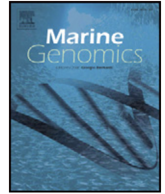
Marine Genomics 28 (2016) 63–70

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen



# Whole genome semiconductor based sequencing of farmed European sea bass (*Dicentrarchus labrax*) Mediterranean genetic stocks using a DNA pooling approach

Francesca Bertolini [a], Claudia Geraci [a], Giuseppina Schiavo [a], Maria Teresa Sardina [b], Vincenzo Chiofalo [c], Luca Fontanesi [a,*]

[a] *Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy*
[b] *Department of Agricultural and Forestry Sciences, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy*
[c] *Department of Veterinary Medicine, Division of Animal Production, University of Messina, Polo Universitario dell'Annunziata, 98168 Messina, Italy*

ARTICLE INFO

*Article history:*
Received 22 February 2016
Received in revised form 14 March 2016
Accepted 14 March 2016
Available online 25 March 2016

*Keywords:*
Cultivated sea bass
Next generation sequencing
Ion Proton
SNP

ABSTRACT

European sea bass (*Dicentrarchus labrax*) is an important marine species for commercial and sport fisheries and aquaculture production. Recently, the European sea bass genome has been sequenced and assembled. This resource can open new opportunities to evaluate and monitor variability and identify variants that could contribute to the adaptation to farming conditions. In this work, two DNA pools constructed from cultivated European sea bass were sequenced using a next generation semiconductor sequencing approach based on Ion Proton sequencer. Using the first draft version of the *D. labrax* genome as reference, sequenced reads obtained a total of about 1.6 million of single nucleotide polymorphisms (SNPs), spread all over the chromosomes. Transition/transversion (Ti/Tv) was equal to 1.28, comparable to what was already reported in Salmon species. A pilot homozygosity analysis across the *D. labrax* genome using DNA pool sequence datasets indicated that this approach can identify chromosome regions with putative signatures of selection, including genes involved in ion transport and chloride channel functions, amino acid metabolism and circadian clock and related neurological systems. This is the first study that reported genome wide polymorphisms in a fish species obtained with the Ion Proton sequencer. Moreover, this study provided a methodological approach for selective sweep analysis in this species.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

*Dicentrarchus labrax* L, the European sea bass (or European seabass), is a teleost fish of the family *Moronidae* mainly distributed along the coasts of the North-Eastern Atlantic Ocean and of the Mediterranean and Black Seas. The European sea bass is a eurythermal and euryhaline species; for these reasons it is adapted to a broad range of environmental conditions and regions, from coastal areas to coastal lagoons or estuary and rocky zones and occasionally rivers (Pickett and Pawson, 1994; Volckaert et al., 2008). This species displays an unusual dual pattern of activity behavior: diurnal in spring and summer time then switching to a nocturnal behavior during wintertime (Sánchez-Vázquez et al., 1998). *D. labrax* is one of the most important marine species for commercial and sport fisheries and aquaculture production, raising attentions and concerns on one hand for conservation and management of natural genetic resources and on the other hand for adaptation to aquaculture conditions, domestication and selective breeding. The expansion of the aquaculture sea bass industry, mainly in Mediterranean countries like Greece, Turkey, Italy, Spain, Croatia and Egypt, has reached an estimated worldwide production of cultivated European sea bass of about 180,000 t (FAO, 2012). Compared with other economically relevant cultivated fish species, such as salmon and trout, genetic information available for *D. labrax* is still in its infancy. Two generations of genetic maps, developed using microsatellites and AFLPs, were developed (Chistiakov et al., 2005, 2008). A radiation hybrid map was obtained by Guyon et al. (2010) and BAC sequencing data and preliminary single nucleotide polymorphism discovery were reported (Kuhl et al., 2010, 2011a, 2011b). Only recently, the publication of the first version of the assembled *D. labrax* whole genome has opened new opportunities to evaluate and monitor genetic stock variability (Tine et al., 2014).

Among the currently available next generation sequencing (NGS) platforms, Ion Torrent has been proposed as a new potentially cost effective NGS technology (Rothberg et al., 2011; Merriman et al., 2012). Ion Torrent sequencing is based on the detection in a specifically constructed chip of $H^+$ that are released during the DNA elongation steps (Rothberg et al., 2011). Recently, the Ion Torrent technology has been implemented in a new machine (Ion Proton) that can sequence in a single run several Gbp (up to 10 Gbp with the Ion Proton chip II), increasing the throughput if compared with the first Ion Torrent

* Corresponding author.
*E-mail address:* luca.fontanesi@unibo.it (L. Fontanesi).

http://dx.doi.org/10.1016/j.margen.2016.03.007
1874-7787/© 2016 Elsevier B.V. All rights reserved.

Personal Genome Machine (PGM) (Boland et al., 2013). We already have used the Ion Torrent PGM and Ion Proton machine for SNP discovery and analysis in complex mammalian genomes (i.e. pig, rabbit and donkey), demonstrating that semiconductor based sequencing can identify with high reliability polymorphisms useful for subsequent applications (Bertolini et al., 2014, 2015; Bovo et al., 2015).

In this work, using an Ion Proton sequencer, we sequenced the genome of farmed European sea bass stocks derived by two Mediterranean hatcheries, applying a DNA pooling approach to identify single nucleotide polymorphisms (SNPs) that were used to design a pilot strategy to detect putative signatures of selection in this species.

## 2. Materials and methods

### 2.1. Sample collection and DNA extraction

Six sea bass were randomly collected from each of two different Mediterranean hatcheries located in two Italian islands: hatchery 1 (H01) was located in Lampedusa and hatchery 2 (H02) was located in Sicily, province of Siracusa. The two hatcheries do not exchange genetic material and have developed their reproductive stocks based on their breeding programs running for about 25 years and based mainly on mass selection for production and efficiency traits. During this period, limited restocking from wild sea bass was carried out a few times, as a common practice in European sea bass farms, to include additional genetic variability in the cultivated lines but precise records on these events were not available. DNA was extracted from liver using the NucliSENS® Nucleic Acid Extraction kit (Biomerieux, Marcy l'Etoile, France), following the manufacturer's protocol. Integrity, quality and quantity of extracted DNA was assessed by i) visual inspection of 5 μg of DNA electrophoresed in 0.8% agarose gels in TBE buffer stained with 1× GelRed Nucleic Acid Gel Stain (Biotium Inc., Hayward, CA, USA) and ii) by measurements with a NanoPhotometer P-330 (Implen GmbH). Then, two DNA pools were constructed with equimolar DNA from each of the six fish of the two different hatcheries (H01 and H02).

### 2.2. Ion Proton sequencing

Two Ion Proton libraries (one for each of the two DNA pools: H01 and H02) were prepared following the manufacturer protocols (Thermo Fisher Scientific/Life Technologies). Briefly, for each library, 1.2 μg of pooled DNA was enzymatically sheared, end repaired, adapter ligated and barcoded using the Ion Xpress™ Plus Fragment Library kit (Thermo Fisher Scientific/Life Technologies) and the Ion Xpress™ Barcode Adapters 1–16 kit (Thermo Fisher Scientific/Life Technologies). The resulting sheared and adapter ligated DNA was size selected using the e-gel system (Invitrogen, Carlsbad, CA, USA). To avoid bias due to the enzymatic shearing, two bands corresponding to 200 bp and 250 bp respectively (considering the insert size and the adapters) were collected and merged for the preparation of each library. Then, the two barcoded libraries, containing each the 2 size-select fragments, were i) separately quantified by qPCR using a StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific/Life Technologies) and then ii) merged in equimolar concentration, to reach the final concentration of 11 pM. The resulted mix of barcoded libraries was attached to the Ion Sphere Particles (ISP) and clonally amplified using the Ion PI™ Template OT2 200 kit v2 and Ion PI™ Template OT2 200 kit v3 (Thermo Fisher Scientific/Life Technologies). Then this clonally amplified material was purified and sequenced on the Ion Proton sequencer using two Ion PI Chip v2 (Thermo Fisher Scientific/Life Technologies) for a total of two sequencing runs. In this way, both libraries prepared from the two different DNA pools, were sequenced twice considering the two different runs (Table 1).

### 2.3. Filtering of sequencing data and alignment on the European sea bass reference genome

Ion Proton reads obtained from the two sequencing runs were first automatically processed with the Ion Torrent Suite (TS) v4.1 on the Ion Torrent Server (Thermo Fisher Scientific/Life Technologies). Preliminary filtering steps included: i) exclusion of polyclonal and low quality sequences; ii) the trimming of adapters and low quality 3′-ends from the high quality reads. Then, a Python script was used to retain for subsequent analyses only trimmed reads with length ≥ 30. Alignment to the reference sequence of *D. labrax* (accession number: GCA_000689215.1; Tine et al., 2014) was carried out using the option *mem* of Burrows–Wheeler Aligner (BWA; Li, 2013), with default parameters. The resulted *sam* file was converted into binary files, and duplicated reads were removed with *Samtools* software (Li et al., 2009). The filtered *bam* files were deposited in the EMBL-EBI European Nucleotide Archive (ENA) with the project accession number PRJEB9487.

### 2.4. Identification of single nucleotide polymorphisms

Single nucleotide polymorphism (SNP) data were extracted by analyzing sequencing reads from all runs and libraries (SNP discovery from the whole experiment) or comparing sequencing reads obtained separately from the two DNA pools (SNP identification and comparison between the two hatcheries).

Variants were identified using the *Samtools mpileup* function and *bcftools* (Li et al., 2009). Several criteria were used to filter and retain variants. In particular, insertions/deletions (indels) were not considered in this study due to the low reliability of NGS technologies (in particular Ion Torrent technology) to call this type of variability (i.e. Boland et al., 2013). Only biallelic SNPs of quality ≥20 and SNPs not included in homo-polymeric regions (more than 3 nucleotides of the same type before or after the mutation point) were retained. Minimum depth at SNP positions was set at 4×. Single nucleotide polymorphisms with the lowest SNP quality (from 20 to 30) were validated by visual inspection of all read alignments with the help of Integrative Genomic Viewer (IGV) software (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

Single nucleotide polymorphism density was calculated in 1-Mbp windows across the reference genome. A Python script was used to assign SNPs to coding or non-coding regions using the annotation map data of the European sea bass reference genome obtained by Tine et al. (2014) and available at http://seabass.mpipz.mpg.de/DOWNLOADS/.

### 2.5. Implementation of a selective sweep pilot methodology

Allele frequency estimation of SNPs was obtained from the separated and the merged DNA pool data by counting the number of reads with the two alternative nucleotides obtained from the two runs (excluding

**Table 1**
Ion Proton reads obtained from the different runs and DNA pools (H01 and H02).

| Items/sequencing runs | Run 1 | | Run 2 | |
|---|---|---|---|---|
| | H01 | H02 | H01 | H02 |
| No. of sequenced nucleotides | 914,742,848 | 1,844,257,760 | 2,707,420,490 | 5,219,189,457 |
| No. of sequenced nucleotides (Q20) | 804,959,215 | 1,650,684,153 | 2,397,959,759 | 4,700,424,669 |
| No. of reads | 7,188,270 | 15,781,188 | 21,222,148 | 44,695,250 |
| Mean read length (bp) | 127 | 116 | 127 | 116 |

reads mapping in the Superscaffold). For the merged data, that were used for selective sweep, only SNP positions with sequenced depth ranging from 12 to 50 were considered, to avoid regions with low coverage (considering as lower limit the number of sampled chromosomes) from one hand and from the other hand to filter out repetitive regions that might produce biases in the allele frequency estimation (Rubin et al., 2010). The major and the minor alleles (the most and the less frequent alleles) were indicated as $n_{MAJ}$ and $n_{MIN}$, respectively. *Fst* (Fixation index) analysis between the two hatcheries-derived datasets was carried out considering the same alleles as defined in the merged dataset. Single nucleotide polymorphism positions were used to define selective sweeps in 50 kbp genomic windows. The 50 kbp window was selected after testing windows of variable sizes (from 30 kbp to 100 kbp) and plotting the distribution of SNP counts within these windows and observing that this window size yielded few windows with <10 SNPs (less than 10%), similarly to the procedure described by Rubin et al. (2010) (data not shown). A total of 11,585 windows along the European sea bass genome were analyzed and only 1064 were discarded due to a low number of SNPs (<10).

*Fst* values were calculated only to obtain indication about windows in which allele frequencies were in the same direction in the two hatcheries. *Fst* values comparing the two hatcheries were not reported due to the limited depth of the separated datasets. For each window, *Fst* was calculated according to the formula adapted from Karlsson et al. (2007):

$$F = Fst = \frac{\sum_{k=1}^{K} Nk}{\sum_{k=1}^{K} Dk}$$

where

1, 2, $k \ldots K$ is the number of biallelic markers in the two populations and the variant allele for marker $k$ has frequency $p_1^{[k]}, p_2^{[k]}$ in the two populations and the reference allele $k$ in population $i$ is

$q_i^{[k]} = 1 - p_1^{[k]}$ and

$$N_k = p_1^{[k]}\left(q_2^{[k]} - q_1^{[k]}\right) + p_2^{[k]}\left(q_1^{[k]} - q_2^{[k]}\right)$$

$$D_k = p_1^{[k]}q_2^{[k]} + q_1^{[k]}p_2^{[k]} = N_k + p_1^{[k]}q_1^{[k]} + p_2^{[k]}q_2^{[k]}.$$

To estimate $N_k$ and $D_k =$ we used estimators:

$$\widehat{N_k} = (a_1/n_1 - a_2/n_2)^2 - h_1/n_1 - h_2/n_2$$

$$\widehat{D_k} = \widehat{N_k} + h_1 + h_2$$

$$h_i = [a_i(n_i - a_i)]/[n_i(n_i - 1)]$$

where $i = 1,2$, and $a_i$ and $n_i$ are the allele counts and total number of alleles for population i.

The pooled heterozygosity ($H_p$) was calculated using the following formula described by Rubin et al. (2010, 2012):

$$Hp = 2\sum n_{MAJ}\sum n_{MIN}/\left(\sum n_{MAJ} + \sum n_{MIN}\right)^2$$

where $\Sigma n_{MAJ}$ and $\Sigma n_{MIN}$ are the sums of $n_{MAJ}$ and $n_{MIN}$ for all SNPs in the window, respectively. Individual $H_p$ values were then Z-transformed: $ZH_p = (H_p - \mu H_p) / \sigma H_p$ and the results were used for the homozygotes analysis.

The same transformation was done for the *Fst* values. Only windows that had a negative value of the *ZFst* in the comparison between the two hatcheries (that means that there was no difference between the two hatcheries in terms of allele distribution) were used to evaluate the $ZH_p$ values. In this way it was possible to obtain a combined dataset with reads from H01 and H02 in which the *ZHp* values might represent driving forces acting in the same direction on the two cultivated hatcheries. Based on this pre-evaluation analysis, $ZH_p$ value was calculated for a total of 5888 windows. A threshold of $ZH_p \leq -6$ was applied for putative selective sweeps because windows below this threshold represent the extreme lower end of the distribution (Fig. S1), following also what was reported by Rubin et al. (2010).

Genes included in the windows with $ZH_p \leq -6$ were retrieved from the annotated European sea bass reference genome (http://seabass.mpipz.mpg.de/DOWNLOADS/; Tine et al., 2014) and analyzed with the Enrichr platform (http://amp.pharm.mssm.edu/Enrichr/enrich) for gene functional enrichment analyses (Chen et al., 2013) using the GO Biological Process and the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/; KEGG 2015) and PANTHER (http://www.pantherdb.org/pathway/) pathway analyses.

## 3. Results and discussion

### 3.1. Sequencing and SNP data

Ion Proton sequencing runs produced a total of ~10.7 billion of sequenced nucleotides, of which about 90% had quality score equal or higher than Q20. These data were obtained from a total of ~88.8 million of reads with averaged length of 116–127 bp. Sequenced nucleotides and related metrics separated for the two DNA pools constructed from European sea bass of the two hatcheries (H01 and H02) are reported in Table 1.

About 26% of reads were not considered for further analyses because read length was <30 nucleotides. The remaining 65,586,049 reads were aligned to the European sea bass reference genome sequence (Tine et al., 2014). The current European sea bass reference genome accounts for about 675 Mbp of assembled sequences. Sequence assembly of the reference genome was obtained from Adriatic sea bass through a combination of three different sequencing technologies (Sanger sequencing, pyrosequencing and Illumina sequencing) with the help of different approaches (whole genome sequencing assembly, radiation hybrid map, genetic linkage maps, BAC-end and 454 20-kb mate pairs mapped on scaffold groups to construct chromosomal sized superscaffolds). This genome version was constituted by ~86% of ordered and orientated contigs along 24 chromosomes and by a superscaffold of unordered genomic pieces (UN; Tine et al., 2014). Using this assembly, including the unordered genomic superscaffold, only a total of 3,248,077 Ion Proton reads that we obtained were not aligned, confirming the general good quality of the European sea bass genome version available from Tine et al. (2014). The total mean genome coverage that was produced by our filtered reads was 97.8% (93.6% for H01 and 96.2% for H02; Fig. 1) and the overall mean depth was 11.4× (more details are presented in Table S1). As expected, the UN superscaffold regions largely differ from all other chromosomes in terms of coverage and depth, probably due to assembly problems that prevented to assign these regions in a more precise way.

Detection of polymorphisms from the Ion Proton reads was obtained applying a stringent approach that excluded the main biases derived by the Ion Torrent technology that produces a high frequency of indels and a high error rate in homo-polymeric regions (Quail et al., 2012; Boland et al., 2013). In addition, we considered a nucleotide coverage in the SNP positions ≥4× that, from our previous studies, was demonstrated to reduce the false positives, after the pre-filtering step that already eliminated most of the putative sequencing errors produced by the semiconductor based sequencing technology (Bertolini et al., 2014; Bovo et al., 2015). Using these parameters and filters, combining data from the two hatcheries, we obtained on the whole 1,569,357 SNPs. Among these SNPs, 247,302 where homozygous in both pools for the alternative allele of the nucleotide present in the reference genome. The number of SNPs identified in the different chromosomes and in the UN is shown in Fig. 2 and reported in Table S1. The overall mutation rate, as determined by comparing the reference genome, was 0.002.

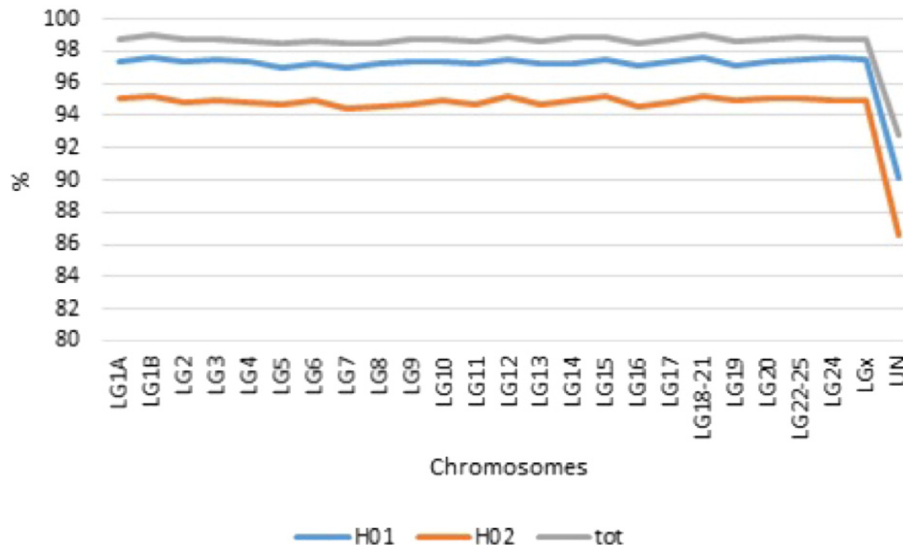



**Fig. 1.** Mean coverage (%) of the sequenced reads against the European sea bass reference genome across the whole chromosomes.

The density of SNPs was higher in LG15 and LG16 as reported in the overview including all chromosomes (Fig. S2) and in the box plots defined with one Mbp sliding window approach (Fig. S3). The higher number of SNPs in the superscaffold seems derived mainly by assembly problems that might have inflated this value. It could be also possible that the difficulties that limited the assembly of the superscaffold are derived by the high number of polymorphisms in these regions, explaining from another perspective the high number of SNPs we identified in the UN.

The current version of the European sea bass genome has 26,719 annotated genes (Tine et al., 2014). Among these genes, 5681 are still located in the superscaffold, while the others are placed along the assembled chromosomes. Among the whole identified filtered SNPs, 169,148 mutations were located in the coding region of a total of 20,914 genes (the 78% of the annotated genes), 147,598 of these SNPs were heterozygous. The distribution of the genes in which we reported variants are included in all chromosomes without any apparent bias (Fig. 3).

The SNP transition/transversion (Ti/Tv) ratio was calculated from the whole dataset (Ti/Tv = 1.28) and from the SNPs called separately from the two DNA pools (H01, Ti/Tv = 1.29; H02, Ti/Tv = 1.32). In mammals this value is reported to be around 2 (that means, a double number of transitions than transversions) in different species, including also SNPs deduced from next generation semiconductor-based sequencing data (Bovo et al., 2015; DePristo et al., 2011; Liu et al., 2012; Molnár et al., 2014). In fish, few studies reported this value for non-mitochondrial genes. Smith et al. (2005) reported that in several *Oncorhynchus* species, the Ti/Tv ratio ranged from 0.95 to 1.49 in resequenced nuclear genes but no whole genome analyses have been reported so far. Taken together from these two studies (Smith et al., 2005; and our survey in sea bass), it seems that in *Teleostei* the Ti/Tv ratio could be lower than in mammals.

### 3.2. Selective sweep analysis in cultivated European sea bass

Considering the obtained SNP data, we designed a pilot strategy (as it was based on a limited number of animals) to take advantage from the genome wide information produced from the two DNA pools. The potential of using pool DNA-seq approaches has been recently discussed by Schlötterer et al. (2014). In our case, the experiment was set up in a cost effective way, applying a few filtering and statistical approaches that, even considering a low number of sampled genomes (but combined from two hatcheries), were able to identify putative selective sweep signals. Rubin et al. (2010, 2012) used a low number of individuals in a DNA pooling strategy to identify selection signatures in chicken and pig genomes. The recently assembled reference genome of *D. labrax* derived from Adriatic sea bass was used in a comparative way to identify putative selective sweeps in the cultivated genetic stocks. The approach does not need to evaluate many different pools as it is based on homozygosity analysis that can be simply obtained comparing a

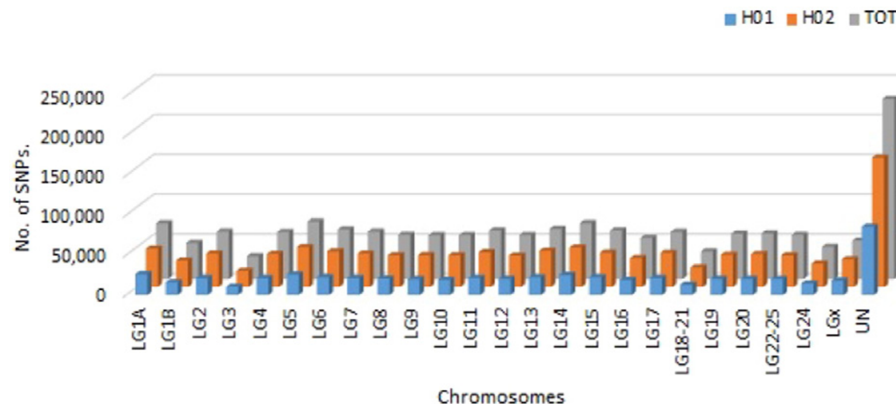



**Fig. 2.** Distribution of single nucleotide polymorphisms in the different chromosomes for the two DNA pools (H01 and H02) and from the merged information.

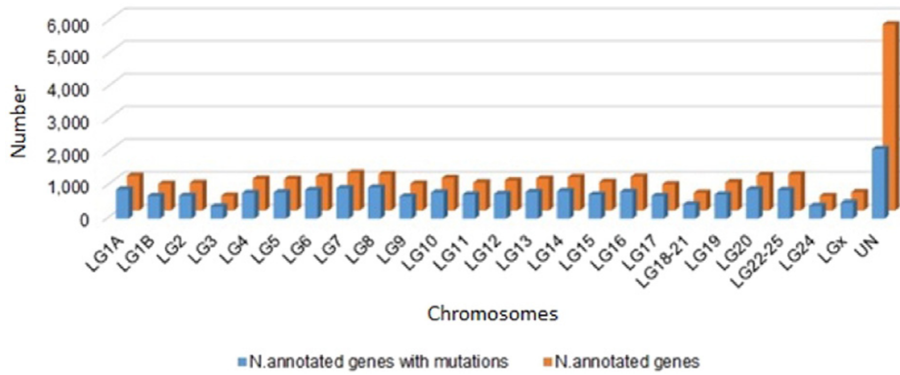



**Fig. 3.** Distribution of the genes in which at least a single nucleotide polymorphism was detected (considering only the coding regions) in the different chromosomes (in blue) compared to the number of annotated genes (in orange) as reported by Tine et al. (2014).

reference genome despite its origin (from a wild or a cultivated European sea bass).

Reads from the two DNA pools were merged to increase coverage and depth and to capture common putative domestication signatures that might have shaped the genome of the two stocks in the same direction (Rubin et al., 2012; Cutter and Payseur, 2013). *Fst* analysis was preliminarily used to filter away signals that might have been derived by unequal depth and sampling and sequencing biases. Then, to identify selective sweeps we used the approach described by Rubin et al. (2010) and calculated *ZHp* in 50 kb genome windows (excluding the superscaffold). Fig. 4 reports a genome wide *ZHp* plot. A total of 17 regions in 11 different chromosomes reached the threshold of −6. Two significant windows were identified in several chromosomes: LG5, LG6, LG8, LG10, LG11 and LG14 (Table 2). At least a few Mbp separated the two significant windows located in the same chromosomes, except for the two signals identified on LG14 that were separated by 150 kb only, suggesting that they could be eventually merged to constitute a larger selective sweep region. The three most significant windows were on LG14 at position 15.60–15.65 Mbp ($ZHp = -8.47$), on LG6 at position 9.25–9.30 Mbp ($ZHp = -7.62$) and on LG5 at position 29.65–29.70 Mbp ($ZHp = -7.36$).

According to the preliminary annotation of the European sea bass draft genome, a total of 31 different genes were located within the 17 significant windows (Table 2). Gene Ontology (GO) analysis for Biological Processes indicated an enrichment (significant at nominal value, even if not significant after correction for multiple testing) for a few GO terms (anion transmembrane transporter activity, GO:0008509; S-adenosylmethionine-dependent methyltransferase activity, GO:0008757; voltage-gated chloride channel activity, GO:0005247; and chloride transmembrane transporter activity, GO:0015108, among the most relevant) that are related to ion transports and nutritional-related metabolism (Table 3).

Gene Ontology related to ion transport and chloride channel activities was mainly due to two genes in the significant windows (chloride channel, voltage-sensitive 2, *CLCN2*; and solute carrier family 12, member 9, *SLC12A9*) that might play an important role in the osmotic adaptation of the euryhaline *D. labrax* to cultivated conditions, mainly in the juvenile stages (Lam et al., 2014; Nebel et al., 2005; Varsamos et al., 2001, 2002). *CLCN2* encodes for a transmembrane protein that regulates chloride ion homeostasis in various cells controlling their volumes that in turn, may affect several essential biological functions in which osmoregulation is relevant (Jentsch, 2015). SLC12A9 is included

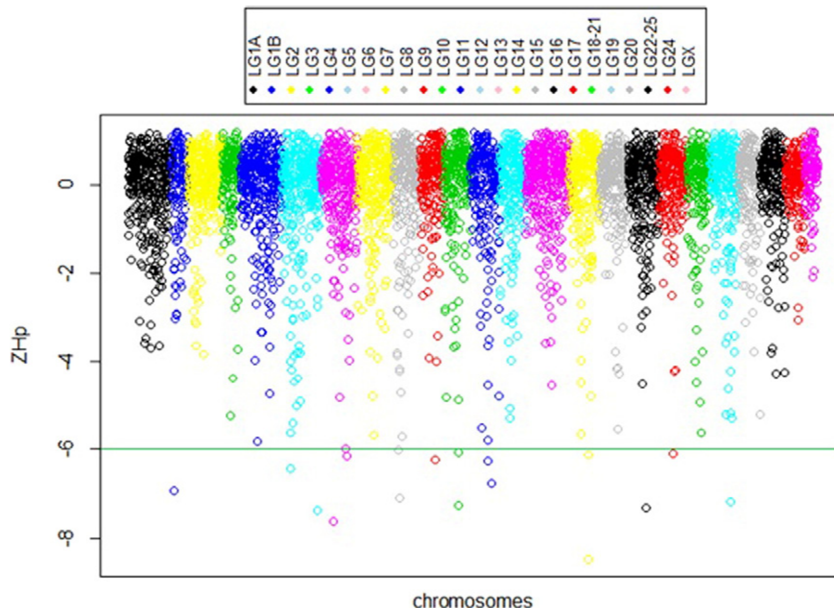



**Fig. 4.** Manhattan plot of the *ZHp* values calculated in 50 kb genomic windows divided by chromosomes (identified with different colors). The threshold (green line) is set at −6.0.

**Table 2**
List of 50 kb genomic windows with *ZHp* < −6.0 and annotated genes included in corresponding region.

| Chr.[a] | Genomic window[b] | No. of SNPs[c] | *ZHp* | Gene symbol | Gene name |
|---|---|---|---|---|---|
| LG1B | 1,350,000–1,400,000 | 14 | −6.9171 | *PEMT* | Phosphatidylethanolamine n-methyltransferase |
| LG5 | 7,250,000–7,300,000 | 31 | −6.4337 | *CD276* | cd276 antigen-like |
| LG5 | 29,650,000–29,700,000 | 22 | −7.3606 | – | – |
| LG6 | 9,250,000–9,300,000 | 21 | −7.6198 | *GRM3* | Metabotropic glutamate receptor 3-like |
| LG6 | 20,250,000–20,300,000 | 18 | −6.1522 | – | – |
| LG8 | 3,050,000–3,100,000 | 22 | −6.0127 | na[d] | Zinc finger protein 27-like |
| | | | | *MYO15A* | Unconventional myosin-xv |
| LG8 | 4,400,000–4,450,000 | 42 | −7.0889 | *PALM2* | Uncharacterized protein loc101469015 isoform ×1 |
| | | | | *C28H19ORF21* | Uncharacterized protein c19orf21 homolog |
| LG9 | 14,350,000–14,400,000 | 20 | −6.2349 | *MYLIP* | e3 ubiquitin-protein ligase mylip |
| LG10 | 12,600,000–12,650,000 | 32 | −7.2727 | *CLCN2* | Chloride channel protein 2-like |
| | | | | *ZP2.6* | Zona pellucida glycoprotein |
| | | | | *THPO* | Thrombopoietin precursor |
| | | | | *SLC12A9* | Solute carrier family 12 member 9-like |
| | | | | *PER2* | Period circadian protein homolog 2-like |
| LG10 | 11,800,000–11,850,000 | 25 | −6.0743 | *EXTL2* | Exostosin-like 2-like |
| | | | | *MGC80525* | Uncharacterized oxidoreductase-like |
| | | | | *MGC80525* | Uncharacterized oxidoreductase-like |
| | | | | *RC3H1* | Low quality protein: roquin-like |
| | | | | *KIAA1614* | Uncharacterized protein kiaa1614 homolog |
| | | | | *BC034090* | Uncharacterized protein kiaa1614 |
| | | | | *TAL1* | T-cell acute lymphocytic leukemia protein 1 homolog |
| LG11 | 17,100,000–17,150,000 | 28 | −6.2743 | *NPHP1* | Nephrocystin-1 isoform 1 |
| | | | | *MAL* | Myelin and lymphocyte |
| | | | | *FBXO5* | F-box only protein 5 |
| | | | | *MRPS5* | 28s ribosomal protein mitochondrial |
| LG11 | 19,450,000–19,500,000 | 12 | −6.7757 | – | – |
| LG14 | 15,400,000–15,450,000 | 11 | −6.1268 | *CADM2* | Cell adhesion molecule 2-like |
| LG14 | 15,600,000–15,650,000 | 11 | −8.4682 | – | – |
| LG16 | 13,250,000–13,300,000 | 15 | −7.3058 | *SERINC1* | Serine incorporator 1-like |
| | | | | *HIVEP3* | Transcription factor hivep3-like isoform ×1 |
| LG17 | 11,800,000–11,850,000 | 51 | −6.0971 | *DPYSL5A* | Dihydropyrimidinase-related protein 5-like |
| | | | | *ZMYM1* | Zinc finger mym-type protein 1 |
| | | | | *DNMT3A* | DNA (cytosine-5)-methyltransferase 3a-like |
| LG19 | 17,800,000–17,850,000 | 19 | −7.1685 | *TNKS* | Tankyrase-1-like |

[a] Chromosomes (as reported in Tine et al., 2014).
[b] Genomic windows are reported with the first and last nucleotide position in the corresponding chromosome.
[c] No. of single nucleotide polymorphisms included in the corresponding genomic window.
[d] Gene symbol not available.

in the protein family encoding electroneutral cation-coupled chloride cotransporters that are important for many physiological processes that control cell volume, neuronal chloride concentration, intercellular ion movement and blood pressure (Arroyo et al., 2013).

Pathway gene analysis using KEGG 2015 suggested that selection signatures might affect amino acid metabolism (confirming what was already evidenced for the GO Biological Processes analysis), including the *DNMT3* and *PEMT* genes, and circadian clock and related neurological

**Table 3**
Gene enrichment functional analyses for genes included in the significant genomic windows of the *ZHp* analysis. Terms and related metrics are reported for Gene Ontology for Biological Processes, KEGG and PANTHER pathway analyses. Only terms with P-value < 0.10 are reported.

| Terms | P-value | Adjusted P-value | Z-score | Combined Score | Genes |
|---|---|---|---|---|---|
| *Gene ontology — biological processes* | | | | | |
| Anion transmembrane transporter activity (GO:0008509) | 0.005 | 0.121 | −2.433 | 5.146 | *SERINC1*, *CLCN2*, *SLC12A9* |
| S-adenosylmethionine-dependent methyltransferase activity (GO:0008757) | 0.011 | 0.121 | −2.394 | 5.065 | *DNMT3A*, *PEMT* |
| Voltage-gated chloride channel activity (GO:0005247) | 0.024 | 0.167 | −2.817 | 5.048 | *CLCN2* |
| Chloride transmembrane transporter activity (GO:0015108) | 0.007 | 0.121 | −2.321 | 4.910 | *CLCN2*, *SLC12A9* |
| Histone deacetylase binding (GO:0042826) | 0.005 | 0.121 | −2.314 | 4.894 | *PER2*, *TAL1* |
| Inorganic anion transmembrane transporter activity (GO:0015103) | 0.012 | 0.121 | −2.287 | 4.839 | *CLCN2*, *SLC12A9* |
| Voltage-gated anion channel activity (GO:0008308) | 0.031 | 0.167 | −2.631 | 4.713 | *CLCN2* |
| RNA polymerase binding (GO:0070063) | 0.032 | 0.167 | −2.627 | 4.706 | *PER2* |
| Peptidase activator activity involved in apoptotic process (GO:0016505) | 0.031 | 0.167 | −2.626 | 4.705 | *MAL* |
| Glutamate receptor activity (GO:0008066) | 0.039 | 0.167 | −2.537 | 4.545 | *GRM3* |
| Transcription cofactor binding (GO:0001221) | 0.015 | 0.132 | −2.244 | 4.537 | *PER2* |
| Neutral amino acid transmembrane transporter activity (GO:0015175) | 0.035 | 0.167 | −2.479 | 4.442 | *SERINC1* |
| *KEGG pathway* | | | | | |
| Methionine metabolism | 0.033 | 0.109 | −1.621 | 3.597 | *DNMT3A* |
| Circadian rhythm | 0.022 | 0.109 | −1.409 | 3.126 | *PER2* |
| Glycine serine and threonine metabolism | 0.080 | 0.180 | −1.691 | 2.902 | *PEMT* |
| Heparan sulfate biosynthesis | 0.036 | 0.109 | −1.299 | 2.883 | *EXTL2* |
| *PANTHER pathway* | | | | | |
| Metabotropic glutamate receptor group II pathway | 0.031 | 0.072 | −1.715 | 4.508 | *GRM3* |
| Ionotropic glutamate receptor pathway | 0.043 | 0.072 | −1.294 | 3.402 | *GRM3* |
| Circadian clock system | 0.009 | 0.047 | −0.949 | 2.907 | *PER2* |

systems (as also confirmed by PANTHER analysis), listing the period circadian clock 2 (*PER2*) and the glutamate receptor, metabotropic 3 (*GRM3*) genes (Table 3). These genes are located in two different significant windows (LG10 and LG6, respectively). PER2 is a key element regulating circadian rhythms of activity, metabolism and behavior (Pegoraro and Tauber, 2011). Pituitary *PER2* gene expression in *D. labrax* was reported to be mostly antiphasic, with high values during the day and low value during the night, and in the liver its daily expression differed depending on the feeding regime, confirming its alternant expression level of circadian systems (Herrero and Lepesant, 2014). Studies in model animals are supporting an important role of metabotropic glutamate receptor group II (that include the *GRM3* gene) functions in sleep and circadian regulation. Pritchett et al. (2015) reported that deletion of *Grm3* disrupted sleep and activity and increased the sensitivity of the circadian system to light in mice supporting a role of this gene in photic entrainment and sleep regulation pathways.

Modification of circadian rhythms and behavioral traits are considered forms of adaptation to faming or agronomical conditions and driven by major biological mechanisms that have been at the basis of the domestication processes of animal and plant species (Price, 1999; Kasahara et al., 2010; Bendix et al., 2015; Carneiro et al., 2014, 2015). Benhaïm et al. (2013) reported behavioral differences between hatchery and wild-caught European sea bass. Our results in European sea bass provided a preliminary indication that similar processes are occurring in this species in which domestication is recent and still under consolidation, considering also the current breeding practices based on restocking.

## 4. Conclusions

In this study we re-sequenced the European sea bass genome from DNA pools using a next generation semiconductor-based sequencing technology. We identified about 1.6 millions of SNPs covering the whole genome, obtaining a large catalog of variants, including polymorphisms in most of the annotated genes of the first version of the *D. labrax* assembled genome. Then we designed a pilot approach to capture in a cost effective manner putative selection signatures in the genome of this teleost species. Interestingly, selection signals were in regions including genes whose function might be relevant in terms of adaptation to farming conditions. The obtained results seem in line to what has been described in domestic species in which artificial directional selection, even if in a quite short period of time, is able to shape the genome of a population (Fontanesi et al., 2015). Results obtained with this pilot study should be confirmed, designing investigations including more fishes and more farmed hatcheries. The methodological approaches that we developed can be easily implemented in more complex experimental designs that will use DNA pooling and whole genome sequencing for the identification of selection signatures not only in *D. labrax* but also in many other marine and aquaculture species.

## Conflicts of interest

No conflicts of interest are declared.


## Acknowledgments

The authors thank personnel of the two hatcheries, Dr. Fabio Marino, Dr. Enrico D'Alessandro (University of Messina) and Dr. Marco Ciro Ghionda for their collaboration in this study and for providing samples. This study was supported by University of Bologna RFO2004 and PanLab Project PONa3_00166/F1 funds.


## Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2016.03.007.

## References


Arroyo, J.P., Kahle, K.T., Gamba, G., 2013. The SLC12 family of electroneutral cation-coupled chloride cotransporters. Mol. Asp. Med. 34, 288–298.

Bendix, C., Marshall, C.M., Harmon, F.G., 2015. Circadian clock genes universally control key agricultural traits. Mol. Plant 8, 1135–1152.

Benhaïm, D., Bégout, M.L., Lucas, G., Chatain, B., 2013. First insight into exploration and cognition in wild caught and domesticated sea bass (*Dicentrarchus labrax*) in a maze. PLoS One 8, e65872.

Bertolini, F., Schiavo, G., Scotti, E., Ribani, A., Martelli, P.L., Casadio, R., Fontanesi, L., 2014. High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. Anim. Genet. 45, 304–307.

Bertolini, F., Scimone, C., Geraci, C., Schiavo, G., Utzeri, V.J., Chiofalo, V., Fontanesi, L., 2015. Next generation semiconductor based sequencing of the donkey (*Equus asinus*) genome provided comparative sequence data against the horse genome and a few millions of single nucleotide polymorphisms. PLoS One 10, e0131925.

Boland, J.F., Chung, C.C., Roberson, D., Mitchell, J., Zhang, X., Im, K.M., He, J., Chanock, S.J., Yeager, M., Dean, M., 2013. The new sequencer on the block: comparison of life technology's proton sequencer to an Illumina HiSeq for whole-exome sequencing. Hum. Genet. 132, 1153–1163.

Bovo, S., Bertolini, F., Schiavo, G., Mazzoni, G., Dall'Olio, S., Fontanesi, L., 2015. Reduced representation libraries from DNA pools analysed with next generation semiconductor based-sequencing to identify SNPs in extreme and divergent pigs for back fat thickness. Int. J. Genomics 2015, 950737.

Carneiro, M., Rubin, C.J., Di Palma, F., Albert, F.W., Alföldi, J., Barrio, A.M., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J.M., Barrell, D., Bolet, G., Boucher, S., Burbano, H.A., Campos, R., Chang, J.L., Duranthon, V., Fontanesi, L., Garreau, H., Heiman, D., Johnson, J., Mage, R.G., Peng, Z., Queney, G., Rogel-Gaillard, C., Ruffier, M., Searle, S., Villafuerte, R., Xiong, A., Young, S., Forsberg-Nilsson, K., Good, J.M., Lander, E.S., Ferrand, N., Lindblad-Toh, K., Andersson, L., 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science 345, 1074–1079.

Carneiro, M., Piorno, V., Rubin, C.J., Alves, J.M., Ferrand, N., Alves, P.C., Andersson, L., 2015. Candidate genes underlying heritable differences in reproductive seasonality between wild and domestic rabbits. Anim. Genet. 46, 418–425.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.

Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A., Volckaert, F.A., 2005. A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. Genetics 170, 1821–1826.

Chistiakov, D.A., Tsigenopoulos, C.S., Lagnel, J., Guo, Y.M., Hellemans, B., Haley, C.S., Volckaert, F.A., Kotoulas, G., 2008. A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. Anim. Genet. 39, 623–634.

Cutter, A.D., Payseur, B.A., 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14, 262–274.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

FAO, 2012. FishStat Database.

Fontanesi, L., Schiavo, G., Scotti, E., Galimberti, G., Calò, D.G., Samorè, A.B., Gallo, M., Russo, V., Buttazzoni, L., 2015. A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. J. Anim. Breed. Genet. 132, 239–246.

Guyon, R., Senger, F., Rakotomanga, M., Sadequi, N., Volckaert, F.A., Hitte, C., Galibert, F., 2010. A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: synteny analysis with model fish genomes. Genomics 96, 228–238.

Herrero, M.J., Lepesant, J.M., 2014. Daily and seasonal expression of clock genes in the pituitary of the European sea bass (*Dicentrarchus labrax*). Gen. Comp. Endocrinol. 208, 30–38.

Jentsch, T.J., 2015. Discovery of CLC transport proteins: cloning, structure, function and pathophysiology. J. Physiol. 593, 4091–4109.

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C., Anderson, N., Biagi, T.M., Patterson, N., Pielberg, G.R., Kulbokas 3rd., E.J., Comstock, K.E., Keller, E.T., Mesirov, J.P., von Euler, H., Kämpe, O., Hedhammar, A., Lander, E.S., Andersson, G., Andersson, L., Lindblad-Toh, K., 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat. Genet. 39, 1321–1328.

Kasahara, T., Abe, K., Mekada, K., Yoshiki, A., Kato, T., 2010. Genetic variation of melatonin productivity in laboratory mice under domestication. Proc. Natl. Acad. Sci. U. S. A. 107, 6412–6417.

Kuhl, H., Beck, A., Wozniak, G., Canario, A.V., Volckaert, F.A., Reinhardt, R., 2010. The European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low coverage shotgun sequencing. BMC Genomics 11, 68.

Kuhl, H., Tine, M., Beck, A., Timmermann, B., Kodira, C., Reinhardt, R., 2011a. Directed sequencing and annotation of three *Dicentrarchus labrax* L. chromosomes by applying Sanger- and pyrosequencing technologies on pooled DNA of comparatively mapped BAC clones. Genomics 98, 202–212.

Kuhl, H., Tine, M., Hecht, J., Knaust, F., Reinhardt, R., 2011b. Analysis of single nucleotide polymorphisms in three chromosomes of European sea bass *Dicentrarchus labrax*. Comp. Biochem. Physio.l Part D Genomics Proteomics 6, 70–75.

Lam, S.H., Lui, E.Y., Li, Z., Cai, S., Sung, W.K., Mathavan, S., Lam, T.J., Ip, Y.K., 2014. Differential transcriptomic analyses revealed genes and signaling pathways

involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, *Oreochromis mossambicus*. BMC Genomics 15, 921.
Li, H., 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM (arXiv preprint arXiv, 1303.3997).
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
Liu, Q., Guo, Y., Li, J., Long, J., Zhang, B., Shyr, Y., 2012. Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. BMC Genomics Suppl 8, S8.
Merriman, B., Ion Torrent R&D Team, Rothberg, J.M., 2012. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis 33, 3397–3417.
Molnár, J., Nagy, T., Stéger, V., Tóth, G., Marincs, F., Barta, E., 2014. Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. BMC Genomics 15, 761.
Nebel, C., Romestand, B., Nègre-Sadargues, G., Grousset, E., Aujoulat, F., Bacal, J., Bonhomme, F., Charmantier, G., 2005. Differential freshwater adaptation in juvenile sea-bass *Dicentrarchus labrax*: involvement of gills and urinary system. J. Exp. Biol. 208, 3859–3871.
Pegoraro, M., Tauber, E., 2011. Animal clocks: a multitude of molecular mechanisms for circadian timekeeping. Wiley Interdiscip. Rev. RNA 2, 312–320.
Pickett, G.D., Pawson, M.G., 1994. Sea Bass. Biology, Exploitation and Conservation vol. 12. Chapman & Hall.
Price, E.O., 1999. Behavioral development in animals undergoing domestication. Appl. Anim. Behav. Sci. 65, 245–271.
Pritchett, D., Jagannath, A., Brown, L.A., Tam, S.K., Hasan, S., Gatti, S., Harrison, P.J., Bannerman, D.M., Foster, R.G., Peirson, S.N., 2015. Deletion of metabotropic glutamate receptors 2 and 3 (mGlu2 & mGlu3) in mice disrupts sleep and wheel-running activity, and increases the sensitivity of the circadian system to light. PLoS One 10, e0125523.
Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13, 341.
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.
Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348–352.
Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, O., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591.
Rubin, C.J., Megens, H.J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A., Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. U. S. A. 109, 19529–19536.
Sánchez-Vázquez, F.J., Azzaydi, M., Martínez, F.J., Zamora, S., Madrid, J.A., 1998. Annual rhythms of demand-feeding activity in sea bass: evidence of a seasonal phase inversion of the diel feeding pattern. Chronobiol. Int. 15, 607–622.
Schlötterer, C., Tobler, R., Kofler, R., Nolte, V., 2014. Sequencing pools of individuals mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15, 749–763.
Smith, C.T., Elfstrom, C.M., Seeb, L.W., Seeb, J.E., 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol. Ecol. 14, 4193–4203.
Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192.
Tine, M., Kuhl, H., Gagnaire, P.A., Louro, B., Desmarais, E., Martins, R.S., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5, 5770.
Varsamos, S., Connes, R., Diaz, J.P., Barnabé, G., Charmantier, G., 2001. Ontogeny of osmoregulation in the European sea bass *Dicentrarchus labrax* L. Mar. Biol. 138, 909–915.
Varsamos, S., Diaz, J.P., Charmantier, G., Flik, G., Blasco, C., Connes, R., 2002. Branchial chloride cells in sea bass (*Dicentrarchus labrax*) adapted to fresh water, seawater, and doubly concentrated seawater. J. Exp. Zool. 293, 12–26.
Volckaert, F.A.M., Batargias, C., Canario, A., Chatziplis, D., Chistiakov, D., Haley, C., Libertini, A., Tsigenopoulos, C., 2008. European Sea Bass. Genome Mapping and Genomics in Fishes and Aquatic Animals. Springer, Berlin Heidelberg, pp. 117–133.