074

[**] $\alpha = 0.01$; se = standard error

the high fidelity restriction enzyme *Sbf*I that recognises the CCTGCA|GG motif (New England Biolabs; NEB) using 6 U *Sbf*I per µg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of about 1 µg DNA per 50 µL reaction volume. The samples (12 µL final volume) were then heat-inactivated at 65 °C for 20 min. Individual specific P1 adapters, each with a unique 5 bp barcode (see Additional file 1: Table S1), were ligated to the *Sbf*I digested DNA at 22 °C for 60 min by adding 1.8/0.6 µL (parental/offspring DNA samples respectively) 100 nmol/L P1 adapter, 0.45/0.15 µL 100 mmol/L rATP (Promega), 0.75/0.25 µL 10× Reaction Buffer 2 (NEB), 0.36/0.12 µL T4 ligase (NEB, 2000 U/µL) and reaction volumes made up to 45/15 µL with nuclease-free water for each parental/offspring sample. After heat-inactivation at 65 °C for 20 min, the ligation reactions were slowly cooled to room temperature (over 1 h), then combined in appropriate multiplex pools (see Additional file 1: Table S1). Shearing (Covaris S2 sonication) and initial size selection (250 to 550 bp) by agarose gel separation [24] was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation and library amplification, exactly as in the original RAD protocol [18, 23]. 150 µL of each amplified library (14 to 16 PCR cycles depending on library) was size-selected (about 350 to 650 bp) by gel electrophoresis [24]. Following a final gel elution step into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen), 16 libraries (7 animals in the parental library, 12 animals in each of 15 progeny libraries) were sent to the Edinburgh Genomics facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing. Libraries were accurately quantified by qPCR (Kapa Library) and run in four lanes of an Illumina HiSeq 2000, using 100 base paired-end reads (v3 chemistry). Raw reads were processed using RTA 1.12.4.2 (Illumina). Reads were deposited at the EBI Sequence Read Archive (SRA) study ERP004018.

Genotyping RAD alleles
Reads of low quality (i.e., with a quality score less than 30, while the average quality score was equal to 37), that

lacked the restriction site or had ambiguous barcodes were discarded. Retained reads were sorted into loci and genotypes using Stacks software 1.02 described in Catchen et al. [26] The likelihood-based SNP calling algorithm [27] implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals and statistically differentiates true SNPs from sequencing errors. Reads were aligned to a draft assembly of the sea bass genome (dicLab v1.0c June 2012) using Bowtie 2 [28] and the generated SAM files were passed to the Stacks wrapper program 'ref_map.pl' to curate RAD loci and call SNPs. Values for the major Stacks parameters were as follows: minimum stack depth (m) = 30; distance between stacks (M) = 2; distance between catalog loci (n) = 1.

Parentage assignment – general statistics
Parentage assignment was performed with Vitassign V8-5.1 [29] using 200 SNPs and with R/hsphase using all discovered SNPs and allowing for a maximum genotyping error of 3 % [30]. R v.3.0.1 was used for chi-square tests to detect significant deviations from the equal sex ratio and for calculating correlations between phenotypic sex and weight or length. SAS-GENMOD was used to test for sire and dam effects (and their interaction) on sex, using a probit link function and a binomial distribution [31].

Construction of the linkage map
The linkage map was constructed with Lep-Map [32]. SNPs with a minor allele frequency (MAF) less than 0.05 and SNPs that deviated from the expected Mendelian segregation ($P < 0.001$) were excluded. Linkage groups were formed with a minimum LOD value of 10 using the SeparateChromosomes module of Lep-Map. SNPs within each linkage group were ordered by applying the OrderMarkers module. Map distances were estimated in centiMorgans (cM) using the Kosambi mapping function.

Heritability estimation
Heritability of phenotypic sex was estimated on the liability scale using R/MCMCglmm [33]. The observed

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68

Page 4 of 10

phenotype was considered to follow a binomial distribution. The probit link function was used as follows:

$$y_i \sim B\left(\text{probit}^{-1}(l_i)\right),$$

where y_i is the observed phenotype coded as a binary trait and l_i is the latent variable. The latent variable was modelled as follows:

$$l_i = \mu + u_i + e_i$$

where μ is the intercept, u_i animal random effect that follows $\sim N(0, A\sigma_g^2)$, with A the pedigree-based relationship matrix and σ_g^2 the additive genetic variance, and e_i the residual error. The additive genetic variance was estimated by Markov chain Monte Carlo (MCMC) estimation, using a prior following the χ^2 (1 df) distribution (10 000 000 iterations; 1 000 000 burn-in; 500 thin).

Heritability was estimated using the following formula:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 + 1},$$

where σ_g^2 is the previous estimated additive genetic variance and σ_e^2 is the residual variance that was fixed to 1 due to the identifiability issue with binary data [33].

QTL mapping

Sex-determining QTL were identified by Haley-Knott regression mapping using GridQTL [34, 35]. A half-sib model and a F_2 sib regression analysis model of the entire pedigree were used, allowing for the existence of one QTL at tested intervals that were 1 cM apart.

The models used had the following general form:

$$y_{ij} = m_i + \sum_{z=1}^{z=n} \mu_z \times p_z + e_{ij},$$

where y_{ij} is the phenotypic sex of individual j belonging to full/half sib family i, with males coded as 1 and females as 2, m_i is the F2 – half sib family specific intercept, μ_z is the effect of marker z, n is the number of possible QTL genotypes of the tested position, p_z is the probability of the inferred genotype (g_z) based on flanking markers ($p_z = \text{Pr}(g_z \,|M)$, M flanking markers) and e_{ij} is the residual error. Confidence intervals (95 %) were estimated using bootstraps with resampling (10 000 iterations). Two levels of significance were calculated based on chromosome ($\alpha = 0.01$) or genome-wide thresholds ($\alpha = 0.05$) by performing 1000 permutations, with the detected QTL referred to as suggestive or significant, respectively [35–37].

GWAS approach

A GWAS was performed using R/rrBLUP [38] in order to test for SNPs associated with phenotypic sex. The model used was based on Yu et al. [39] and had the following format:

$$y = X\alpha + Zu + e,$$

where y is the vector of the phenotypes, α is the vector of marker effects, u is the vector of animal random effects $\sim N(0, G\sigma_g^2)$ and e is the vector of residuals. The matrix G represents the relationship matrix calculated from SNPs and σ_g^2 the additive genetic variance estimated using REML. X and Z are incidence matrices relating y to α and u, respectively. The Bonferroni correction was used to correct for multiple testing (Type I error rate = 0.05).

Prediction of phenotypic sex or length based on estimated breeding values

A preliminary study was conducted to test whether breeding values that were estimated from the additive effects of SNPs, could be used as a predictive measure of phenotypic sex or length, respectively. SNPs with more than 5 % missing data were removed. Missing values of the remaining 4881 SNPs were imputed with R/synbreed [40]. Additive SNP effects were estimated using RRBLUP, BayesA, BayesB [41], BayesC [42] or Bayesian Lasso [43] using R/BGLR [44]. Pedigree-based BLUP [45] was applied using the same software.

The general form of the fitted models was the following:

$$y = \eta + \varepsilon,$$

where y is the vector of phenotypic records, η the linear predictor and ε the vector of residuals. In the scenario for the prediction of phenotypic sex, the probit link function was used to connect the underlying latent variable with the linear predictor.

The linear predictor η in the case of RRBLUP, BayesA, BayesB, BayesC and Bayesian Lasso had the following general form:

$$\eta = 1\mu + X_1\beta_1 + X_2\beta_2,$$

where μ is the intercept, X_1 and X_2 are design matrices relating the phenotypes to the included fixed effects and markers, respectively, β_1 is the vector of regression coefficients for the included fixed effects (length or sex and tank, respectively for each study), and β_2 is the vector of marker effects with corresponding priors depending on the model used.

The linear predictor η in the case of pedigree-based BLUP had the following form:

$$\eta = 1\mu + X_1\beta_1 + Zu,$$

where u is the animal random effect that follows $\sim N(0, V\sigma_g^2)$ where V is the pedigree-based relationship matrix. Z is the incidence matrix relating y to u.

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68

Page 5 of 10

The parameters of the above models were estimated by MCMC sampling (100 000 iterations; burn-in: 10 000; thin: 100). Convergence of the resulting posterior distributions was assessed both visually and analytically by the Geweke diagnostic using R/boa v1.1.7 [46].

The dataset was randomly split into a training set (150 individuals) and a test set (25 individuals). This was repeated 100 times to record prediction accuracies for each model tested. The prediction of phenotypic sex was tested by the following naïve approach: animals in the test set that were predicted by the model to have a probability of being female greater than 0.5 were considered as females, while the rest were considered as males. The number of correctly assigned individuals for each model tested was recorded. T-tests (right-tailed) were performed to evaluate whether the mean number of correctly assigned individuals was significantly larger ($\alpha = 0.05$) from that expected by chance alone.

Results

RAD reads

In total, 1 156 659 542 raw reads (100 bases long) were produced (578 329 771 paired-end reads, EBI-SRA study ERP004018). After removing low-quality sequences (i.e., with a quality score less than 30), ambiguous barcodes, and orphaned paired-end reads, 76.7 % of the raw reads were retained (886 927 866 reads). Then, assembly and grouping of the sequences into the RAD loci for each individual were performed with the Stacks package [26] and 56 696 unique RAD-tags were retrieved

(see Additional file 1: Table S1). In order to maximise the number of informative SNPs and minimise the amount of missing or erroneous data, we used the RAD-tags that were retrieved in at least 75 % of the samples in each family, and that carried only one or two SNPs.

Parentage assignment- general statistics

All 175 progeny (88 females and 87 males) were assigned to a unique parental pair using Vitassign [29], by allowing a maximum of five mismatches in a test panel of 200 SNPs. Resulting parentage assignments were confirmed with R/hsphase [30], allowing for a maximum genotypic error of 3 %. Sire 1 had the smallest number of progeny, while Sire 4 had the largest number, for both dams. A correlation of 0.96 was found between weight and length, while the correlation between phenotypic sex and weight or length was equal to 0.23. Significant deviations from an equal sex ratio within full-sib families were observed only for the parental pair Dam 2 and Sire 3 (Table 1). The sire effect on sex-ratio was highly significant ($P < 0.001$) but the dam effect was not ($P > 0.50$). The sire-dam interaction was not significant ($P > 0.50$), which indicated that the genetic variation in sex-ratio was additive.

Linkage map

The constructed linkage map consisted of 6706 SNPs that were grouped in 24 linkage groups, in accordance with the number of chromosomes in the *D. labrax* karyotype, with a total length of 4816 cM (Fig. 1;

Fig. 1 Sea bass linkage map. Heatmap on the right side provides scale of colour coding for the size of SNP clusters

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68
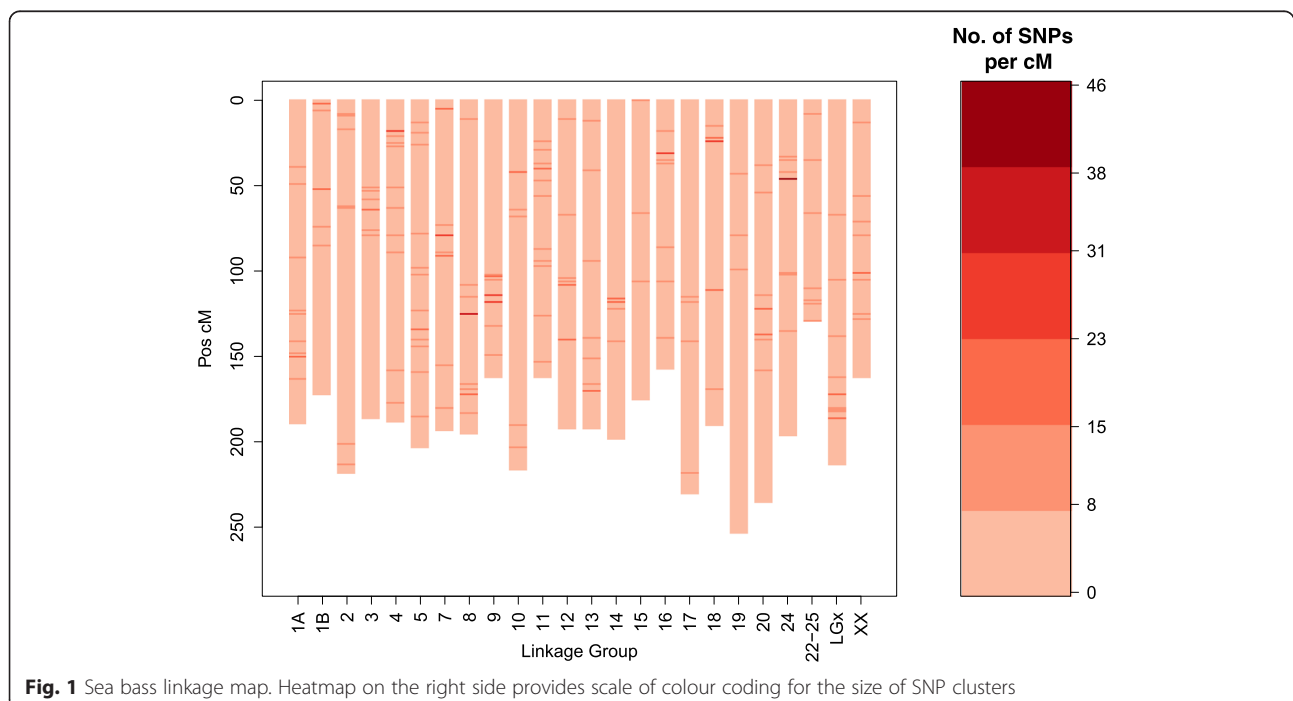
Page 6 of 10

Table 2). Each linkage group corresponded to a different chromosome, as confirmed by comparison with the sea bass genome sequence [47]. In addition, the linkage map included 852 SNPs that were located in unanchored contigs of the sea bass genome (see Additional file 2: Table S2).

Estimation of sex heritability, QTL mapping and GWAS

The estimated heritability for phenotypic sex was equal to 0.47 (95 % density interval: 0.15 to 0.79). In the F_2 sib regression analysis, sex-determining QTL that were significant at the genome-wide level were detected on linkage groups 6, 11 and 18 to 21 (Fig. 2; Table 3), with F values of 19.19, 20.28 and 20.17, respectively. The significance threshold at the genome-wide level had a value of F = 17.8 (10 000 permutations, $\alpha = 0.05$). The QTL confidence intervals (95 %) spanned regions between 3 and 147 cM in linkage group 6, between 12 and 143 cM in linkage group 11, and between 30 and 237 cM in linkage groups 18 to 21. In addition, one suggested sex-related

QTL (significant at the chromosome level; $\alpha = 0.01$) was detected on linkage group 12 (Table 3).

The maternal half-sib regression model detected suggestive sex-determining QTL in four linkage groups. Two were detected in the first maternal half-sib panel (Dam 1 - linkage groups: 19 and 24; Table 4) and four in the second maternal half-sib panel (Dam 2 - linkage groups: 12, 14, 19 and 24; Table 4). The genome-wide significant threshold F was equal to 19.58 (10 000 permutations, $\alpha = 0.05$). The paternal half-sib regression model did not detect significant QTL at either the genome-wide level ($\alpha = 0.05$) or the chromosome level ($\alpha = 0.01$).

The GWAS was not able to identify individual SNPs that were significantly associated with phenotypic sex (Bonferroni threshold $P < 10^{-6}$). The SNPs with the lowest p-values ($10^{-4} < P < 10^{-3}$) were located in linkage groups 6 (lowest p-value: $10^{-3.5}$), 12, and 16 (Fig. 3).

Prediction of phenotypic sex and length based on estimated breeding values

The convergence diagnostics of Geweke and MCMC related graphs did not provide evidence of non-convergence of the estimated parameters (posterior distributions). For prediction of phenotypic sex, all models led to a significantly larger number of correctly assigned animals than expected by chance (Table 5). The application of SNP-based models resulted in a slightly better prediction (mean correct assignment 66 to 67 %) than that achieved using pedigree-based BLUP (mean correct assignment 64 %). For prediction of body length, (Table 6), prediction accuracy was lowest with the pedigree-based BLUP model (0.32) and ranged from 0.41 to 0.44 with the other models.
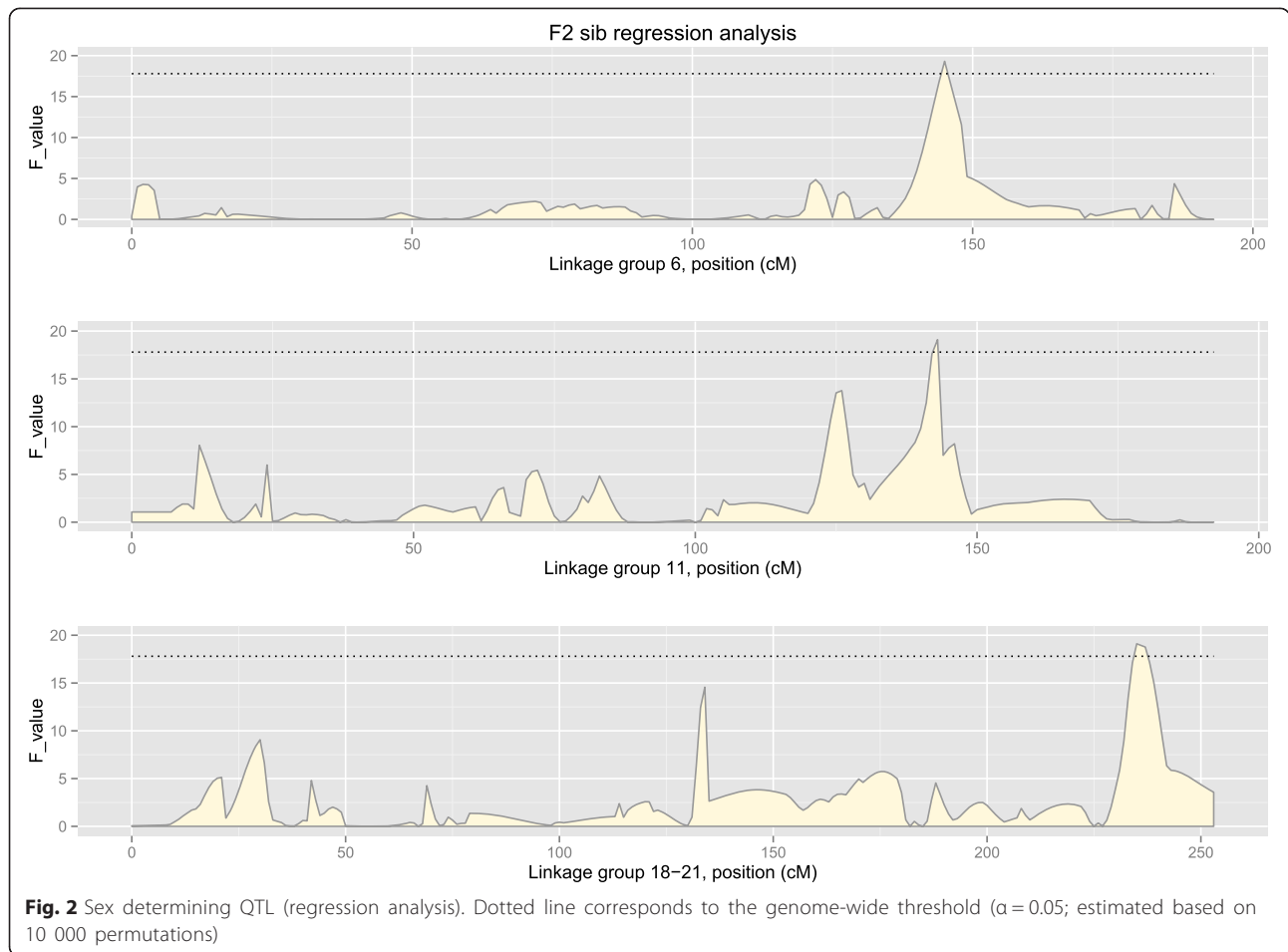
Discussion

D. labrax is one of the most important species in Mediterranean aquaculture. The observed sexual dimorphism in growth and the need to control sex-ratio in selective breeding programmes are practical concerns that would benefit from a better understanding of the sex determination system in sea bass. Current evidence, based on the variance of sex-ratio between families, suggests the existence of a polygenic sex determination system [17]. Use of genetic markers offers the potential to directly assess genetic variance and its distribution between putative QTL and a polygenic background.

For *D. labrax*, relatively rich genomic resources are available and a reference genome has recently been publicly released [47]. However, linkage maps that are available for *D. labrax* are mainly based on microsatellites and AFLP (amplification fragment length polymorphisms) and the most recent map consists of 190

Table 2 Details of the sea bass linkage map

Corresponding chromosome (seabass_v1.0)	Number of markers	Unique positions	Length (cM)
LG1A	302	125	191.61
LG1B	238	84	163.33
LG2	216	96	204.47
LG3	219	89	174.05
LG4	315	119	191.19
LG5	363	147	206.11
LG6	299	133	243.5
LG7	330	126	245.49
LG8	301	90	156.75
LG9	226	108	259.7
LG10	305	108	218.23
LG11	255	98	192.49
LG12	292	105	181.85
LG13	280	140	188.76
LG14	217	102	216.2
LG15	298	110	173.17
LG16	259	137	227.55
LG17	318	104	199.54
LG18-21	286	147	210.52
LG19	272	111	202.01
LG20	333	113	229.37
LG24	203	76	164.14
LG22-25	321	117	194.33
LGx	258	91	182.57
	6706	2676	4816.93

Palaiokostas et al. Genetics Selection Evolution (2015) 47:68

Page 7 of 10

Fig. 2 Sex determining QTL (regression analysis). Dotted line corresponds to the genome-wide threshold (α = 0.05; estimated based on 10 000 permutations)

microsatellites, 176 AFLP and two SNPs [48, 49]. In this study, we present the first high-density linkage map of *D. labrax* based on 6706 SNPs (2676 unique positions). The number of linkage groups corresponded to the number of chromosomes in the *D. labrax* karyoptype. The accurate grouping of markers was further confirmed by comparison to the reference genome (see Additional file 2: Table S2). The genetic map presented here spans 4816 cM, while the map of Chistiakov et al. [49] has a total length of 1373 cM. We hypothesize that this large increase in size is mainly due to the larger number of markers used. This had already been observed between
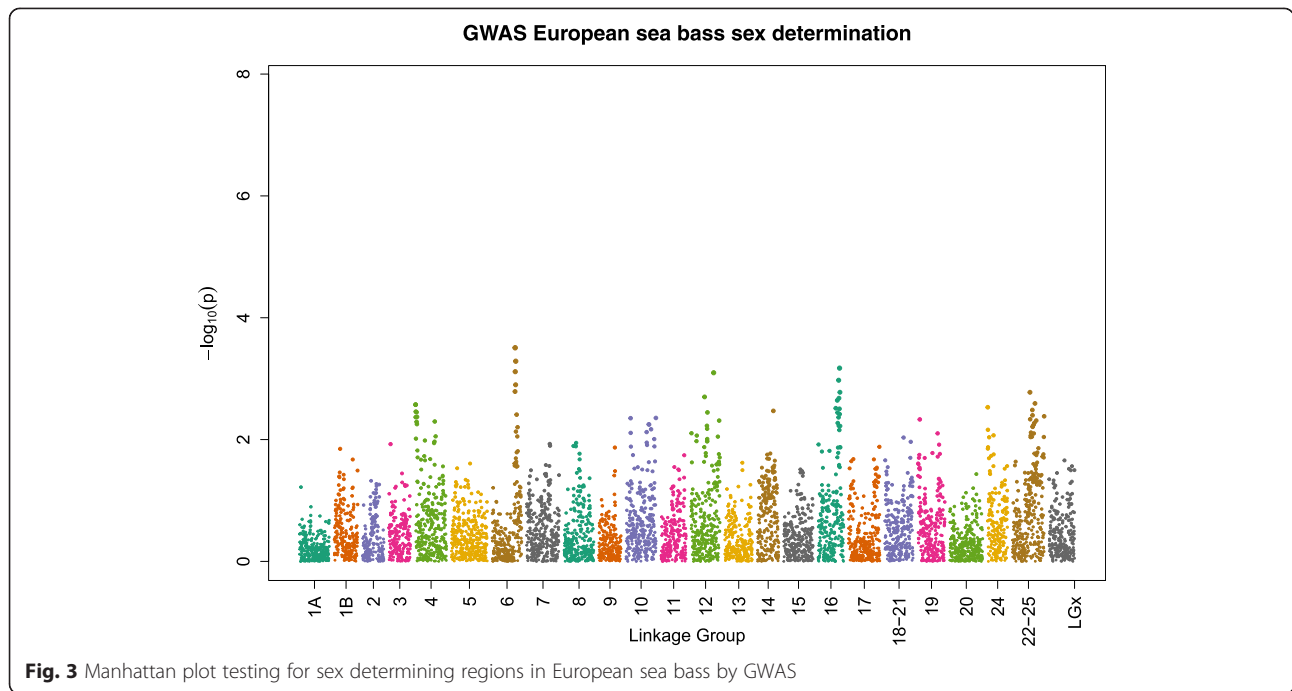
the map of Chistiakov et al. [49] (368 markers) and the first *D. labrax* linkage map, which was based on 174 microsatellites and spanned 814 cM [48]. The 856 SNPs on our map that are located in unanchored contigs of the current sea bass reference genome (*seabass_V1.0*) should help to improve future versions of the assembly.

A moderate correlation of phenotypic sex with weight and length (r = 0.23) was found. In contrast,

Table 3 Mapped sex-determining QTL using F_2 sib regression analysis

LG	Position (cM)	LOD	F
6	146	21.28[**]	4.3
11	143	19.19[**]	3.9
12	192	16.19[*]	3.3
18-21	236	20.17[**]	4.1

*Chromosome wide significant (α = 0.01); **Genome-wide significant (α = 0.05)

Table 4 Mapped sex-determining QTL using half-sib regression analysis

Half-sib family	LG	Position (cM)	F
Dam_1	19	138	16.65[*]
	24	22	16.01[*]
Dam_2	12	186	18.42[*]
	14	121	18.69[*]
	19	109	17.67[*]
	24	90	16.03[*]

*Chromosome wide significant (P < 0.01)

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68

Page 8 of 10

Fig. 3 Manhattan plot testing for sex determining regions in European sea bass by GWAS

Vandeputte et al. [17] reported a relatively strong genetic correlation between weight and sex ($r_A = 0.5$). Successive grading has been shown to produce dominantly female (larger fish) and dominantly male (smaller fish) populations [16, 50], while Diaz et al. [15] showed that a clear relationship exists between growth rate at stages prior to sex differentiation (3 to 4 cm) and sex ratio in sea bass. We estimated a heritability of 0.47 for sex, which is within the range of values reported for sea bass elsewhere, i.e., 0.12 in [51], and 0.62 in [17]. However, different methods were used to estimate heritabilities in these studies, which prevents making meaningful conclusions. It should also be noted that, in our study, the number of parents was small and thus, the estimated heritability cannot be generalized outside the studied pedigree. For comparison, in the study of

Vandeputte et al. [17], heritability was estimated using data on 5893 animals from 253 full-sib families.

The F_2 sib regression analysis detected three genome-wide significant sex-determining QTL on linkage groups 11, 16 and 18. Among these, the best candidate QTL is in linkage group 16, since this region had the highest statistical significance both in the QTL mapping approach and the GWAS. However, since the supporting F-values only just exceeded the estimated genome-wide threshold and since QTL mapping studies tend to overestimate the QTL effect due to the Beavis effect [52], these results should be considered with caution. In addition, the fact that no individual SNP was significant after Bonferroni correction ($\alpha = 0.05$) requires these QTL to be confirmed in a larger dataset.

Evidence that supports the hypothesis that a polygenic sex determination system exists in *D. labrax* was obtained by testing the efficiency of predicting phenotypic

Table 5 Proportions of offspring with sex correctly assigned in the validation sets (25 animals; 100 replicates) and testing of prediction deviations from those expected by chance using t-tests

Model	Assigned correctly (%)	P-value
pBLUP	64	$<10^{-16}$
RRBLUP	66	$<10^{-16}$
BayesA	67	$<10^{-16}$
BayesB	67	$<10^{-16}$
BayesC	67	$<10^{-16}$
BayesLasso	67	$<10^{-16}$

Table 6 Mean accuracies of predicted length for animals in the validation sets (25 animals; 100 replicates)

Model	Mean accuracy (s.e)
pBLUP	0.32 (0.02)
RRBLUP	0.41 (0.03)
BayesA	0.43 (0.03)
BayesB	0.44 (0.03)
BayesC	0.41 (0.03)
BayesLasso	0.41 (0.03)

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68

Page 9 of 10

sex based on estimated breeding values. The genomic-based model gave only slightly better predictions than the pedigree-based BLUP. This is most probably due to a combination of factors; primarily the analysis of a relatively small dataset and the response variable being a binary trait (requiring the fitting of generalized linear mixed models). The latter raises additional issues, e.g., the residual variance has to be fixed in order to achieve identifiable estimated parameters. In comparison, when the same models were tested for the prediction of a continuous trait (total body length), the genomic models were clearly more efficient than pedigree-based BLUP. Another issue is that some QTL may be homozygous in some of the parents and thus remain undetected. The fact that paternal half-sib regression does not detect any QTL while most of the variation in sex-ratio is between sire half-sib families would advocate this. In this case, homozygous QTL may contribute to variation between sires, but remain undetected since they do not segregate and thus do not contribute to genomic prediction. Finally, the fact that all genomic models gave similar prediction accuracies, a phenomenon often observed in the study of polygenic traits [41, 53–55], would indirectly support the fact that, in *D. labrax*, sex is a polygenic trait, as hypothesised previously [17].

Although firm conclusions cannot be drawn from this study, the fact that all tested models gave significantly better predictions than that expected by chance alone indicates that further improvement could be possible by increasing genotyping efforts.

Conclusions

This study presents the first high-density linkage map for the European sea bass. Based on the large number of SNPs (856) that are located in unanchored contigs of the recently published reference genome and thus, that could be positioned, this map will help to improve the existing genome assembly. Overall, the study supports the polygenic hypothesis of sex determination in sea bass of Vandeputte et al. [17]. The families used in this study originated from the West Mediterranean region where significant differentiation exists between populations of *D. labrax*. It is likely that if the unstable nature of the polygenic determinism of sex in *D. labrax* evolves to QTL with larger effects, as suggested by theory [56, 57], this evolution could differ between populations. Searching for population-specific QTL using the same technique as that in this study is, therefore, the next logical step to help unravel the complex genetic sex determination system of this species. Finally, the preliminary genomic prediction results indicate that selection for increased female sex ratios and sizes should be possible.

Additional files

Additional file 1: Table S1. Origin of samples and barcodes. This file provides details for each sample used i.e., sample ID, family, gender, barcode used, number of raw reads paired-ended and number of RAD-tags. (CSV 23 kb)

Additional file 2: Table S2. Detailed positions of individual SNPs in the constructed genetic map. (CSV 281 kb)

Authors' contributions
CP, MB, MV, BC, JBT, KG, BJM and DJP conceived and designed the experiments. CP, MB, MV and JBT performed the experiments. CP, MB and MV analyzed the data. CP, MV, JBT and KG contributed reagents/materials/analysis tools. CP, MV, JBT and DJP wrote the paper. BC, MV, MB, JBT, KG, DJP and BJM contributed to editing. All authors read and approved the final manuscript.

Author details
[1]Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK. [2]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland, UK. [3]Edinburgh Genomics, Ashworth Laboratories, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK. [4]Ifremer, Chemin de Maguelone, F-34250 Palavas-les-Flots, France. [5]INRA, UMR1313 Génétique animale et Biologie intégrative, F-78350 Jouy-en-Josas, France.

References
1. Devlin RH, Nagahama Y. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture. 2002;208:191–364.
2. Penman DJ, Piferrer F. Fish gonadogenesis. Part I: Genetic and environmental mechanisms of sex determination. Rev Fish Sci. 2008;16:16–34.
3. Volff JN, Nanda I, Schmid M, Schartl M. Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev. 2007;1:85–99.
4. Piferrer F, Guiguen Y. Fish gonadogenesis. Part II: Molecular biology and genomics of sex differentiation. Rev Fish Sci. 2008;16:35–55.
5. Federation of European Aquaculture Producers. Annual Report. 2015. http://www.feap.info/default.asp?SHORTCUT=582. Accessed: 08/08/2015
6. Vandeputte M, Quillet E, Chatain B. Are sex ratios in wild European sea bass (*Dicentrarchus labrax*) populations biased? Aquat Living Resour. 2012;25:77–81.
7. Zanuy S, Carrillo M, Felip A, Rodríguez L, Blázquez M, Ramos J, et al. Genetic, hormonal and environmental approaches for the control of reproduction in the European sea bass (*Dicentrarchus labrax* L.). Aquaculture. 2001;202:187–203.
8. Gorshkov S, Gorshkova G, Meiri I, Gordin H. Culture performance of different strains and crosses of the European sea bass (*Dicentrarchus labrax*) reared under controlled conditions at Eilat. Israel J Appl Ichthyol. 2004;20:194–203.
9. Conover DO. Adaptive significance of temperature-dependent sex determination in a fish. Am Nat. 1984;123:297–313.
10. Piferrer F, Blázquez M, Navarro L, González A. Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). Gen Comp Endocrinol. 2005;142:102–10.

Palaiokostas *et al. Genetics Selection Evolution* (2015) 47:68

Page 10 of 10

11. Navarro-Martín L, Blázquez M, Viñas J, Joly S, Piferrer F. Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*). Aquaculture. 2009;296:347–58.

12. Peruzzi S, Chatain B, Saillant E, Haffray P, Menu B, Falguiere JC. Production of meiotic gynogenetic and triploidsea bass, *Dicentrarchus labrax* L. 1. Performances, maturation and carcass quality. Aquaculture. 2004;230:41–64.

13. Blazquez M, Carrillo M, Zanuy S, Piferrer F. Sex ratios in offspring of sex-reversed sea bass and the relationship between growth and phenotypic sex differentiation. J Fish Biol. 1999;55:916–30.

14. Francescon A, Barbaro A, Bertotto D, Libertini A, Cepollaro F, Richard J, et al. Assessment of homozygosity and fertility in meiotic gynogens of the European sea bass (*Dicentrarchus labrax* L.). Aquaculture. 2005;243:93–102.

15. Díaz N, Ribas L, Piferrer F. The relationship between growth and sex differentiation in the European sea bass (*Dicentrarchus labrax*). Aquaculture. 2013;408–9:191–202.

16. Saillant E, Fostier A, Haffray P, Menu B, Thimonier J, Chatain B. Temperature effects and genotype-temperature interactions on sex determination in the European sea bass (*Dicentrarchus labrax* L.). J Exp Zool. 2002;292:494–505.

17. Vandeputte M, Dupont-Nivet M, Chavanne H, Chatain B. A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. Genetics. 2007;176:1049–57.

18. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 2008;3:e3376.

19. Anderson JL, Rodríguez Marí A, Braasch I, Amores A, Hohenlohe P, Batzel P, et al. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. PLoS ONE. 2012;7:e40701.

20. Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB, et al. Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. BMC Genomics. 2013;14:566.

21. Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, McAndrew BJ. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. PLoS ONE. 2013;8:e68389.

22. Palaiokostas C, Bekaert M, Khan MG, Taggart JB, Gharbi K, McAndrew BJ, et al. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). BMC Genomics. 2015;16:171.

23. Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol. 2011;772:157–78.

24. Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A, et al. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. BMC Genomics. 2012;13:244.

25. Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, et al. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PLoS ONE. 2011;6:e19315.

26. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. *Stacks*: Building and genotyping Loci de novo from short-read sequences. G3 (Bethesda). 2011;1:171–82.

27. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 2010;6:e1000862.

28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

29. Vandeputte M, Mauger S, Dupont-Nivet M. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Mol Ecol Notes. 2006;6:265–67.

30. Ferdosi MH, Kinghorn BP, van der Werf JHJ, Lee SH, Gondro C. hsphase: an R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. BMC Bioinformatics. 2014;15:172.

31. Orelien JG. Model fitting in PROC GENMOD. In Proceedings of the 26th Annual SAS Users Group International Conference: 22–25 April 2001; Long Beach.2001:Paper 264–26. http://www2.sas.com/proceedings/sugi26/p264-26.pdf.

32. Rastas P, Paulin L, Hanski I, Lehtonen R, Auvinen P. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. Bioinformatics. 2013;29:3128–34.

33. Hadfield J. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J Stat Softw. 2010;33:1–22.

34. Seaton G, Hernandez J, Grunchec J, White I, Allen J, De Koning D et al. GridQTL: A Grid portal for QTL mapping of compute intensive datasets. In Proceedings of the 8th World Congress on Genetics Applied to Livestock Production: 13–18 August 2006; Belo Horizonte; 2006. http://www.gridqtl.org.uk/publications/wcgalp27_633-916.pdf.

35. Knott SA, Elsen JM, Haley CS. Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theor Appl Genet. 1996;93:71–80.

36. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138:963–71.

37. Baranski M, Moen T, Våge DI. Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). Genet Sel Evol. 2010;42:17.

38. Endelman JB. Ridge regression and other kernels for genomic selection with R Package rrBLUP. Plant Genome J. 2011;4:250–5.

39. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203–8.

40. Wimmer V, Albrecht T, Auinger HJ, Schön CC. Synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics. 2012;28:2086–7.

41. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.

42. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.

43. Park T, Casella G. The Bayesian Lasso. J Am Stat Assoc. 2008;103:681–6.

44. Pérez P, de Los CG. Genome-wide regression and prediction with the BGLR statistical package. Genetics. 2014;198:483–95.

45. Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975;31:423–47.

46. Smith BJ. boa: An R Package for MCMC output convergence. Assessment and Posterior Inference. J Stat Softw. 2007;21:1–37.

47. Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 2014;5:5770.

48. Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigenopoulos CS, Kotoulas G, et al. A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. Genetics. 2005;170:1821–6.

49. Chistiakov DA, Tsigenopoulos CS, Lagnel J, Guo YM, Hellemans B, Haley CS, et al. A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. Anim Genet. 2008;39:623–34.

50. Papadaki M, Piferrer F, Zanuy S, Maingot E, Divanach P, Mylonas CC. Growth, sex differentiation and gonad and plasma levels of sex steroids in male- and female-dominant populations of *Dicentrarchus labrax* obtained through repeated size grading. J Fish Biol. 2005;66:938–56.

51. Chatziplis D, Batargias C, Tsigenopoulos CS, Magoulas A, Kollias S, Kotoulas G, et al. Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): The BASSMAP pilot study. Aquaculture. 2007;272:S172–82.

52. Xu S. Theoretical basis of the Beavis effect. Genetics. 2003;165:2259–68.

53. Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics. 2010;185:623–31.

54. Kizilkaya K, Fernando RL, Garrick DJ. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J Anim Sci. 2010;88:544–51.

55. Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S. Improved Lasso for genomic selection. Genet Res (Camb). 2011;93:77–87.

56. Bulmer MG, Bull JJ. Models of polygenic sex determination and sex ratio control. Evolution. 1982;36:13–26.

57. Van Dooren TJM, Leimar O. The evolution of environmental and genetic sex determination in fluctuating environments. Evolution. 2003;57:2667–77.