UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Biomedicina Comparata e Alimentazione

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE VETERINARIE

INDIRIZZO COMUNE

CICLO XXVIII

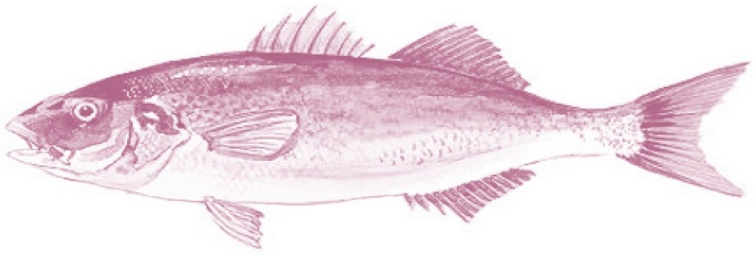
Development of high-throughput technologies for species of veterinary relevance. Investigation of the genetic basis of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Direttore della Scuola: Ch.mo Prof. Gianfranco GABAI

Coordinatore d'indirizzo: Ch.mo Prof. Giuseppe RADAELLI

Supervisore: Ch.mo Prof. Luca BARGELLONI

Dottorando: Massimiliano BABBUCCI



A mia nipote,

Carlotta.

Table of Contents

Riassunto

Negli ultimi 20 anni, l'allevamento di specie ittiche marine ha avuto un rapido incremento, grazie principalmente al continuo sviluppo e miglioramento delle tecniche di produzione. Rimangono comunque ancora da risolvere diversi problemi nei processi di produzione, come ad esempio la presenza di malformazioni scheletriche, l'elevata mortalità delle larve o la suscettibilità allo stress e le malattie. Le anomalie scheletriche nei pesci d'allevamento incidono in maniera rilevante anche sul benessere e la salute degli animali, causando notevoli perdite economiche per gli allevatori. Tuttavia, il costante progresso delle tecniche di biologia molecolare e di genetica hanno permesso di ampliare notevolmente le nostre conoscenze sui meccanismi molecolari alla base di molti caratteri produttivi di rilevanza economica nelle specie allevate. L'obiettivo principale di questo studio è stato quello di indagare le basi genetiche del prognatismo mandibolare (MP), una malformazione scheletrica presente, con una moderata incidenza, negli allevamenti intensivi di branzino (*Dicentrarchus labrax*). A questo scopo, abbiamo applicato un particolare protocollo di analisi, il 2bRAD (type IIB endonucleases restriction-site associated DNA), ampiamente utilizzato per lo sviluppo di marcatori genetici affiancato a tecniche di sequenziamento massivo NGS (Next Generation Sequencies).

L'affidabilità e la riproducibilità del protocollo 2bRAD è stata prima testata su due organismi di una certa rilevanza veterinaria: il tonno pinna gialla (*Thunnus albacares*), un pesce importante sia dal punto di vista biologico che economico, e la *Listeria monocytogenes*, un batterio patogeno che può causare la listeriosi. Nel tonno pinna gialla sono stati identificati 6772 marcatori SNP (Single Nucleotide Polymorphism) in popolazioni campionate su tutto l'areale di distribuzione negli Oceani Atlantico, Indiano e Pacifico. L'analisi discriminante delle componenti principali (DAPC) ha permesso di rilevare la presenza di popolazioni distinte di tonno pinna gialla nei tre Oceani. Questi risultati hanno dimostrato quindi l'efficacia del 2bRAD nello studio della divergenza genetica in un pesce marino ad alto potenziale di dispersione. Un totale di 1279 loci SNP sono stati identificati per *Listeria monocytogenes*, e questo lavoro rappresenta un primo esempio di applicazione di tecnologie "RAD-like" nel campo della genetica microbica. I risultati ottenuti suggeriscono che il 2bRAD potrebbe rivelarsi uno strumento estremamente utile per l'epidemiologia molecolare, così come in altri settori di studio come la filogenesi, la tassonomia, la genetica di popolazione e la salute pubblica. Nel lavoro sul prognatismo mandibolare del branzino, è stata inizialmente sviluppata una mappa di linkage ad alta densità utilizzata per la mappatura dei QTL potenzialmente associati alla patologia; successivamente è stato fatto uno studio di associazione (GWAS) scansionando l'intero genoma di branzino per trovare marcatori SNP putativamente associati al prognatismo. Sono stati utilizzati 298 esemplari giovanili in totale, di cui 148 appartenenti a quattro famiglie full-sib. Un totale di 7362 marcatori SNP sono stati genotipizzati in più del 80% della popolazione sperimentale.

Tre QTL significativi sono stati rilevati per il prognatismo utilizzando un'analisi di regressione *half-sib*. Il primo è stato mappato sul gruppo di linkage LG18, il secondo su LG20 e il terzo sul gruppo di linkage LG22; ogni QTL spiegava circa l'11-13% della variazione fenotipica. Due SNP associati al MP sono stati identificati con l'analisi GWAS. I marcatori candidati sono stati individuati sul cromosoma ChrX e sul cromosoma Chr17, in stretta vicinanza ai due QTL più significativi. In particolare, il marcatore SNP su Chr17 è posizionato all'interno della regione codificante del gene Sobp, che svolge un ruolo fondamentale nello sviluppo craniofacciale. Inoltre, l'analisi dei geni differenzialmente espressi nei branzini prognati ha evidenziato lo "sviluppo del sistema nervoso" come un *pathway* cruciale nella formazione del MP. In particolare, Zic2, un gene chiave per la morfogenesi craniofacciale nelle specie modello, risulta significativamente sotto-espresso negli animali affetti da prognatismo. Il lavoro svolto in questo studio quindi, integrando la trascrittomica e l'analisi di marcatori molecolari sviluppati sull'intero genoma, fornisce validi risultati per comprendere meglio i meccanismi molecolari che sono alla base dell'insorgenza del prognatismo mandibolare nei branzini di allevamento.

Abstract

In the last 20 years, the production of farmed marine fish species has increased rapidly, mainly as a consequence of improved breeding methods and technologies. Skeletal malformations and others severe production bottlenecks such as high larval mortality or susceptibility to stress and disease, remain to be solved. Skeletal anomalies in farmed fish are a relevant issue affecting animal welfare and health and cause significant economic losses. The constant progress of genomic technologies promises to rapidly increase our knowledge on molecular mechanisms underlying productive traits of economic relevance in farmed species.

The main objective of this PhD research was to investigate the molecular basis underlying the mandibular prognathism (MP), a skeletal malformation with a moderate incidence in intensively reared European seabass (*Dicentrarchus labrax*). To this aim, we applied a 2bRAD (type IIB endonucleases restriction-site associated DNA) pipeline, a widely used NGS (Next Generation Sequencing) technique for genome-wide genotyping. The reliability and reproducibility of the 2bRAD protocol has been preliminary tested on two organisms of veterinary relevance: the yellowfin tuna (*Thunnus albacares*), a fish with great biological and economic importance at global scale and the *Listeria monocytogenes*, a pathogenic bacterium that causes the infection listeriosis. In total, 6772 high-quality genome-wide SNPs (Single Nucleotide Polymorphism) were identified for the yellowfin tuna, in across population samples collected from the Atlantic, Indian and Pacific oceans and covering the entire distribution area. Discriminant Analysis of Principal Components (DAPC) endorsed the presence of genetically discrete yellowfin tuna populations among three oceanic pools. These results showed the efficiency of this genotyping technique in assessing genetic divergence in a marine fish with high dispersal potential.

A total of 1279 unique SNPs loci were identified for *Listeria monocytogenes*. This research represents a first example of application of RAD-like technologies in microbial genetics. Results obtained for *L. monocytogenes* suggest that 2b-RAD is an effective tool for molecular epidemiology and public health, as well as proved in other areas such as phylogenetics, taxonomy, population genetics and biosafety. Once the 2bRAD technique was optimized as described above, a high-density genetic map of European seabass for QTL mapping of jaw deformity was constructed and a genome-wide association study (GWAS) was carried out on a total of 298 juveniles, 148 of which belonged to four full-sib families. A total of 7362 SNP markers was genotyped in more than 80% of the experimental population.

Three significant QTLs were detected as significantly associated to MP by applying a half-sib regression analysis. The first QTL was located on linkage group LG18, the second one on LG20 and the third on LG22, each of them explaining 11-13% of the phenotypic variation. Two SNPs associated

with MP were identified with GWAS analysis. Candidate markers were located on chromosome ChrX and chromosome Chr17, both in close proximity with the peaks of the two most significant QTLs. Notably, the SNP marker on Chr17 was positioned within the Sobp (Sine Oculis Binding Protein) gene coding region, which plays a pivotal role in craniofacial development. The analysis of differentially expressed genes in jaw-deformed animals highlighted the "nervous system development" as a crucial pathway in MP. In particular, Zic2, a key gene for craniofacial morphogenesis in model species, was significantly down-regulated in MP-affected animals.

By integrating transcriptomic and GWA methods, the analysis developed during this PhD study, provides evidence for putative mechanisms underlying seabass jaw deformity.

1. Introduction

1.1 A brief history of aquaculture in Europe

Aquaculture refers to the breeding, rearing and harvesting of aquatic organisms, mainly fish crustaceans and mollusks, in all types of water environments including ponds, rivers, lakes, and ocean. Marine aquaculture has ancient roots, it is present in human history since the dawn of civilization. Fishery and aquaculture remain important sources of food and income for hundreds of millions of people around the world. Worldwide, as reported in "World review of fisheries and aquaculture 2016", since 1995 the global production of farmed fish has roughly tripled in volume reaching the milestone of over 80 million tons as recorded in 2014. According to recent FAO estimates, the per capita consumption of fish has increased from 10 kg in 1960 to more than 20 kg in 2014. Fishery and aquaculture are the main means of livelihood for more than 12% of the world population. The expansion of aquaculture has thus contributed to both improve the quality of the diet of many people, especially in poor rural areas, and support consumption in western countries in view of the constant reduction of natural fish stocks.

In Europe 50% of aquaculture products is made from shellfish, mostly mussels and oysters. The marine fish, such as sea bass (*Dicentrarchus labrax*), sea bream (*Sparus aurata*) and salmon (*Salmo salar*) account for about 27% of farmed fish, while freshwater species such as trout (*Oncorhynchus mykiss*) and carp (*Cyprinus carpio*) for about 23% (European Commission for Maritime Affairs and Fisheries).

Depending on the type of management, aquaculture can mainly be “extensive” or “intensive”. Typically, the extensive aquaculture entails fish-farming in natural valleys or lagoons. Human activity is limited to preparation of the basins by checking the quality and condition of water bed and embankments and to seed and select juveniles including health checks. Fish grow depending on the stocking density and environmental conditions. Food is not provided and fish obtain it from the surrounding environment as in natural conditions. In intensive aquaculture, the stocking density is increased beyond the natural productivity of the basin, and the feeding availability is artificially increased with extra food. For high-value species, a recycle aquaculture system (RAS) technology is often used, where all the production parameters are controlled. By recycling it, very little water is used per unit of production but the process does have high capital and operating costs. The higher cost structures mean that RAS is only economical for high-value products such as broodstock for egg production, sturgeon production and research animals (Avnimelech et al., 2008; Weaver, 2006).

1.2 Seabass farming in the Mediterranean Sea

Dicentrarchus labrax (Linnaeus, 1758), seabass, is a marine teleost belonging to the family of *Moronidae*. It naturally occurs in the Atlantic Ocean (from Norway to Canaries), the Mediterranean Sea and Black Sea. It is adapted to both open sea and brackish waters and often moves between these habitats following the tidal flow. It preferentially occurs along the sea coasts between 10 and 100 m of deepth.

At average, individuals caught in nature can be 45 cm long and 5 kg heavy although adults can reach one meter in length and 12 kg in weight (Porcellotti, 2005). Adults of seabass generally live in pairs or isolated, while juveniles and sub-adults are gregarious. In spring, juveniles perform trophic migrations from coast to warmer brackish waters in lagoons. In winter (January-February), young adults perform reproductive migration from freshwater and brackish to the sea (Agbayani, 1999).

The seabass is an economically valuable species with a role of primary importance in aquaculture, especially in the Mediterranean (in France, Italy and Spain) (Chavanne et al., 2009).

In the last decade, the introduction of farming in sea cages has significantly incremented the yield. The final product is almost entirely marketed as fresh, whole fish, while only a small part is further processed. In the past, the seabass was farmed mostly in coastal lagoons and the supply of fry depended totally on the number of specimens caught in estuarine areas of rivers. Between the 60ies and the 70ies in France and Italy, the knowledge and technologies for mass production of fry were acquired. The ensuing standardization of techniques for the production and transport of eggs, larvae and juveniles enabled the enhancing of aquaculture of seabass.

Usually, the farming process consists in positioning of a series of barriers in a lagoon to catch the fish during the autumn migration towards open sea. In these enclosures seabass reaches the weight of 300-500g in 37 months. Often used are also cages in the sea especially in the Mediterranean. In these systems, larger breeding volumes than those utilized by the intensive installations allow reduction of costs with no need to pump water from the outside. In almost all farms using cages, juveniles are previously pre-fatten as fry (2-3 months) in facilities on land. Phases of the process are: A) production of juveniles: every hatchery has its own stock of carefully selected brood stock age cohorts. Females reach the best reproductive performance typically between 5 and 8 years while for males the optimal age is between 2 and 4 years. At sexual maturity, breeding stock is placed in tanks, usually with a male to female ratio of 2:1. After hatching, larvae are reared in small modules in high density (30-150 larvae/l) and fed live prey or alternatively, to reduce costs, with microencapsulated inert food. About two months, at weight of 2-5g fry is transferred to the companies for fattening. B) fattening: generally, it takes place in floating cages installed at a short distance from the coast. Alternatively,

some companies use tanks positioned on the ground, powered with a recirculation system which allows to control the water temperature. Farmed juveniles are collected at the weight of 300-500g.
C) consumption: the seabass is generally sold fresh and cleaned, usually by large retailers and restaurants. To date, the annual production of seabass from aquaculture facilities is greater than 155,000 tons per year (FAO fishstat, http://fao.org/fishery/species/2291/en).

1.3 Malformations in farmed fish

1.3.1 Deformities

The skeletal abnormalities in teleost fish are still a severe problem in aquaculture sector with serious consequences both economically and biologically. The incidence of deformities varies among farmed species, farming conditions and batches within the same breeding event or even within the same batch of eggs (Boglione et al., 2013). Recent advances in the field with improved methods of incubation of larvae and the scrupulous and constant monitoring of the diseases have greatly increased the quality of fry and their rate survival. Despite this, deformities still occur. This incidence of skeletal deformities in farmed fish prompts to better understand the knowledge of the genetic and epigenetic factors that causing or contributing to malformations. The scenario is complicated by the following observations: (i) several non-genetic factors can produce the same abnormality in different species; (ii) the same factor can induce abnormalities in different species (Boglione and Costa, 2011): (iii) abnormalities can be induced by different factors in different cohorts of the same species (Kause et al., 2007); (iv) the sensitivity to a causative factor may significantly change during ontogeny (Mazurais et al., 2009); (v) a single causative factor can be compensated by the contribute of other different factors (Sfakianakis et al., 2006); (vi) factors are strongly correlated with abnormalities in specific regions of the body only in some species and not in others (Koumoundouros, 2010); (vii) in the same individual, a causative factor can trigger abnormalities in some parts of the skeleton and not in others although characterized by the same skeletal tissue and ossification processes (Fernández and Gisbert, 2011).
Deformities bring about significant economic consequences for farmers. Skeletal abnormalities do not always reduce the performance in terms of growth and resistance to stress conditions (manipulation, anesthesia, overcrowding, infections), but have a major impact on production costs and on commercial value of farmed fish. Usually, the rate of malformed individuals is kept in check under than 4-5% by applying a manual selection of fry. This approach is obviously highly time consuming and expensive in terms of manpower involved. It also implies a not negligible problem of animal welfare (Boglione et al., 2013).

Multiple triggers are implied in developmental abnormalities: environmental factors (e.g. density of the embryos during incubation, stress during vitellogenesis, contaminants in the water, excessive turbulence in hatching tanks), nutritional factors (deficiencies in the absorption of the ascorbic acid, excess of tryptophan or vitamin A and D in food) and toxicity factors such as the presence of contaminants of various origin (e.g. the material used to make the tanks).

Only recently, scientists have pointed out the importance of genetic factors in etiology of deformity. The development of innovative techniques in molecular biology and genetic analysis using markers such as microsatellites and SNPs (Single Nucleotide Polymorphism) have suggested new scenery in understanding the genetic basis of malformations in farmed fish. Several studies have been conducted to determine the heritability of some skeletal abnormalities. These studies have specifically investigating the potential existence of specific mutations occurring in malformed fish, the effects on the phenotype and the correlation with ploidy level and inbreeding rate (Boglione et al., 2013; Castro et al., 2006, 2004; Ferguson and Danzmann, 1998).

(Kause et al., 2007) suggest that the genetic predisposition to develop skeletal malformations becomes visible only in the presence of specific environmental conditions (e.g. strong variations in temperature or rearing system's failure). In their study, they report that in farmed salmon the heritability of skeletal malformations was zero when no malfunctions occurred in the rearing system. As a failure occurred, the incidence of abnormalities increased unusually resulting in high heritability of the trait.

1.3.2 Mandibular prognathism

Mandibular prognathism is a development malformation associated with a defined dental malocclusion resulting in a particularly distinctive facial phenotype (underbite).

This malformation can be found in many vertebrate classes (reptilia, aves, mammalia, actinopterygii and amphibia) as well as in some artificial breeds of brachycephalic dogs (Bouillon et al., 2013; Kuzir et al., 2004; Ritchie et al., 1997). This is one of the most widespread malformations in horses. It leads to important consequences in the correct masticatory function and, especially in racehorses, can hamper the positioning of the bite (Signer-Hasler et al., 2014).

In teleost fish the anatomical structures of the jaws are developed from the first visceral arch that, during the organogenesis, originates the cranial portion of splanchnocranium splitting into palate-square cartilage and Meckel-cartilage. From the Meckel-cartilage and articulated through the square bone to the cranium, the dental bone develops frontally and the angular bone posteriorly. These structures constitute the base of the jaws. An exaggerated extension of the skeleton cartilaginous or

an impaired ossification of dental portion involved in the formation of the mandibular body could likely be implied in the development of prognathism. In sea bass larvae, cartilaginous structures of the oral cavity become visible only at completion of yolk sac absorption. The mandibular ossification begins after larvae hatch and takes about 80 days (Kuzir et al., 2004).

Although it is established that environmental factors and genetic factors both contribute to the occurrence of the malformation, there is still little information on their mutual interaction. The transmission seemingly follows a Mendelian pattern although polygenic and multifactorial causes are probably coexisting (Cruz et al., 2008). Recent extensive studies on affected families showed a dominant autosomal pattern with incomplete penetrance (Cruz et al., 2008; El-Gheriani et al., 2003; Wolff et al., 1993). A study on 42 Korean and Japanese families showed a significant association between the occurrence of the malformation and three loci located on different chromosomes (Yamaguchi et al., 2005). The corresponding genes encode for alkaline phosphatase, heparan-sulfate proteoglycan II and the matrilin-1 (protein of the cartilage matrix). These proteins are involved in growth of cartilage tissue (heparan sulfate proteoglycan II), in its maturation (matrilin-1), and in the mineralization of bone tissue (alkaline phosphatase). The matrilin-1 gene (*matn1*) and EPB41 (*Erytrocyte membrane protein band 41*) have been recently associated with the development of an underbite (Jang et al., 2010; Xue et al., 2010). Experiments on zebrafish larvae (*Danio rerio*) have revealed a potentially interesting role of the *Hh Hedgehog* genes (*Sonic Hedgehog a and b*) in the normal development of the jaw. Proteins coded by *Hh* genes act as morphogenic factors and are involved in different signal pathways of organogenesis of vertebrates. Mutations of Sonic Hedgehog or treatments with toxic compounds such as *2,3,7,8-tetrachlorodibenzo-p-dioxin* (TCDD) altering the *Hh* signal pathway seem to trigger alterations in the normal mandibular development (Teraoka et al., 2006). Other compounds delivered via a wrong diet have resulted in the emergence of various deformities. The excess of vitamin A in sea bass juveniles causes a major deficit in the development of the premaxillary bones, maxillary, ethmoid and parasfenoid dish This deformation may trigger prognathism as it brings about an underdevelopment of the maxillary arch and the projection of the mandibular arch (Mazurais et al., 2009). As far as the lipids are concerned, fatty acids have been suggested as being involved in the modulation of the transcription of genes taking part in their metabolism and playing a role in the pathway of the retinoic acids (RAR). Retinoic acids control the expression of morphogenetic genes belonging to different families: Hox genes, genes encoding BMP (*Bone Morphogenetic Protein*) and IGF (*Insulin-like Growth Factor*). They can also interact with the expression of Hedgehog genes (Suzuki et al., 1999).Malformations such as the mandibular prognathism in seabass juveniles (Fig.1) are totally compatible with life and rise with no significant change in growth rate during the fattening stage, or increase in susceptibility to disease.

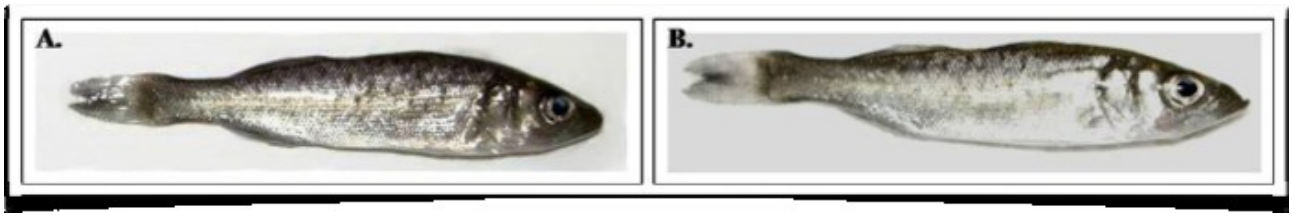


Figure 1. Normal (A.) and prognathic (B.) seabass to 90 days after hatching.

The underbite phenotype becomes less evident with the transition from fry to adult. Nonetheless, breeders reckon this feature aesthetically unpleasant as far as customers are concerned. The sea bass is generally sold as a whole and not filleted, therefore the morphology of the head is one of the main criterion for choice from the customer. Consequently, rearing prognathic individuals is a loss in terms of farmed space and labor employed in the manual selection process and leads to a significant fall in price of the larvae batches sold to fattening centers. All these reasons clearly prompt the scientific community to identify the causes of skeletal malformations, and in this specific case prognathism.

1.4 Genomic applications in aquaculture.

Genetic improvement refers to all those techniques used to increase the productive and reproductive performance of livestock by evaluating and subsequently selecting the brood stock. Most of the aspects related to the conformation (phenotype) are associated to a genotype and then transmitted over generations. All phenotypic (mostly morphologic) characteristics are genetically coded by one or more genes or resulting from the interaction between genotype and environment. Taking advantage of the heritability of characters, it is therefore possible to improve traits of commercial interest (e.g. growth rate, body size, disease resistance) by specific crosses, once the corresponding genetic background has been identified. The recent methodological advancement has allowed to understand the genetic nature of some traits of interest, and to identify specific gene regions or individual loci influencing these features (Andersson, 2001).

MAS (Marker Assisted Selection) allows to identify DNA regions associated with genes of interest (molecular markers) to select efficiently specific qualitative and quantitative traits (e.g. productivity, disease resistance, abiotic stress tolerance) (Duhnam, 2004). The MAS technique is therefore based on the association between a trait and a marker, regardless of phenotype-environment interaction, and eventually enables a genetic-guided selection of individuals at a very early development stage, rather than waiting until phenotypic expression.

Molecular markers are available from a wide variety of genetic studies of aquatic species and are mainly used to trace the genetic profile of single individuals, to allocate parents or to determine the

genetic diversity to select economically relevant genetic trait (Huete-Pérez and Quezada, 2013). Owing to their practical use in the analysis of trait-genotype associations, SNPs (Single Nucleotide Polymorphisms) have become largely popular in aquaculture for genetic improvement programs and, in general, for genome research. SNPs are frequently the first choice when it comes to trace genotypic associations with phenotypic traits owing to their considerable genome coverage, and throughput of assay. The most common approaches used to analyze these datasets are GWAS (Genome-Wide Association Study) and QTL (Quantitative Trait Loci) mapping. GWAS examines the association between widespread genetic variants and specific traits in the genome, transcriptome or proteome. Thanks to the advances in high-throughput genotyping technologies, GWAS may also be used to enhance the efficiency of breeding and selection in aquaculture.

GWAS and QTL analysis are very similar, although GWAS tests at marker positions, whereas QTL analysis examines regions between markers (Huete-Pérez and Quezada, 2013).

Among several genomic tools for the cost-efficient generation of SNP, the recent development of restriction site-associated DNA sequencing (RAD-Seq) has facilitated marker discovery and genotyping at large scale. RAD-Seq markers have proved informative for association and QTL mapping as well as for population genetics and evolutionary research (Houston et al., 2012). All RAD-Seq methods entail the use of restriction enzymes and construction and sequencing of DNA libraries of short fragments. To answer to the biological problems brought up in this thesis, we focused on a particularly cost-effective and flexible RAD-Seq method called 2bRAD (Wang et al., 2012). This method is unique among the RAD-Seq protocols (Fig.2) in that it uses IIB restriction enzymes to produce short fragments that are of equal size across all loci (32–36 bp). Of all the RAD-like methods, 2bRAD produces the shortest reads, so this technique is not recommended for "de novo" locus identification or in the case of large and complex genomes, as the read length is essentially too short to enable reliable mapping (Andrews et al., 2016).

This method is very flexible, cheap and extensively validated (Dou et al., 2016; Guo et al., 2014; Pauletto et al., 2016; Pecoraro et al., 2016). 2bRAD allows to easily adjust the number of genetic markers targeted therefore being well-suited for studies involving large sample sizes or species with large genomes.

All RAD-Seq methods typically produce a large amount of data that can be analyzed with several basic steps of bioinformatics pipelines. If a reference genome is available, loci can then be identified by aligning sequence reads to the reference genome. Alternatively, loci can be assembled de novo by clustering similar sequence reads together assuming that variation among reads at each locus represents allelic variation.

STACKS (Catchen et al., 2013, 2011) is a bioinformatics tool designed to analyze data generated by 2bRAD protocol. STACKS contains several flexible modules to conduct all parts of the analysis, from quality filtering, locus identification (either reference-aligned or de novo) and genotyping to estimation of population genetic parameters (Fig. 3).

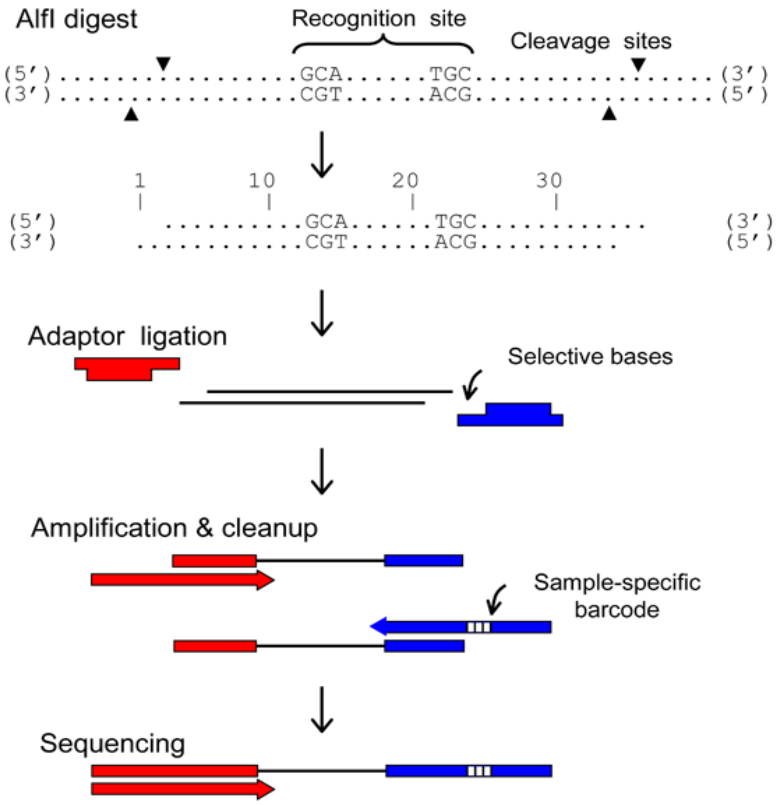


Figure 2. Sample preparation for 2bRAD genotyping is accomplished by restriction digest of genomic DNA, cohesive end ligation of partially double-stranded adaptors with compatible (NN) overhangs, and incorporation of barcodes for multiplex sequencing by PCR. (http://nature.com/protocolexchange/protocols/2356#/).

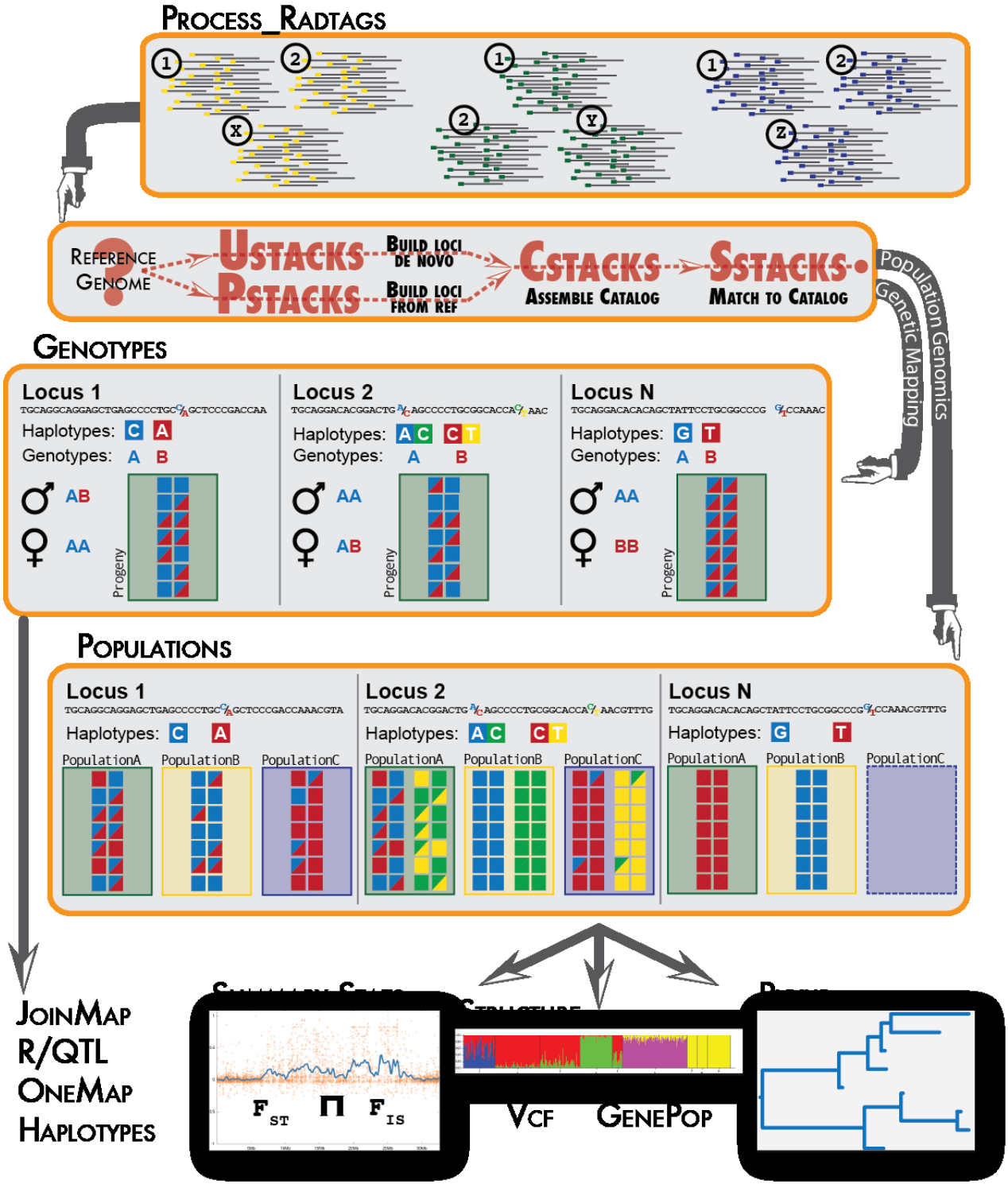


Figure 3. Stacks pipeline workflow. Stacks proceeds in five major stages. First, reads are demultiplexed and cleaned by the process_radtags program. The next three stages comprise the main Stacks pipeline: building loci (ustacks/pstacks), creating the catalog of loci (cstacks), and matching against the catalog (sstacks). In the fifth stage, either the populations or genotypes program is executed, depending on the type of input data (http://catchenlab.life.illinois.edu/stacks/).

1.5 Summary of the papers included in the thesis

The whole PhD project included three independent but related research lines.
During the first two years, the reliability and reproducibility of the 2bRAD protocol have been tested on two species: (i) the yellowfin tuna (*Thunnus albacare*s), a fish of great commercial interest (publication 1, Pecoraro et al., 2016), and (ii) the *Listeria monocytogenes*, a pathogenic bacterium that causes the infection listeriosis and is therefore an organism of great veterinary importance (publication 2, Pauletto et al., 2016). In the third year, the method was successfully applied to investigate the genetic bases of the mandibular prognathism in the farmed seabass (publication 3, Babbucci et al., 2016).

In the first work, we demonstrated the usefulness of 2bRAD genotyping technique to investigate population genetic diversity in species with high gene-flow. A total of 6,772 high-quality genome-wide SNPs were identified across several tuna's population samples.
In the second work, the 2bRAD protocol was applied to characterize *Listeria monocytogenes* strains, and the method compared to the traditional Sanger sequencing approach MultiLocus Sequence typing (MLST). Results demonstrate that 2bRAD predicts MLST types and is often more informative on population structure than MLST.

After these two validation steps, 2bRAD protocol was applied to study the genetic basis of mandibular prognathism (MP) in European seabass (*Dicentrarchus labrax*) using an integrated genomic approach. The MP is a skeletal malformation affecting farmed seabass and bringing about a considerable economic loss to aquaculture facilities. The integrated transcriptomic and genome-wide analysis (GWA) methods provided evidence for putative mechanisms underlying seabass jaw deformity.

1.6 References

Agbayani, E., 1999. **www.fishbase.org/Reproduction/SpawningList, marzo 2009.**

Andersson, L., 2001. **Genetic dissection of phenotypic diversity in farm animals.** Nat. Rev. Genet. 2, 130–138. doi:10.1038/35052563

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. **Harnessing the power of RADseq for ecological and evolutionary genomics.** Nat Rev Genet 17, 81–92. doi:10.1038/nrg.2015.28

Avnimelech, Y., Verdegem, M.C.J., Keshavanath, P., Kurup, B.M., 2008. **Sustainable land-based aquaculture: rational utilization of water, land and feed resources.** Mediterranean Aquaculture Journal 1, 45–55.

Boglione, C., Costa, C., 2011. **Skeletal Deformities and Juvenile Quality**, in: Pavlidis, M.A., Mylonas, C.C. (Eds.), Sparidae. Wiley-Blackwell, Oxford, UK, pp. 233–294.

Boglione, C., Gisbert, E., Gavaia, P., E. Witten, P., Moren, M., Fontagné, S., Koumoundouros, G., 2013. **Skeletal anomalies in reared European fish larvae and juveniles. Part 2: main typologies, occurrences and causative factors.** Reviews in Aquaculture 5, S121–S167. doi:10.1111/raq.12016

Bouillon, K., Kivimaki, M., Hamer, M., Sabia, S., Fransson, E.I., Singh-Manoux, A., Gale, C.R., Batty, G.D., 2013. **Measures of frailty in population-based studies: an overview.** BMC Geriatrics 13, 64. doi:10.1186/1471-2318-13-64

Castro, J., Bouza, C., Presa, P., Pino-Querido, A., Riaza, A., Ferreiro, I., Sánchez, L., Martínez, P., 2004. **Potential sources of error in parentage assessment of turbot (Scophthalmus maximus) using microsatellite loci.** Aquaculture 242, 119–135. doi:10.1016/j.aquaculture.2004.09.010

Castro, J., Pino, A., Hermida, M., Bouza, C., Riaza, A., Ferreiro, I., Sánchez, L., Martínez, P., 2006. **A microsatellite marker tool for parentage analysis in Senegal sole (Solea senegalensis): Genotyping errors, null alleles and conformance to theoretical assumptions.** Aquaculture 261, 1194–1203. doi:10.1016/j.aquaculture.2006.09.001

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. **Stacks: an analysis tool set for population genomics.** Mol. Ecol. 22, 3124–3140. doi:10.1111/mec.12354

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H., 2011. **Stacks: Building and genotyping Loci de novo from short-read sequences.** G3 (Bethesda) 1. doi:10.1534/g3.111.000240

Chavanne H, Chatain B, Haffray P: **Review on Breeding and Reproduction of European aquaculture species.** Aqua Breeding 2009, www.aquabreeding.eu.

Cruz, R.M., Krieger, H., Ferreira, R., Mah, J., Hartsfield, J., Oliveira, S., 2008. **Major gene and multifactorial inheritance of mandibular prognathism.** Am. J. Med. Genet. A 146A, 71–77. doi:10.1002/ajmg.a.32062

Dou, J., Li, X., Fu, Q., Jiao, W., Li, Y., Li, T., Wang, Y., Hu, X., Wang, S., Bao, Z., 2016. **Evaluation of the 2b-RAD method for genomic selection in scallop breeding.** Scientific Reports 6, 19244. doi:10.1038/srep19244

Dunham, R.A., 2004. **Aquaculture and fisheries biotechnology-genetic approaches.** CABI publishing, Oxford

El-Gheriani, A.A., Maher, B.S., El-Gheriani, A.S., Sciote, J.J., Abu-Shahba, F.A., Al-Azemi, R., Marazita, M.L., 2003. **Segregation analysis of mandibular prognathism in Libya.** J. Dent. Res. 82, 523–527.

Ferguson, M.M., Danzmann, R.G., 1998. **Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting?** Canadian Journal of Fisheries and Aquatic Sciences 55, 1553–1563. doi:10.1139/f98-096

Fernández, I., Gisbert, E., 2011. **The effect of vitamin A on flatfish development and skeletogenesis: A review.** Aquaculture 315, 34–48. doi:10.1016/j.aquaculture.2010.11.025

Guo, Y., Yuan, H., Fang, D., Song, L., Liu, Y., Liu, Y., Wu, L., Yu, J., Li, Z., Xu, X., Zhang, H., 2014. **An improved 2b-RAD approach (I2b-RAD) offering genotyping tested by a rice (Oryza sativa L.) F2 population.** BMC Genomics 15, 956. doi:10.1186/1471-2164-15-956

Houston, R.D., Davey, J.W., Bishop, S.C., Lowe, N.R., Mota-Velasco, J.C., Hamilton, A., 2012. **Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon.** BMC Genomics 13. doi:10.1186/1471-2164-13-244

Huete-Pérez, J.A., Quezada, F., 2013. **Genomic approaches in marine biodiversity and aquaculture.** Biol. Res. 46, 353–361. doi:10.4067/S0716-97602013000400007

Jang, J.Y., Park, E.K., Ryoo, H.M., Shin, H.I., Kim, T.H., Jang, J.S., Park, H.S., Choi, J.Y., Kwon, T.G., 2010. **Polymorphisms in the Matrilin-1 Gene and Risk of Mandibular Prognathism in Koreans.** Journal of Dental Research 89, 1203–1207. doi:10.1177/0022034510375962

Kause, A., Ritola, O., Paananen, T., 2007. **Changes in the expression of genetic characteristics across cohorts in skeletal deformations of farmed salmonids.** Genetics Selection Evolution 39, 529–543. doi:10.1051/gse:2007019

Koumoundouros, G., 2010. **Morpho-anatomical abnormalities in Mediterranean marine aquaculture.** Recent Advances in Aquaculture Research.

Kuzir, S., Kozaric', Z., Nejedli, S., 2004. **Development of mandibular arch in European sea bass, Dicentrarchus labrax (Linnaeus, 1758) from the "Cenmar" hatchery, Croatia.** Veter arc 74.

Mazurais, D., Glynatsi, N., Darias, M., Christodoulopoulou, S., Cahu, C., Zambonino-Infante, J.L., Koumoundouros, G., 2009. **Optimal levels of dietary vitamin A for reduced deformity incidence during development of European sea bass larvae (Dicentrarchus labrax) depend on malformation type.** Aquaculture 294. doi:10.1016/j.aquaculture.2009.06.008

Pauletto, M., Carraro, L., Babbucci, M., Lucchini, R., Bargelloni, L., Cardazzo, B., 2016. **Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.** Molecular Ecology Resources 16, 823–835. doi:10.1111/1755-0998.12495

Pecoraro, C., Babbucci, M., Villamor, A., Franch, R., Papetti, C., Leroy, B., Ortega-Garcia, S., Muir, J., Rooker, J., Arocha, F., Murua, H., Zudaire, I., Chassot, E., Bodin, N., Tinti, F., Bargelloni, L., Cariani, A., 2016. **Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (Thunnus albacares).** Marine Genomics 25, 43–48. doi:10.1016/j.margen.2015.12.002

Porcellotti, S., 2005. **Pesci d'Italia, ittiofauna delle acque dolci.** Plan, Firenze

Ritchie, B., Harrison, G., Harrison, L., Zantop, D., 1997. **Avian Medicine: Principles and Application.** Abridged Edition

Sfakianakis, D.G., Georgakopoulou, E., Papadakis, I.E., Divanach, P., Kentouri, M., Koumoundouros, G., 2006. **Environmental determinants of haemal lordosis in European sea bass, Dicentrarchus labrax (Linnaeus, 1758).** Aquaculture 254, 54–64. doi:10.1016/j.aquaculture.2005.10.028

Signer-Hasler, H., Neuditschko, M., Koch, C., Froidevaux, S., Flury, C., Burger, D., Leeb, T., Rieder, S., 2014. **A Chromosomal Region on ECA13 Is Associated with Maxillary Prognathism in Horses.** PLoS ONE 9, e86607. doi:10.1371/journal.pone.0086607

Suzuki, T., Oohara, I., Kurokawa, T., 1999. **Retinoic acid given at late embryonic stage depresses sonic hedgehog and Hoxd-4 expression in the pharyngeal area and induces skeletal malformation in flounder (Paralichthys olivaceus) embryos.** Development, Growth and Differentiation 41, 143–152. doi:10.1046/j.1440-169x.1999.00420.x

Teraoka, H., Dong, W., Okuhara, Y., Iwasa, H., Shindo, A., Hill, A.J., Kawakami, A., Hiraga, T., 2006. **Impairment of lower jaw growth in developing zebrafish exposed to 2,3,7,8-**

tetrachlorodibenzo-p-dioxin and reduced hedgehog expression. Aquatic Toxicology 78, 103–113. doi:10.1016/j.aquatox.2006.02.009

Wang, S., Meyer, E., McKay, J.K., Matz, M.V., 2012. 2b-RAD: **a simple and flexible method for genome-wide genotyping.** Nature Methods 9, 808–810. doi:10.1038/nmeth.2023

Weaver, D.E., 2006. **Design and operations of fine media fluidized bed biofilters for meeting oligotrophic water requirements.** Aquacultural Engineering, Design and Selection of Biological Filters for Freshwater and Marine Applications 34, 303–310. doi:10.1016/j.aquaeng.2005.07.004

Wolff, G., Wienker, T.F., Sander, H., 1993. **On the genetics of mandibular prognathism: analysis of large European noble families.** J Med Genet 30, 112–116.

Xue, F., Wong, R., Rabie, A.B.M., 2010. **Identification of SNP markers on 1p36 and association analysis of EPB41 with mandibular prognathism in a Chinese population.** Arch. Oral Biol. 55, 867–872. doi:10.1016/j.archoralbio.2010.07.018

Yamaguchi, T., Park, S.B., Narita, A., Maki, K., Inoue, I., 2005. **Genome-wide linkage analysis of mandibular prognathism in Korean and Japanese patients.** J. Dent. Res. 84, 255–259.

2. Publications

Carlo Pecoraro#, **Massimiliano Babbucci**#, Adriana Villamor, Rafaella Franch, Chiara Papetti, Bruno Leroy, Sofia O. Garcia, Jeff Muir, Jay Rooker, Freddy Arocha, Hilario Murua, Iker Zudairej, Emmanuel Chassot, Nathalie Bodin, Fausto Tinti, Luca Bargelloni, Alessia Cariani (2015).
Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*).
Marine Genomics, **25**. DOI: 10.1016/j.margen.2015.12.002.
#equally contributing authors

Marianna Pauletto, Lisa Carraro, **Massimiliano Babbucci**, Rosaria Lucchini, Luca Bargelloni, Barbara Cardazzo (2015).
Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Molecular Ecology Resources, **16**. DOI: 10.1111/1755-0998.12495.

Massimiliano Babbucci[1], Serena Ferraresso, Marianna Pauletto[1], Rafaella Franch, Chiara Papetti, Tomaso Patarnello, Paolo Carnier, Luca Bargelloni. (2016)
An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).
Scientific Reports, **8**. DOI: 10.1038/srep38673

Publication I

Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*).

Carlo Pecoraro#, **Massimiliano Babbucci#**, Adriana Villamor, Rafaella Franch, Chiara Papetti, Bruno Leroy, Sofia Ortega-Garcia, Jeff Muir, Jay Rooker, Freddy Arocha, Hilario Murua, Iker Zudairej, Emmanuel Chassot, Nathalie Bodin, Fausto Tinti, Luca Bargelloni, Alessia Cariani.

#equally contributing authors.

Abstract

Global population genetic structure of yellowfin tuna (*Thunnus albacares*) is still poorly understood despite its relevance for the tuna fishery industry. Low levels of genetic differentiation among oceans speak in favour of the existence of a single panmictic population worldwide of this highly migratory fish. However, recent studies indicated genetic structuring at a much smaller geographic scales than previously considered, pointing out that YFT population genetic structure has not been properly assessed so far. In this study, we demonstrated for the first time, the utility of 2b-RAD genotyping technique for investigating population genetic diversity and differentiation in high gene-flow species. Running *de novo* pipeline in *Stacks*, a total of 6,772 high-quality genome-wide SNPs were identified across Atlantic, Indian and Pacific population samples representing all major distribution areas. Preliminary analyses showed shallow but significant population structure among oceans (F_{ST} =0.0273; P-value < 0.01). Discriminant Analysis of Principal Components endorsed the presence of genetically discrete yellowfin tuna populations among three oceanic pools. Although such evidence needs to be corroborated by increasing sample size, these results showed the efficiency of this genotyping technique in assessing genetic divergence in a marine fish with high dispersal potential.

Introduction

Yellowfin tuna (*Thunnus albacares*, YFT) has relevant biological and economic importance at the global scale, being an apex predator in oceanic ecosystem and representing the second largest tuna fishery worldwide (FIGIS, 2010-2015). Currently, YFT is managed in four distinct stocks under the jurisdiction of four independent Regional Fisheries Management Organizations (RFMOs). Although a proper fish stock management needs accurate knowledge on the stock structure and its genetic variation with respect to environmental and ecological conditions (Papetti et al., 2013), YFT genetic population structure has not been resolved yet. Different studies provided discordant patterns of YFT global-scale genetic differentiation (Ward et al., 1997; Ely et al., 2005; Appleyard et al., 2001), together with a genetic structuring detected at the regional level (Dammanagoda et al., 2008; Kunal et al., 2013; Li et al., 2015). This discordance was likely due to the YFT life history traits (e.g. high fecundity, large population sizes), which make detecting patterns of genetic differentiation among population samples very difficult (Ely et al., 2005; Juan-Jordá et al., 2013). Moreover, population genetic studies reporting significant differences relied upon a relatively small number of molecular markers, hence, covering only a very limited portion of the genome (Appleyard et al., 2001; Díaz-Jaimes and Uribe-Alcocer, 2006). Failing to detect population structure, due to limited genetic resolution of classical markers, can potentially be misleading for management purposes, driving to local overfishing and severe stock decline (Ying et al., 2011).

According to the uncertainty about both population structure and size of YFT stocks, there is an evident need for developing alternative approaches based on genomics, that allow screening a larger number of markers across the entire genome, including neutral and non-neutral loci. This might enable detecting YFT population structure, quantifying the extent of spatial demographic changes and discover imprints of local adaptation, which represent priority focus for implementing any effective management plan.

The rapid advent of next-generation sequencing (NGS)-based genotyping methods has significantly improved our ability to analyse thousands of Single Nucleotide Polymorphism (SNP) markers across the entire genome, increasing the precision in detecting small genetic differentiation among geographical populations (Waples 2008; Allendorf et al., 2010; Davey et al., 2011; Narum et al., 2013; Andrews and Luikart 2014). Although SNPs are characterized by a low diversity due to the only four possible allelic states, this limitation is largely outweighed by their abundance, being as frequent as one SNP every few hundred base pairs (Morin et al., 2004; 2009). Moreover, SNPs are becoming the marker of choice for many applications in population ecology, evolution and conservation genetics, having a high potential for genotyping efficiency, data quality and low-scoring error rates, genome-wide coverage and analytical simplicity (Milano et al., 2014).

Here, for the first time, we applied the 2b-RAD Genotyping-By-Sequencing (GBS) technique (Wang et al., 2012) for testing its potential for investigating population genetic structure in a non-model, large pelagic and highly migratory fish species. This novel genomic tool is based on sequencing reduced representation libraries produced by type IIB restriction endonucleases, which cleave genomic DNA upstream and downstream of their target site, generating tags of uniform length that are ideally suited for sequencing on existing NGS platforms (Wang et al., 2012). This method permits parallel and multiplexed sample sequencing of tag libraries for the rapid discovery of thousands of SNPs across the entire individuals' genome, with a very cost-effective procedure resulting in high genome coverage. The 2b-RAD method allows to screen in parallel almost every restriction site in the genome, whereas other GBS methods can only target a subset of total restriction sites to counterbalance loss of PCR amplification and sequencing efficiency due to large size of restriction fragments. This technique also allows fine-tuning the marker density by means of selective adapters in order to sequence fewer loci with higher coverage, for applications such as population genetics (Puritz et al., 2014; Andrews and Luikart 2014). Given these attributes, the 2b-RAD method has the potential to discriminate the existence of genetic differentiation with a high statistical power, generating genome-wide data for genetic structure analysis at different spatial scales for YFT populations.

In this study, we: i) first examine the utility of Technical Replicates (TRs) for optimizing genotyping procedure, comparing the results obtained running the *denovo_map.pl* and the *ref_map.pl* programs in *Stacks* (Catchen et al., 2011, 2013); and ii) finally assess the applicability of 2b-RAD for future investigations in this highly migratory species.

Results and discussion

A similar number of reads was obtained among TRs, before and after quality filtering (Table 1), which underlines the reliability of this technique in genotyping individuals. Among the different *Stacks* settings considered, the *-m* value was the parameter that most affected the genotyping results, in particular the number of detected SNPs. Sensitivity tests performed on the TRs showed a decrease in the number of SNPs, from 5,753 to 4,490, when increasing *-m* from 5 to 15 (Fig. 1 and Supplementary Material 1 for values with associated Standard Error). The percentage of error rate varied approximately from 1% to 5%, with a decreasing trend when increasing the *–m* values (Fig. 1 and Supplementary Material 1). The percentage of heterozygous SNPs remained constant with increasing *-m* values (Fig. 1 and Supplementary Material 1)

.

Sample ID	Oceanic origin	gDNA ng/µL	Library nm/µL	N° raw reads	N° filtered reads	% retained reads
34_2_Y_2R1	Atlantic Ocean	333.80	185.38	2,276,239	1,772,927	78 %
34_2_Y_2R2	Atlantic Ocean	333.80	207.83	2,672,917	1,914,181	72 %
34_2_Y_2R3	Atlantic Ocean	333.80	197.03	2,309,039	1,805,181	78 %
77_2_Y_15R1	Pacific Ocean	218.02	238.81	2,212,559	1,788,988	81 %
77_2_Y_15R2	Pacific Ocean	218.02	240.53	2,292,834	1,850,871	81 %
77_2_Y_15R3	Pacific Ocean	218.02	208.91	2,085,658	1,802,820	86 %
51_1_Y_7R1	Indian Ocean	177.54	166.05	2,330,292	1,767,345	76 %
51_1_Y_7R2	Indian Ocean	177.54	170.81	2,300,342	1,824,635	79%

Table 1. Details on the technical replicates: acronym (Sample ID), Oceanic origin, genomic DNA concentration in ng/µL, library concentration in nm/µL, number of raw reads obtained, retained reads after quality filtering, and their corresponding percentage

.

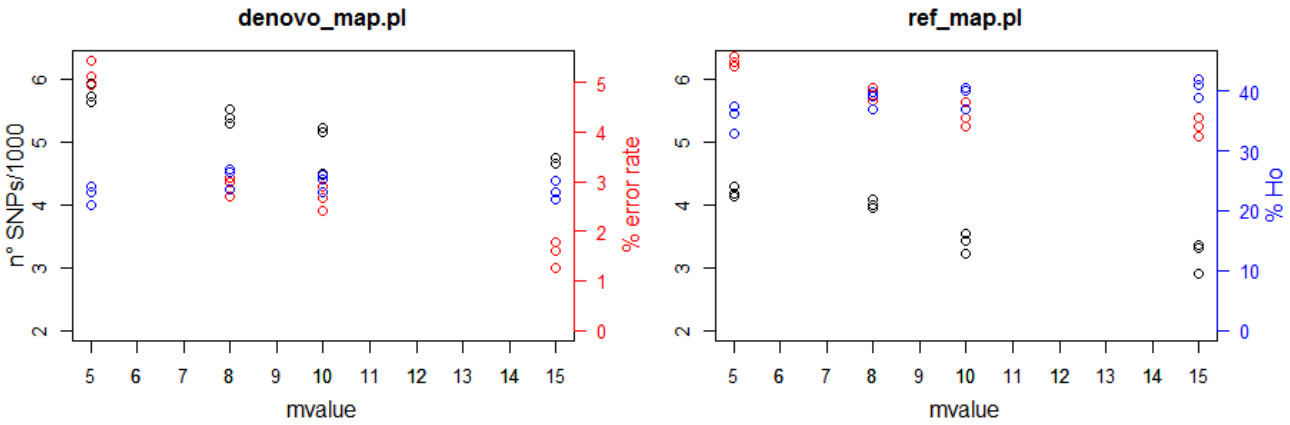


Fig. 1. Comparison between d*enovo_map.pl* (Left panel) and ref_map.pl (Right panel) performance in terms of (a) number of SNPs (black dots) (b) error rate (red dots) and (c) percentage of heterozygous loci (blue dots), using different -m values. Each dot represents the average value among the three individual's TRs. The three y axes (n° SNPs/1000, % error rate, % Ho) are shared between the two plots.

An increase in true heterozygous SNPs calls was observed using the *bounded SNP calling model* compared to the *default SNP model* and reducing the upper bound values, in agreement with the results obtained by Mastretta-Yanes et al. (2014). In fact, reducing the upper bound on the maximum-likelihood of ε decreases the possibility of calling a homozygote instead of a true heterozygous genotype (Catchen et al. 2013). The proper genotype calling was further checked for a sub-sample of the total reads obtained, in the *Stacks* web interface, verifying the sequences alignment and monitoring the genotyping inference when the results were exported. This procedure was repeated each time when changing the different model's *upper bound* values.

By relaxing the number of mismatches within each locus (*-n*) and among loci (*-M*), an increase in the number of SNPs and error rate was observed (Supplementary Material 2).

Mapping 2b-RAD reads against the genome of *T. orientalis* allowed a high percentage of successfully mapped sequences (86.59%). The outputs obtained on the mapped data from TRs with the *ref_map.pl* program, confirmed the trends observed with the *denovo_map.pl* program (Fig. 1). However, the absolute number of SNPs was lower than that obtained with the *denovo_map.pl* program, likely due to the incompleteness of the reference genome used (the only *Thunnus sp.* genome available to date, Nakamura et al., 2013) and the phylogenetic distance between YFT and *Thunnus orientalis*.

Aligning reads to the reference genome, before calling a locus, can filter out erroneous stacks generated by contaminants (e.g. bacteria) possibly present in very small amount in the starting gDNA sample. Moreover the error rate also showed a less evident decreasing pattern when increasing *-m*, confirming however a low error rate in the genotyping call (<5%). On the contrary, the percentage of

heterozygous SNPs identified using *T. orientalis* genome as reference, showed a slight increase from 35.6% to 40.7%, when higher values of *m* where used (Fig.1).

Based on the number of SNPs identified with the two different approaches (with or without using the reference genome), the low error rates and the consistent percentage of heterozygous SNPs obtained among TRs, the *denovo_map.pl* program was run applying the following parameter settings *-m*= 8, *-M* = 3, *-n* = 2 and a bounded *SNP calling model* with an upper bound of 0.1 (all remaining *Stacks* settings as default), to obtain the final dataset (see Supplementary Material 3 for details about each individual). The AMOVA results (Supplementary Material 4) revealed that pooling samples into the three major oceanic regions produced the highest percentage of variation explained by groups subdivision (2.73% P-value <0.001) and at the same time very low and not significant differences were observed among populations within groups (0.97% P-value >0.01). Pooling YFT individuals in these three groups corresponding to the three oceans (Table 2), allowed to increase the sample size and to obtain more robust and reliable inferences of population structure.

Location	**N° individuals**	**Raw reads (mln)**	**Filtered reads (mln)**	**% of reads retained**	**Unique tags**	**Polymorphic SNPs**	**SNPs found**
Atlantic Ocean	40	3.80 (± 0.29)	3.09 (± 0.37)	80%	30,776	3,264	5,693
Indian Ocean	20	3.98 (± 0.46)	2.91 (± 0.15)	79%	31,430	3,695	6,516
Pacific Ocean	40	3.30 (± 0.26)	2.78 (± 0.24)	84%	31,573	3,363	5,906

Table 2. Summary statistics of the three *Thunnus albacares* oceanic groups. The table reports: the sampling origin (location) the sample size (N° individuals), the mean number (millions) of raw reads with the associated standard error (SE), the corresponding mean number (millions) of filtered reads (with SE), the percentage of reads retained, the mean value of unique tags, polymorphic SNPs and number of SNPs found.

The optimization process combining different *-r* and *-p* parameters' values of the *Stacks* population module, led to *-r = 0.7* and *p* = 6 as best middle way between the number of SNPs and the percentage of missing values obtained (Table 3).

		-p					
		4		6		9	
		SNPs	NA %	SNPs	NA %	SNPs	NA %
-r	0.4	8,158	14.3	7,871	11.49	7,560	9.33
	0.7	7,049	9.33	6,772	6.3	6,430	4.69
	0.9	5,981	8.62	5,673	5.44	5,187	2.58

Table 3. Number of SNPs and percentage of missing value (NA %) obtained for the entire dataset according to the *-r* and *-p* parameters' values of the *Stacks* population program

This set of parameters produced a panel of 6,772 SNPs. Pairwise F_{st} distances, calculated with this dataset, were highly significant, suggesting genetic differences occurring among oceanic groups (Tab. 4).

	Atlantic	**Indian**	**Pacific**
Atlantic	*	**<0.01**	**<0.01**
Indian	0.04736	*	**<0.01**
Pacific	0.02932	0.01714	*

Table 4._Pairwise F_{st} values calculated among geographic pools (from Atlantic, Indian and Pacific Ocean) of yellowfin tuna are reported (below diagonal) with their associated P-values (below diagonal). Significant values after Bonferroni standard correction are in bold (nominal significant threshold $\alpha = 0.01$).

The DAPC confirmed the genetic differentiation among oceanic basins. The graph of the Bayesian Information Criterion (BIC) values for increasing values of the number of clusters (k) showed that k=3 corresponded to the lowest associated BIC value. In the data transformation step for PCA analysis, 30 principal components (PCs) were retained, accounting for approximately the 92% of the total genetic variability. The eigenvalues of the DAPC indicated that the first two components explained most of the variation. The resulting scatterplot (Fig. 2) showed three genetic clusters corresponding to Atlantic, Indian and Pacific YFT groups. Moreover, the cross-validation of DAPC performed on our dataset, indicated that the number of PCs associated to the highest mean success and lowest mean squared error corresponded to 35 PCs. This result supported our choice to retain 30 PCs during the dimension-reduction step of the DAPC.

These results are in agreement with previously observed signatures of genetic heterogeneity among oceans found by Ward et al. (1997) by means of significant allele frequency differences at the locus

GPI-A (PGI-F). Although, this scenario necessarily needs to be confirmed by increasing the sample size, it validated 2b-RAD genotyping technique as a powerful tool to assess YFT genetic structure and diversity at the global scale

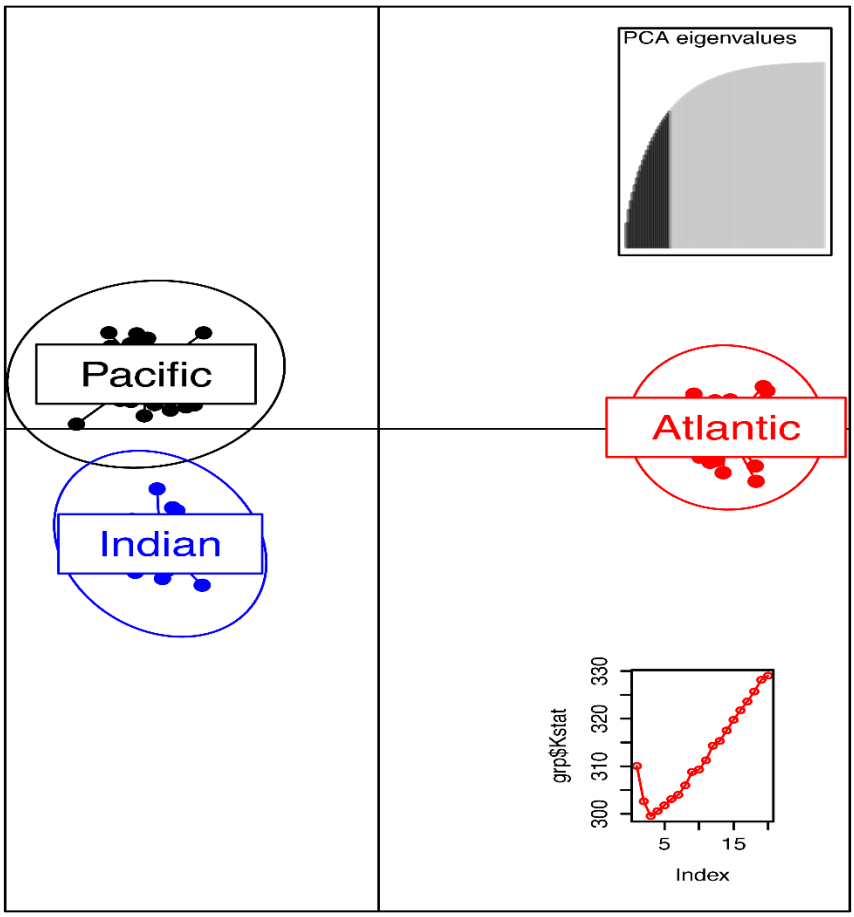


Fig. 2. Scatterplot of the DAPC results identifying three genetic clusters of *Thunnus albacares.*

Conclusions

This methodological study confirmed that TRs are useful for optimizing genotyping procedure and that they are crucial to reduce the amount of statistical error introduced in allele frequency estimation due to PCR artefacts. We unambiguously mapped the TRs' tags against the reference genome of *T. orientalis* with a high percentage of success (86,59%), in spite of the small size of fragments (Puritz et al., 2014), and the evolutionary distance between these two species. The methodological approach showed that the lack of a reference genome, although undesirable, does not evidently compromise the reproducibility and accuracy of the data obtained, underlying the consistence of the technique in genotyping individuals. We preliminarily demonstrated that 2b-RAD is a promising tool to screen a

large set of genomic loci in a marine high gene-flow species, underlying the inter-oceanic population genetic differentiation. Certainly, an increased sample size is needed to address estimates of genetic differentiation among YFT population samples also at a smaller local geographic scale.

Materials and methods

Sampling design, libraries preparation and sequencing

A total of 100 juvenile YFT (35-55 cm of fork length, FL) from Atlantic, Indian and Pacific geographic population samples (Table 5) were analysed, covering the entire species distribution (Fig. 3).

Sampling location	Sample code	Number of individuals
WC Atlantic Ocean	**31_1**	10
EC Atlantic Ocean	**34_1**	10
EC Atlantic Ocean	**34_2**	10
WC Atlantic Ocean	**41_1**	10
WC Indian Ocean	**51_1**	10
WC Indian Ocean	**51_2**	10
WC Pacific Ocean	**71_1**	10
WC Pacific Ocean	**71_2**	10
EC Pacific Ocean	**77_1**	10
EC Pacific Ocean	**77_2**	10

Table 5. The table summarizes the sampling location, sample code and number of individual per each geographic population sample.

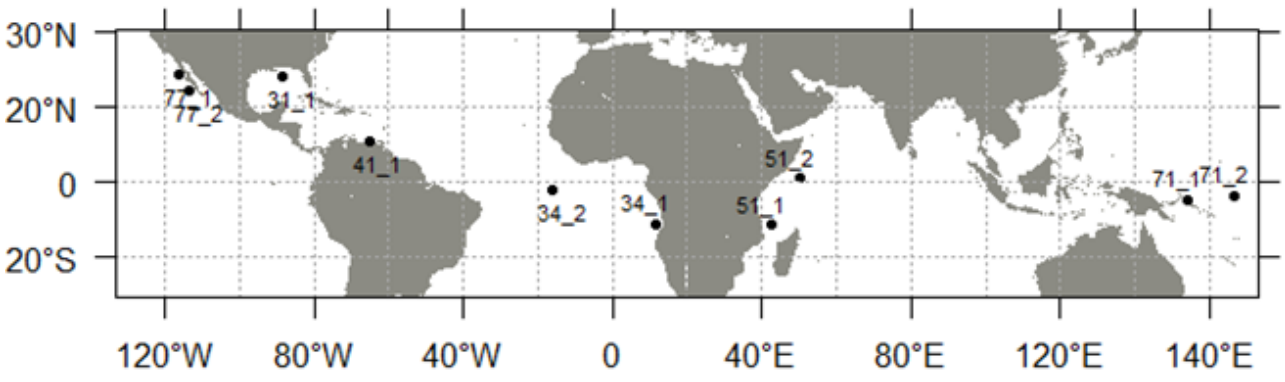


Fig. 3. Location of *Thunnus albacares* geographical population samples analyzed in this study. Sample codes are given as in Table 5.

Genomic DNA (gDNA) was extracted from approximately 20 mg of tissue (skeletal muscle or finclip) using the commercial kit Invisorb® Spin Tissue Mini Kit (Invitek, STRATEC Biomedical, Germany) following the manufacturers' recommendations. Since high-quality gDNA is required in the 2b-RAD genotyping technique, its concentration and purity, in terms of ratios of absorbance at 260/230 nm and at 260/280 nm, were quantified by both a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and a Qubit 2.0 Fluorometer (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA). This procedure ensured to work with high quality samples and comparable DNA concentration.

The 2b-RAD libraries were constructed for each individual following the protocol from Wang et al. (2012) with minor modifications (see below). To assess the robustness of the method and subsequent data analyses, three libraries were replicated (Technical Replicates, TRs) for two individuals (34_2_Y_2 and 77_2_Y_15) and two for an additional third specimen (51_1_Y_7). gDNA (300 ng) was digested with 2 U of the enzyme *CspC*I (New England Biolabs, NEB, Ipswich, Massachusetts, USA) for 1 h at 37°C. The digested DNA was ligated in a 25µL total volume reaction consisting of 0.4 µM for each of the two library-specific adaptors, 0.2 mM ATP (New England Biolabs, NEB, Ipswich, Massachusetts, USA) and 1 U T4 DNA ligase (SibEnzyme Ltd., Academ town, Siberia). To reduce marker density, one adaptor with fully degenerate 3' overhangs NN and one with reduced 3'degeneracy NG were chosen. Sample-specific barcodes were designed with Barcode Generator (http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator) and introduced by PCR with platform-specific barcode-bearing primers. 2b-RAD tags were amplified by PCR in two separate 25 µL-reactions, in order to minimize PCR amplification bias (Mastretta-Yanes et al., 2014). Each amplification consisted of 6.25 µL of ligated DNA, 0.5 µM each primer (P4 and P6-BC, Eurofins Genomics S.r.l, Italy), 0.2 µM each primer (P5 and P7, Eurofins Genomics), 0.3 mMdNTP (New England Biolabs, NEB, Ipswich, Massachusetts, USA), 1X Phusion HF buffer and 1 U TaqPhusion high-fidelity DNA polymerase (NEB). Cycling conditions were: 98°C for 4min; 98 °C for 5 s, 60° C for 20 s, 72° C for 5 s for 14 cycles, 72°C for 5 min. The reduced number of amplification cycles (n=14) is crucial to produce a negligible amount of PCR amplification errors, comparing to those needed to reach the plateau phase.

PCR products were purified with the SPRIselect purification kit (Beckman Coulter, Pasadena, California, USA), to exclude any high-molecular weight DNA remaining after the enzyme digestion and any incorrect constructs that may emerge during PCR amplification. The concentration of purified individual libraries was quantified using Qubit®ds DNA BR Assay Kit (Invitrogen–ThermoFisher Scientific, MA, USA) and Mx3000P qPCR instrument, and the quality checked on an

Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). Individual libraries were pooled into equimolar amounts and resulting pools' quality was re-verified on Agilent 2100 Bioanalyzer. Pooled libraries were sequenced on an Illumina HiSeq2500 platform with a 50 bp single-read module at the Genomix4Life S.r.l. facilities (Baronissi, Salerno, Italy), which also performed data demultiplexing.

Technical Replicates analysis and optimization of genotyping procedure

Demultiplexed reads were returned by the sequencing facility in Fastq format and their quality was checked by FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). After this, a custom-made Perl script was run for quality filtering and adaptors trimming of the reads, obtaining sequences of 34bp (Fasta files available at SRA Bioproject: PRJNA294940). Filtered reads were analyzed with the software *Stacks* v. 1.32 (Catchen et al., 2011, 2013), which allows genotype inference through the identification of SNP loci without a reference genome (*denovo_map.pl* program) or aligning reads against a reference genome (*ref_map.pl* program). Different settings were tested on the TRs dataset to fine-tune the *de novo Stacks* pipeline' parameters and to assess the consistency of results, in terms of total number of identified SNPs; the error rate calculated counting discordant genotypes between TRs and, among the concordant data, the percentage of heterozygous SNPs. Following the *Stacks* author guidelines, multiple combinations were considered for: a) the minimum number of identical reads necessary to call an allele (*-m* value set to: 5, 8,10, 15); b) mismatches between reads within a locus (*-M* value set to: 2, 3, 4, 6); and c) mismatches among loci when comparing across individuals (*-n* value set to: 0, 2, 3, 4, 6). Only one parameter was varied at a time while keeping the others fixed. The default values for *min_het_seqs* and *max_het_seqs* were used in order to check that the genotype was correctly called. If the ratio of the depth of the minor allele to the major allele is bigger than *max_het_seqs* (default of 1/10) a SNP is called as heterozygote. Otherwise, if it is smaller than *min_het_seqs* (default of 1/20) a SNP was called homozygous. If the ratio is in between the two values the genotype is not assigned. In addition, we compared the default and the bounded SNP calling models (*--bound_high* value set to: 0.05, 0.1, 0.15, 0.2, 0. 5) to evaluate the percentage of heterozygous genotypes correctly assessed, in order to make the genotype calling between them as much concordant as possible. In the bounded SNP calling model, *Stacks* employs a multinomial-based likelihood model to identify SNPs and to estimate the maximum-likelihood value of the sequencing error rate ε at each nucleotide position, in order to proper call each possible genotype (for details see Catchen et al. 2011 and Catchen et al. 2013).

The reads were also mapped against the genome of *Thunnus orientalis* (GenBank accession numbers BADN01000001-BADN01133062; Nakamura et al., 2013) using CLC Genomics Workbench v. 5.1

(CLC Bio) program. The following parameters settings were applied *length fraction*= 1.0 and *similarity fraction*= 0.9 (all remaining parameters as default), retaining only uniquely mapped reads. Mapping results were exported in SAM format and were used as input files for *refmap_map.pl* in *Stacks*. To further evaluate the robustness of the approach a similar testing was performed on mapped data, using the same settings as for the *denovo_map.pl* for *-m*, *-n* and *--bound_high* of the bounded SNP calling model (Fig. 2).

Preliminary analysis of YFT population structure

Once identified the *Stacks*' parameter set which minimized differences among TRs, the *denovo_map.pl* program was run on the entire YFT dataset (Table 2). Using the program *populations* available from *Stacks* software, different combinations of *-p* (4, 6, 9) and *-r* (0.4, 0.7, 0.9) parameters were tested, in order to investigate changes in the number of SNPs obtained, and in the percentage of missing values among samples. Following these tests, we selected from the resulting catalogue of loci only those containing one bi-allelic SNP (*-F snps_l*=1 *snps_u*=2), and those values of *-p* and *-r* rendering the highest number of SNPs with the lowest percentage of missing data.

In order to increase the sample size and to improve robustness of the genetic analyses, several grouping of the geographic samples were tested, especially due to their ocean basin distance, performing an *analysis* of molecular variance (*AMOVA*) with the software Arlequin 3.5.1.2 (Excoffier and Lischer, 2010) with 10,000 permutations and $p \leq 0.01$ significance level.

Based on the SNPs dataset and AMOVA results obtained, F_{ST} estimates for pairwise comparison among pooled samples, were calculated with the software Arlequin 3.5.1.2 using the same settings as above.

A preliminary assessment of YFT genetic structure was performed using the Discriminant Analysis of Principal Components (DAPC, Jombart et al., 2010) implemented in the R package Adegenet (Jombart, 2008, R version 3.1.2, R Development Core Team, 2014; http://www.r-project.org). The function *find.clusters was used to identify the optimal number of clusters (k) that maximizes the variation between groups* (Jombart et al., 2010). The cross-validation test was also carried out in order to validate the number of Principal Components (PCs) retained in the first transformation step of DAPC analysis, because a wrong choice of the number of PCs might negatively impact the DAPC results and produce unstable output due to over-parameterization.

References

Appleyard S., Grewe P., Innes B., and Ward R. "Population Structure of Yellowfin Tuna (*Thunnus albacares*) in the Western Pacific Ocean, inferred from microsatellite loci." Marine Biology 139, (2001): 383-93. DOI: 10.1007/s002270100578

Allendorf F. W., Hohenlohe P. A., and Gordon L. "Genomics and the future of conservation genetics." Nature Reviews Genetics 11, (2010): 697–709. DOI: 10.1038/nrg2844

Andrews K. R. and Luikart G. "Recent novel approaches for population genomics data analysis." Molecular Ecology 23, (2014),: 1661–1667. DOI:10.1111/mec.12686

Catchen J., Hohenlohe P. A, Bassham S., Amores A., and Cresko W. A. "Stacks: an analysis tool set for population genomics." Molecular Ecology 22, (2013): 3124–40. DOI: 10.1111/mec.

Catchen J. M., Amores A., Hohenlohe P., Cresko W., and Postlethwait J. H. "Stacks: building and genotyping loci de novo from short-read sequences." G3: Genes, Genomes, Genetics 1, (2011): 171–82.DOI: 10.1534/g3.111.000240

Chow S. and Inoue S. "Intra-and interspecific restriction fragment length polymorphism in mitochondrial genes of *Thunnus tuna* species." Bulletin National Research Institute of Far Seas Fisheries, 30, (1993): 207-225.

Dammannagoda S. T., Hurwood D. A., and Mather P. B. "Evidence for fine geographical scale heterogeneity in gene frequencies in yellowfin tuna (*Thunnus albacares*) from the north Indian Ocean around Sri Lanka." Fisheries Research 90, (2008): 147–57.DOI:10.1016/j.fishres

Davey J. W., , Hohenlohe P. A., Etter P. D., Boone J. Q., Catchen J. M., and Blaxter. M. L. "Genome-wide genetic marker discovery and genotyping using next-generation sequencing." Nature Reviews Genetics 12, (2011): 499–510. DOI:10.1038/nrg3012

Díaz-Jaimes P. and Uribe-Alcocer M. "Spatial differentiation in the eastern Pacific yellowfin tuna revealed by microsatellite variation." Fisheries Science 72, (2006): 590–96. DOI:10.1111/j.1444-2906.2006.01188.x.

Ely B., Viñas J., Alvarado Bremer J. R., Black D., Lucas L., Covello K., Labrie A. V., and Thelen E. "Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*)." BMC Evolutionary Biology 5, (2005): 19.DOI: 10.1186/1471-2148-5-19

Excoffier L., and Lischer H. E. “Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.” Molecular ecology resources, 10,(2010): 564-567.

FAO 2010-2015. Fisheries Global Information System (FAO-FIGIS) - Web site. Fisheries Global Information System (FIGIS). FI Institutional Websites. In: FAO Fisheries and Aquaculture Department [online]. Rome. Updated . http://www.fao.org/fishery/figis/en

Jombart T., Devillard S., Dufour A. B., and Pontier D. “Revealing cryptic spatial patterns in genetic variability by a new multivariate method.” Heredity 101, (2008): 92–103. DOI: 10.1038/hdy.2008.34

Juan-Jordá M. J., Mosqueira I., Freire J., and Dulvy N. K. “Life in 3-d: life history strategies in tunas, mackerels and bonitos.” Reviews in Fish Biology and Fisheries 23, (2013): 135–55. DOI:10.1007/s11160-012-9284-4.

Kunal S. P., Kumar G., Menezes M. R., and Meena R. M. “Mitochondrial DNA analysis reveals three stocks of yellowfin tuna *Thunnus Albacares* (Bonnaterre, 1788) in Indian Waters.” Conservation Genetics 14, (2013): 205–13. DOI: 10.1007/s10592-013-0445-3

Li W., Xinjun C., Qianghua X., Jiangfeng Z., Xiaojie D., and Liuxiong X. “Genetic population structure of *Thunnus Albacares* in the Central Pacific Ocean based on mtDNA COI Gene Sequences.” Biochemical Genetics 53, (2015): 8–22. DOI:10.1007/s10528-015-9666-0.

Mastretta-Yanes A., Arrigo N., Alvarez N., Jorgensen T. H., Piñero D. and. Emerson B. C. “Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference.” Molecular Ecology Resources 15, (2015): 28–41. DOI:10.1111/1755-0998.12291.

Milano I., Babbucci M., Cariani A., Atanassova M., Bekkevold D., Carvalho G. R., Espiñeira M., Fiorentino F., Garofalo G., Geffen A. J., Hansen Jakob. H., Helyar S. J., Nielsen E. E., Ogden, R., Patarnello T., Stagioni M., FishPopTrace Consortium, Tinti F. and Bargelloni L. “Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*).” Molecular Ecology 23, (2014): 118–135. DOI:10.1111/mec.12568

Morin P. A., Luikart G., Wayne R. K. and the SNP workshop Group. “SNPs in ecology, evolution and conservation.” *Trends in Ecology & Evolution* 19, (2004): 208–16. DOI:10.1016/j.tree.2004.01.009

Morin P. A., Martien K. K., and Taylor L. B. “Assessing statistical power of SNPs for population structure and conservation studies.” Molecular Ecology Resources 9, (2009): 66–73.DOI: 10.1111/j.1755-0998.2008.02392.x

Nakamura Y., Mori K., Saitoh K., et al. ″Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna.″ Proc Natl Acad Sci USA 110 , (2013): 11061–6.DOI: 10.1073/pnas.1302051110

Narum S. R., Buerkle C. A., Davey J. W., Miller M. R., and Hohenlohe P. A. “Genotyping-by-Sequencing in ecological and conservation genomics.” Molecular Ecology 22, (2013): 2841–47. DOI: 10.1111/mec.12350

Pedrosa-Gerasmio I.R., Babaran R.P., and Santos M.D. ″Discrimination of juvenile yellowfin (*Thunnus albacares*) and bigeye (*Thunnus obesus*) tunas using mitochondrial DNA control region and liver morphology.″PLoS ONE 7, (2012): e35604. DOI: 10.1371/journal.pone.0035604.

Papetti C., Di Franco A., Zane L., Guidetti P., De Simone V., Spizzotin M., Zorica B., Kec V. C., Mazzoldi C. “Single population and common natal origin for Adriatic *Scomber scombrus* stocks: evidence from an integrated approach.” ICES Journal of Marine Science: Journal Du Conseil, 70, (2013): 387 –398. DOI: 10.1093/icesjms/fss201.

Puritz J. B., Matz M. V., Toonen R. J., Weber J. N., Bolnick D. I., and Bird. C. E. “Demystifying the RAD Fad.” Molecular Ecology 23, (2014): 5937–42. DOI: 10.1111/mec.12965

Ward R. D., Elliot N. G., Innes B. H., Smolenski A. J., and Grewe P. M. ″Global population structure of yellowfin tuna, *Thunnus albacares,* inferred from allozyme and mitochondrial DNA variation”. ″Fishery Bulletin 95, (1997): 566-575. DOI: 10.1007/bf00347499

Wang S., Meyer E., McKay J. K., and Matz M. V. “2b-RAD: a simple and flexible method for genome-wide genotyping.” Nature Methods 9, (2012): 808–10. DOI: 10.1038/nmeth.2023.

Waples R. S., Punt A. E., and Cope J. M. “Integrating genetic data into management of marine resources: how can we do it better?”. Fish and Fisheries 9, (2008): 423–49. DOI: 10.1111/j.1467-2979.2008.00303.x

Ying Y., Chen Y., Lin L., Gao T., and Quinn T. “Risks of ignoring fish population spatial structure in fisheries management.” Canadian Journal of Fisheries and Aquatic Sciences 68, (2011): 2101–20. DOI:10.1139/F2011-116.

Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*)-SUPPLEMENTARY MATERIAL 1-

Supplementary Material 1- Comparison between *denovo_map.pl* (Left table) and *ref_map.pl* (Right table) performance in terms of: (1) number of SNPs obtained (SNPs/1000) with the associated standard error (SE); (2) error rate (SE); and (3) percentage of heterozygous loci (SE), using different *m* values (5, 8, 10, 15)

denovo_map.pl				
ID Sample	mvalue	SNPs/1000	%error rate	%Ho
34_2_Y_2	15	4.08(±0.19)	1.51(±0.39)	22.13(±0.37)
	10	4.50(±0.14)	2.90(±0.15)	23.12(±0.15)
	8	5.27(±0.26)	3.08(±0.26)	23.70(±0.18)
	5	5.92(±0.13)	5.43(±0.29)	21.12(±0.27)
77_2_Y_15	15	4.65(±0.32)	1.27(±0.21)	25.23(±0.24)
	10	5.13(±0.15)	2.42(±0.13)	26.10(±0.11)
	8	5.37(±0.14)	2.71(±0.32)	27.00(±0.34)
	5	5.63(±0.29)	4.95(±0.24)	24.61(±0.31)
51_1_Y_7	15	4.74(±0.27)	1.77(±0.15)	23.23(±0.23)
	10	5.22(±0.29)	2.68(±0.27)	25.31(±0.14)
	8	5.52(±0.12)	2.99(±0.23)	26.42(±0.07)
	5	5.71(±0.09)	5.13(±0.17)	23.21(±0.12)

ref_map.pl				
ID Sample	mvalue	SNPs/1000	%error rate	%Ho
34_2_Y_2	15	2.89(±0.26)	3.91(±0.19)	39.00(±0.33)
	10	3.21(±0.23)	4.11(±0.37)	37.11(±0.41)
	8	4.08(±0.14)	4.63(±0.25)	36.98(±0.25)
	5	4.18(±0.16)	5.53(±0.21)	33.10(±0.24)
77_2_Y_15	15	3.35(±0.08)	4.11(±0.15)	41.00(±0.15)
	10	3.54(±0.13)	4.29(±0.08)	40.00(±0.18)
	8	3.99(±0.25)	4.89(±0.16)	39.80(±0.23)
	5	4.27(±0.28)	5.32(±0.27)	37.40(±0.16)
51_1_Y_7	15	3.30(±0.15)	4.28(±0.14)	42.11(±0.09)
	10	3.42(±0.27)	4.61(±0.18)	40.50(±0.22)
	8	3.94(±0.33)	4.76(±0.23)	39.21(±0.12)
	5	4.13(±0.28)	5.41(±0.24)	36.32(±0.11)

Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*)-SUPPLEMENTARY MATERIAL 2-

Supplementary Material 2- The effects of different *M* values (2, 3, 4, 6; Left Table); and *n* values (0, 2, 3, 4, 6; Right Table) on the: 1) number of SNPs obtained (SNPs/1000); (2) percentage of error rate; and (3) percentage of heterozygous loci).

ID Sample	-M value	SNPs/1000	%error rate	%Ho
34_2_Y_2	2	5.02 (±0.34)	3.02 (±0.43)	21.32 (±0.25)
	3	5.27 (±0.26)	3.08 (±0.26)	23.70 (±0.18)
	4	5.35 (±0.33)	3.43 (±0.47)	23.72 (±0.36)
	6	5.37 (±0.16)	3.57 (±0.21)	22.98 (±0.13)
77_2_Y_15	2	5.24 (±0.21)	2.14 (±0.12)	24.45 (±0.43)
	3	5.37 (±0.14)	2.71 (±0.32)	27.00 (±0.34)
	4	5.43 (±0.37)	2.91 (±0.42)	27.02 (±0.21)
	6	5.50 (±0.26)	3.21 (±0.31)	27.12 (±0.38)
51_1_Y_7	2	5.23 (±0.43)	2.97 (±0.39)	26.31 (±0.45)
	3	5.52 (±0.12)	2.99 (±0.23)	26.42 (±0.07)
	4	5.56 (±0.25)	3.15 (±0.42)	27.02 (±0.13)
	6	5.86 (±0.48)	3.65 (±0.37)	27.31 (±0.57)

ID Sample	-n value	SNPs/1000	%error rate	%Ho
34_2_Y_2	0	5.19 (±0.13)	3.04 (±0.27)	22.63 (±0.31)
	2	5.27 (±0.26)	3.08 (±0.26)	23.70 (±0.18)
	3	5.31 (±0.46)	3.25 (±0.18)	23.82 (±0.34)
	4	5.33 (±0.28)	3.29 (±0.32)	23.84 (±0.21)
	6	5.49 (±0.13)	4.01 (±0.34)	23.96 (±0.27)
77_2_Y_15	0	5.24 (±0.21)	2.58 (±0.27)	26.98 (±0.15)
	2	5.37 (±0.14)	2.71 (±0.32)	27.00 (±0.34)
	3	5.42 (±0.17)	2.92 (±0.28)	27.04 (±0.16)
	4	5.51 (±0.37)	3.18 (±0.19)	27.15 (±0.07)
	6	5.59 (±0.45)	3.31 (±0.09)	27.22 (±0.22)
51_1_Y_7	0	5.41 (±0.04)	2.93 (±0.18)	25.97 (±0.39)
	2	5.52 (±0.12)	2.99 (±0.23)	26.42 (±0.07)
	3	5.54 (±0.45)	3.11 (±0.03)	26.45 (±0.13)
	4	5.63 (±0.45)	3.23 (±0.15)	26.48 (±0.26)
	6	5.67 (±0.31)	3.24 (±0.27)	26.68 (±0.19)

Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*)-SUPPLEMENTARY MATERIAL 3-

Supplementary Material 3- Summary statistics of each *Thunnus albacares* individual. The table reports: the sampling origin (Ocean), the mean number (millions) of raw reads, the corresponding mean number (millions) of filtered reads, the percentage of reads retained, the mean value of unique tags, polymorphic SNPs and the number of SNPs found.

ID Sample	Ocean	Raw reads	Filtered reads	Unique tags	Polymorphic loci	SNPs found
T_31_1_Y_12	Atlantic	3105528	2563455	30838	3235	5699
T_31_1_Y_13	Atlantic	5637152	4557441	33788	3402	5824
T_31_1_Y_14	Atlantic	2644769	2285761	29602	3007	5260
T_31_1_Y_15	Atlantic	3288097	2451980	30016	3107	5365
T_31_1_Y_16	Atlantic	6336440	5450801	34926	3611	6249
T_31_1_Y_23	Atlantic	3654650	3263003	31607	3314	5712
T_31_1_Y_35	Atlantic	2721757	2521783	29673	3205	5565
T_31_1_Y_39	Atlantic	2740004	2261642	29808	3215	5694
T_31_1_Y_43	Atlantic	2603891	1676058	27211	2905	5223
T_31_1_Y_34	Atlantic	2122040	985219	22810	3024	5248
T_34_1_Y_13	Atlantic	3131345	2517429	31023	3344	5891
T_34_1_Y_15	Atlantic	3347614	2590822	31639	3333	5867
T_34_1_Y_16	Atlantic	3019207	2308602	30701	3338	5940
T_34_1_Y_17	Atlantic	4386713	3073122	33114	3537	6198
T_34_1_Y_28	Atlantic	6107652	5081470	37117	3964	6819
T_34_1_Y_29	Atlantic	4130588	3147206	33245	3602	6162
T_34_1_Y_3	Atlantic	2846057	2056458	29254	3092	5644
T_34_1_Y_31	Atlantic	4025468	3185735	34137	3617	6281
T_34_1_Y_35	Atlantic	2717490	2111695	30702	3255	5802
T_34_1_Y_37	Atlantic	3145853	2203273	29257	2927	5185

T_34_2_Y_10	Atlantic	4650572	3494452	15529	1347	2493
T_34_2_Y_11	Atlantic	4032665	3227973	32985	3605	6243
T_34_2_Y_14	Atlantic	3029896	2359699	20612	1991	3342
T_34_2_Y_16	Atlantic	4998909	4301682	35583	3732	6550
T_34_2_Y_17	Atlantic	4023014	3294257	32566	3449	5896
T_34_2_Y_20	Atlantic	2376762	1948145	29748	3208	5647
T_34_2_Y_21	Atlantic	4679065	4050647	35695	3752	6534
T_34_2_Y_23	Atlantic	4300413	3598795	34985	3663	6322
T_34_2_Y_27	Atlantic	3494322	2747041	31605	3358	5864
T_34_2_Y_38	Atlantic	2469862	1839548	29513	3118	5454
T_41_1_Y_13	Atlantic	2885225	2229266	29375	3148	5557
T_41_1_Y_14	Atlantic	4112269	3887296	34064	3574	6042
T_41_1_Y_18	Atlantic	3845104	3104426	32644	3423	5868
T_41_1_Y_19	Atlantic	3750752	3150911	32874	3411	5912
T_41_1_Y_20	Atlantic	7135055	6366517	18207	2407	4535
T_41_1_Y_27	Atlantic	5182575	4736811	35547	3742	6319
T_41_1_Y_30	Atlantic	2463935	2048843	29966	3186	5569
T_41_1_Y_31	Atlantic	3984362	3443014	33264	3422	5852
T_41_1_Y_33	Atlantic	3405123	2981862	31292	3412	5969
T_41_1_Y_36	Atlantic	4903957	4112642	34534	3599	6123
T_51_1_Y_11	Indian	4197832	3271746	30808	3259	5746
T_51_1_Y_12	Indian	5106469	3943972	32425	3204	5508
T_51_1_Y_13	Indian	3807089	2659014	30458	3347	5782
T_51_1_Y_15	Indian	5553686	4150067	33413	3463	5822
T_51_1_Y_16	Indian	4224280	3640016	31420	3275	5564
T_51_1_Y_19	Indian	3013892	2312388	30010	3178	5547
T_51_1_Y_28	Indian	3210746	2295315	28312	2962	5234

T_51_1_Y_3	Indian	2884006	2052075	29288	3368	6223
T_51_1_Y_4	Indian	5210472	3911316	26135	2810	4962
T_51_1_Y_5	Indian	2632718	1931146	28444	2763	4917
T_51_2_Y_1	Indian	3522515	3040289	32751	3432	6043
T_51_2_Y_10	Indian	6441211	6003918	31732	3619	6442
T_51_2_Y_11	Indian	2641836	2141771	30187	3242	5766
T_51_2_Y_27	Indian	2888780	2238918	29561	3295	5923
T_51_2_Y_34	Indian	2543649	2073189	29620	3215	5718
T_51_2_Y_44	Indian	3485004	2841056	35543	5501	9745
T_51_2_Y_45	Indian	3135178	2553943	40053	8862	15727
T_51_2_Y_46	Indian	2404384	1941743	29545	3300	5904
T_51_2_Y_47	Indian	2260180	1850287	33939	3679	6280
T_51_2_Y_50	Indian	3937595	3282788	34957	4138	7470
T_71_1_Y_11	Pacific	3897551	3283367	33373	3534	5988
T_71_1_Y_12	Pacific	3394137	2893282	32166	3287	5842
T_71_1_Y_13	Pacific	3604370	3047045	32116	3432	5895
T_71_1_Y_14	Pacific	4000961	3396357	33478	3500	6093
T_71_1_Y_16	Pacific	4727965	3962636	34144	3593	6179
T_71_1_Y_18	Pacific	3818853	3285112	33031	3446	5970
T_71_1_Y_2	Pacific	3440711	2874443	32188	3408	5963
T_71_1_Y_21	Pacific	2508039	2155027	34264	4617	8628
T_71_1_Y_22	Pacific	4044866	3428059	33271	3502	6084
T_71_1_Y_23	Pacific	5676853	4942237	35273	3714	6391
T_71_2_Y_15	Pacific	2656551	2123387	30040	3279	5745
T_71_2_Y_17	Pacific	3377858	2825637	31733	3351	5814
T_71_2_Y_2	Pacific	3159899	2710391	32016	3435	5923
T_71_2_Y_24	Pacific	4837695	4034112	34695	3706	6474

T_71_2_Y_25	Pacific	2863301	2313035	30574	3221	5797
T_71_2_Y_26	Pacific	3121822	2545630	31697	3416	5977
T_71_2_Y_27	Pacific	3391479	2748050	32031	3498	6099
T_71_2_Y_29	Pacific	3345330	2801175	32150	3377	5895
T_71_2_Y_30	Pacific	3719775	2980475	30729	3235	5603
T_71_2_Y_31	Pacific	3951818	3299857	42885	4454	8100
T_77_2_Y_11	Pacific	3618576	3216232	33632	3618	6333
T_77_2_Y_17	Pacific	2403163	1914776	29096	3118	5607
T_77_2_Y_18	Pacific	3866009	3386384	33704	3516	6125
T_77_2_Y_19	Pacific	2806651	2363857	21917	2321	4269
T_77_2_Y_2	Pacific	2780059	2340856	29617	3101	5546
T_77_2_Y_20	Pacific	2549065	2108849	29540	3193	5573
T_77_2_Y_27	Pacific	3398439	3004995	31710	3308	5713
T_77_2_Y_28	Pacific	2919624	2420989	30259	3243	5710
T_77_2_Y_32	Pacific	4317852	3725254	33004	3452	6035
T_77_2_Y_33	Pacific	4371397	3896432	33500	3513	5920
T_77_1_Y_30	Pacific	5765636	5148952	35233	3706	6160
T_77_1_Y_17	Pacific	2794243	2470957	31793	3342	5927
T_77_1_Y_19	Pacific	2723470	2254595	29008	2990	5105
T_77_1_Y_45	Pacific	2400483	2096522	30429	3239	5747
T_77_1_Y_1	Pacific	2155938	1855371	29363	3100	5584
T_77_1_Y_15	Pacific	2085535	1770279	29350	3182	5679
T_77_1_Y_5	Pacific	1857567	1612989	27970	2936	5210
T_77_1_Y_16	Pacific	2035715	1556718	28019	2930	5242
T_77_1_Y_22	Pacific	1796311	1478481	27735	2953	5322
T_77_1_Y_50	Pacific	1990282	1464790	26208	2779	4986

Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*)-SUPPLEMENTARY MATERIAL 4-

Supplementary Material 4- Results of the *analysis* of molecular variance (*AMOVA*), testing different grouping of the geographic samples: *1) #group1 (34_1, 34_2, 41_1, 31_1); #group2 (51_1 and 51_2); #group3 (71_1, 71_2, 77_1, 77_2); *2) #group1 (34_2, 51_1, 31_1, 77_2); #group2 (71_1, 41_1); #group3 (71_2, 77_1, 34_1, 51_2); *3) #group1 (31_1, 77_2, 51_2, 41_1); #group2 (34_2; 77_1); #group3 (71_2, 34_1, 71_1, 51_2); *4) #group1(34_1, 34_2, 41_1, 31_1); #group2 (51_1, 51_2, 71_1, 71_2, 77_1, 77_2).

*1 Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	2	532.21	2.50242	2.73
P value				<0.01
Among populations within groups	7	741.379	0.88479	0.97
P value				>0.01
Within populations	190	16763.346	88.22814	96.3
P value				<0.01
Total	199	18036.935	91.61535	

*2 Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	2	236.373	-0.47002	-0.52
P value				>0.01
Among populations within groups	7	1037.216	2.99888	3.3
P value				<0.01
Within populations	190	16763.346	88.22814	97.21
P value				<0.01
Total	199	18036.935	90.757	

*3 Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
---				
Among groups	2	226.016	-0.57412 Va	-0.63
P value				<0.01
Among populations within groups	7	1047.573	3.07290 Vb	3.39
P value				<0.01
Within populations	190	16763.346	88.22814 Vc	97.25
P value				>0.01

Total	199	18036.935	90.72692	

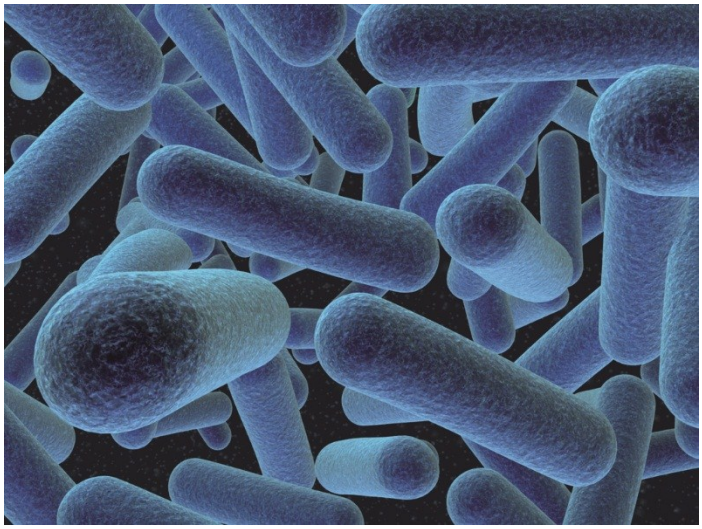
Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
---				
Among groups	1	344.227	2.37453 Va	2.58
P value				<0.01
Among populations within groups	8	929.362	1.39798 Vb	1.52
P value				<0.01
Within populations	190	16763.346	88.22814 Vc	95.9
P value				>0.01

Total	199	18036.935	90.72692	

Publication II

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.

Marianna Pauletto[1*], Lisa Carraro[1*], **Massimiliano Babbucci[1]**, Rosaria Lucchini[2], Luca Bargelloni[1], Barbara Cardazzo[1].



Abstract

The advent of Next-Generation Sequencing (NGS) has dramatically changed bacterial typing technologies, increasing our ability to differentiate bacterial isolates. Despite it is now possible to sequence a bacterial genome in a few days and at reasonable costs, most genetic analyses do not require whole-genome sequencing, which also remains impractical for large population samples due to the cost of individual library preparation and bioinformatics. More traditional sequencing approaches, however, such as MultiLocus Sequence Typing (MLST) are quite laborious and time-consuming, especially for large-scale analyses. In the present study a genotyping approach based on restriction site-associated (RAD) tag sequencing, 2b-RAD, was applied to characterize *Listeria monocytogenes* strains. To verify the feasibility of the method, an *in silico* analysis was performed on 30 available complete genomes. For the same set of strains *in silico* MLST analysis was conducted as well. Subsequently, 2b-RAD and MLST analyses were experimentally carried out on 58 isolates collected from food samples or food-processing sites.

The obtained results demonstrate that 2b-RAD predicts MLST types and often provides more detailed information on population structure than MLST. Moreover, the majority of variants differentiating identical ST isolates mapped against accessory fragments, thus providing additional information to characterize strains. Although MLST still represents a reliable typing method, large-scale studies on molecular epidemiology and public health, as well as bacterial phylogenetics, population genetics and biosafety could benefit of a low cost and fast turnaround time approach such as the 2b-RAD analysis proposed here.

Background

Listeria monocytogenes is a Gram positive foodborne pathogen that is both saprophytic and parasitic (*Parisi et al. 2010*). This bacterium is ubiquitous in soil, vegetation, and other environmental sources. Moreover *Listeria* can contaminate foods of animal origin in processing plants and retail establishments and it is difficult to eradicate, being able to grow at refrigeration temperatures and to resist to acid and high-salt concentrations. *L. monocytogenes* is the causative agent of listeriosis, a severe food-borne disease (*Sauders et al. 2012*), with a relatively low incidence, a mortality rate of 20–30% and a hospitalization rate over 92% (*Ward et al. 2008; Lomonaco et al. 2015*). For such reasons it is recognized as a public health issue and a serious challenge for the food industry (*Ferreira et al. 2014; Rückerl et al. 2014*).

Several methods for typing *L. monocytogenes* isolates are available (*Jadhav et al. 2012*) and the intended application would ultimately determine the most appropriate typing strategy. Traditional

subtyping was based on serotyping, but recently the application of molecular typing methods was recommended by EFSA Panel on Biological Hazards on the management of the major food-borne microbiological hazards (*EFSA BIOHAZ Panel 2104*).

Among such methods, Pulse Field Gel Electrophoresis (PFGE) was established as the gold standard approach for epidemiological studies in foodborne outbreaks (*Jiang et al. 2008*). This method is widely used for small-scale epidemiology, listeriosis surveillance, and outbreak investigations despite practical disadvantages, as it is time-consuming and requires stringent standardization for inter-laboratory data comparison (*Chenal-Francisque et al. 2013*). Furthermore, this method provides only limited information on the phylogenetic relationships among strains, which is a serious limitation to understand the evolution of important phenotypic traits such as virulence. MultiLocus Sequence Typing (MLST) is another well-established method for microbial ecology (*Maiden 2006*). MLST is based on core genome sequences and has been extensively used to investigate broad population structure of several bacterial species. In the case of *L. monocytogenes,* several studies used MLST (*Salcedo et al. 2003; Nightingale et al. 2005; Ragon et al. 2008; Wang et al. 2012b*), the most recent one reporting the analysis of more than 1,000 isolates, across wide temporal, geographical, and source diversity (*Haase et al. 2014*). However, MLST is neither rapid nor cheap and it has demonstrated limited discriminatory power within *L. monocytogenes* (*Chenal-Francisque et al. 2013*). Additional molecular typing methods have been proposed, such as multi-locus variable-number tandem repeat analysis (MLVA) (*Chenal-Francisque et al. 2013*), multi-virulence-locus sequence typing (MVLST) (*Zhang et al. 2004; Chen et al. 2007*), lineages (*Tsai et al. 2011*; *Orsi et al. 2011*), single nucleotide polymorphism (SNP) (*Ward et al. 2008; Lomonaco et al. 2011*) and clustered regularly interspaced short palindromic repeats (CRISPR) (*Sesto et al. 2014*). These molecular methods have been broadly exploited and they all reported that the population structure of *L. monocytogenes* is largely clonal and consists of four lineages (I–IV), of which lineages I and II represent more than 95% of all analysed isolates.

In recent years, High-Throughput Next-Generation Sequencing (HT-NGS) has become increasingly popular, replacing classical microbiology and first-generation bio-molecular technologies in several fields, as HT-NGS delivers sequence data thousands folds more cheaply than Sanger sequencing (*Loman et al. 2012*). Cost-effective high-throughput approaches that combine MLST scheme and NGS have been proposed (*Boers et al. 2012; Singh et al. 2012*), considerably reducing labour and costs compared to MLST using traditional Sanger sequencing. Moreover, thanks to NGS methodologies, several genomic comparative studies have been recently conducted, leading to considerable progresses in diagnostic/epidemiology, phylogenetics, and functional characterization of bacterial strains (*den Bakker et al. 2008; Orsi et al. 2011; Hain et al. 2012*).

By rapidly revealing bacterial genome sequences, whole-genome sequencing (WGS) might soon become the reference for microbial diagnostics and genotyping. On the other hand, it is often neither necessary nor feasible to sequence the whole genome for most genetic analyses in bacteria. In fact, WGS for microbial population genomics requires the preparation of individual strain libraries, which remains relatively expensive. The most relevant hurdle, however, are the intensive computational burden and the need to achieve sufficient coverage at shared genomic locations to ensure accurate genotyping. All such requirements remain still prohibitive for most laboratories, especially in presence of a large amount of isolates to be tested. Hence, selectively capturing a large number of defined genomic regions followed by NGS analysis could provide optimal trade-off for time- and cost-effective high-resolution molecular typing of a sizable set of strains. Restriction site-Associated DNA (RAD) tag sequencing is one of the most exploited genotyping techniques in a broad range of eukaryotic species. RAD-tag sequencing reduces genome complexity by re-sequencing only DNA regions adjacent to recognition sites of a chosen restriction endonuclease (for review, see *Davey et al. 2011*). By sequencing only those tags flanking a restriction site in multiplexed, individually-barcoded samples, RAD sequencing allows efficient high-throughput identification of Single Nucleotide Polymorphisms (SNPs). For such reasons it has proven to be a powerful tool for genetic mapping and analysis of quantitative trait loci (*e. g. Baird et al. 2008; Chutimanitsakun et al 2011; Pujolar et al. 2013; Guo et al. 2014*), adaptation (*e. g. Hohenlohe et al. 2010*) and phylogeography (*e. g. Emerson et al. 2010; Cromie et al. 2013*).

With the aim to develop a simple, rapid, and low cost typing method to investigate bacterial population structure in food environment isolates, a genome-scale approach was successfully applied here to *L. monocytogenes* species. The method extends 2b-RAD, a RAD-tag sequencing technique that was originally developed for genomic analysis of eukaryotic genomes (*Wang et al. 2012a*). Briefly, 2b-RAD is based on sequencing reduced representation libraries and compared to WGS, allows the identification of a lower set of SNPs though at a higher mean coverage and with greatly reduced computational costs.

In order to verify the feasibility of 2b-RAD in bacteria, and more specifically in *Listeria*, firstly an *in silico* analysis was performed on 30 *L. monocytogenes* strains for which the complete genome was available. For the same set of strains *in silico* MLST analysis was conducted as well. 2b-RAD and MLST analyses were then experimentally carried out on a group of 58 *L. monocytogenes* isolates collected from food samples or food processing sites.

A comparative analysis proved that bacterial 2b-RAD genotyping is a streamlined and cost-effective method, achieving sufficient accuracy and providing improved discrimination power to evaluate *L. monocytogenes* population structure.

Results

In silico analysis

MLST typing was carried out *in silico* for all available complete genomes (30), representing lineages I, II and III. A total of 22 unique sequence types (STs) representing lineages I, II and III were identified, by comparing *in silico* allelic profiles to those available in the MLST *L. monocytogenes* database (Table S2). MLST data were analysed with STRUCTURE. Multiple runs in STRUCTURE with different values of K (the number of putative ancestral populations) were carried out and, as reported in Figure S1a, the model probability increased dramatically when K value was equal to 3. The three "ancestral" populations reflected the three *Listeria* lineages represented in the dataset: cluster 1, 2 and 3 corresponded to Lineage I, II and III, respectively. Except for g17 and g28, all the strains showed a 100% of membership to a single ancestral population. Strain g17 (L70 *L. monocytogenes* draft genome, ST18) was assigned to cluster 2 with a 91% of probability and to cluster 1 with a 9% of probability. In strain g28 (SLCC2376, ST71) up to the 82% of the genetic material come from the ancestral lineage III population, while the remaining 18% derived from lineage I cluster.

Likewise, clonal relationships within the *L. monocytogens* population were analysed on MLST profiles using PHYLOViZ software (Figure 1a) by applying a threshold of 7 loci (*i. e.* by cutting those links in the MST diagram that connected strains differing by more than 6 loci), allowed the identification of three distinct groups reflecting the three *Listeria* lineages, each highlighted by a different colour. Since the software automatically removes identical allelic profiles, in Figure 1a strains with the same ST are represented by a single sample. A total of eight samples are removed: g02 (ST1 as g01), g04 (ST2 as g03), g07 and g08 (both ST4 as g06), g13 (ST9, identical to g12), g16 (ST12 like g15), g24 (ST121 as g23) and g30 (ST201, identical to g29).

The total amount of predicted 2b-RAD fragments with both *AlfI* and *CspCI* enzymes is reported in Table S3. The mean number of fragments obtained through *in silico* digestion of complete genomes performed with *AlfI* was higher than that observed after *CspCI* digestion, in agreement with the fact that the recognition site for *CspCI* is composed by seven base pairs (bp), while *AlfI* recognizes six bp.-Both *in silico* and experimental 2b-RAD experiments were conducted using *AlfI* and *CspCI*. Since *CspCI* produced a smaller number of informative tags and analysis of combined *CspCI-AlfI* data yielded results that were overlapping with those obtained just using *AlfI* (results not shown), only the latter data are presented and discussed here.

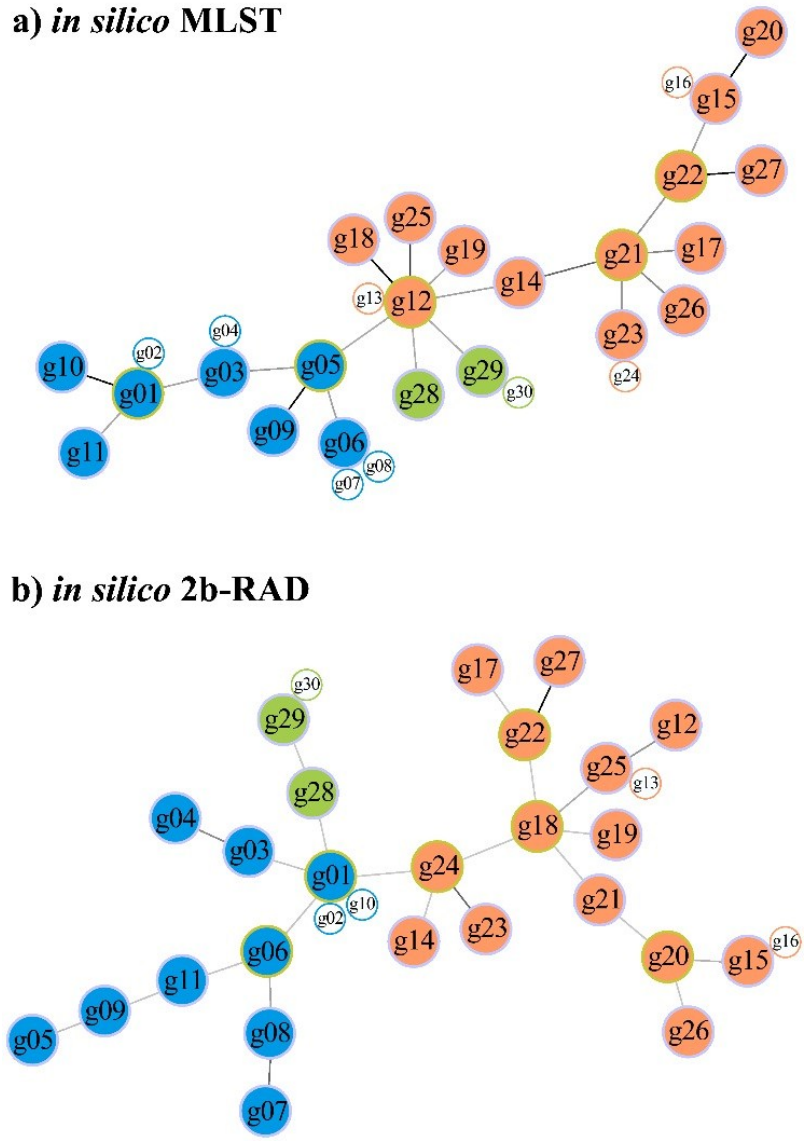


Figure 1. Minimum spanning tree (MST) – *in silico* analysis. MSTs generated from MLST data **(a)** and 2b-RAD analyses **(b)**. Strains are identified by the IDs introduced in Table 1 (g01-g30). Colours refer to the three different lineages: lineage I (blue), lineage II (orange) and lineage III (green). Empty circles identify strains having an allelic profile identical to the closer filled circle

STACKS analysis allowed the identification of 1,279 unique loci (*i. e.* the catalog). A total of 1,084 loci (85%) mapped on *L. monocytogenes* core and accessory regions. Out of them, 261 were polymorphic tags (at least one SNP). The STACKS variant output file was deposited into the Dryad Data Repository (DOI:10.5061/dryad.m1bm5). Percentages of tags mapping on both core and accessory fragments are reported in Table 2.

Table 2. Mapping statistics. The table reports the number of STACKS loci mapping against both core and accessory regions assembled by Panseq.

	in silico **2b-RAD**	**experimental 2b-RAD**
N° loci	1279	923
N° loci mapping	1084 (85%)	745 (81%)
N° loci mapping on core regions	472 (44%)	405 (54%)
N° loci mapping on accessory regions	612 (56%)	340 (46%)
N° polymorphic loci mapping	261 (24%)	203 (27%)
N° polymorphic loci mapping on core regions	125 (48%)	119 (59%)
N° polymorphic loci mapping on accessory regions	136 (52%)	84 (41%)

The analysis conducted with STRUCTURE showed that the probability of the data was maximal at K=3 (Figure S1b) and each cluster included strains belonging to a single lineage. Notably, these results are largely in agreement with those inferred from MLST data (Figure S1a). The genetic material of most of the strains descended from a single population with 100% probability. As seen in the MLST data analysis, the greatest fraction (73%) of g28 genetic material descended from the ancestral population 3 (lineage III), while its proportional membership to cluster 1 (lineage I) and 2 (lineage II) was 14% and 13%, respectively.

MST analysis performed with PHYLOViZ on 2b-RAD data (Figure 1b) recovered a network where three *L. monocytogenes* lineages were clearly distinct (highlighted by different colours). In Figure 1b identical profiles are represented by a single sample. Thus, g02 and g10 (both identical to g01), g16 (=g15), g13 (=g25) and g30 (=g29) have been discarded. A total of 25 distinct allelic profiles (*i. e.* filled circles) were identified, three more than the effective number of STs obtained from the same set of strains (Figure 1a).

Experimental genetic analysis

MLST analysis was carried on a total of 58 samples: 56 field (Table S1) and 2 reference strains. As reported in Table S2, sequence comparison with the MLST database identified a total of 17 unique STs belonging to lineages I and II. The most represented strains were those identified by ST121 (23 isolates) and ST9 (7 isolates). MLST profiles were deposited in the MLST database (http:/www.pasteur.fr/mlst).

Figure 2a illustrates STRUCTURE results with K equal to 2, the fixed value at which the model probability was the highest. Samples were mostly assigned to a single cluster with a 100% of probability, except for L56 (ST14) and L67, L69 and L70 (classified all as ST18), which showed a mixed genetic composition. Nevertheless, these four strains exhibited on average 90% of genetic material from cluster 2 and only a negligible proportional assignment to cluster 1 (less than 10%).

The MST computed by PHYLOViZ showed 17 unique STs, which were organized in a network where each circle represents a unique ST (reported in the circle) and its size is proportional to the number of samples in the dataset belonging to that ST (Figure 3a).

The raw reads obtained by the Illumina sequencing of *AlfI*-digested fragments were deposited in the NCBI Sequence Read Archive under the study accession number SRP058349. Reads were filtered by quality, thus resulting in a total of 9,134,940 reads to be used as input for STACKS. The total number of unique STACKS fragments per sample produced and sequenced with 2b-RAD is reported in Table S4. STACKS catalog consisted on 923 unique loci. Out of the 923 unique fragments obtained by *AlfI* digestion, 356 presented at least one SNP. The STACKS variant output file was deposited into the Dryad Data Repository (DOI:10.5061/dryad.m1bm5). Consistency across replicated samples was assessed by checking both unique STACKS tags and the input matrix used for population structure analysis. The comparison of the number of unique STACKS tags between TRs demonstrated a maximal coefficient of variation of 2.2%, measured in the two L18 technical replicates. Among the 356 polymorphic loci identified by STACKS, the maximum number of tags showing a different allele between each pair of technical replicates was 3. Notably all these incongruences were ascribed to presence/absence status of a tag rather than a different base (*i. e.* sequencing error).

A total of 754 (81%) out of the 923 unique loci mapped on *L. monocytogenes* core and accessory regions determined by a Panseq analysis with user-defined parameters (see section 2.1). Among the mapped loci, 203 (27%) were polymorphic (SNPs $\geq$ 1). Detailed statistics of the number of tags and percentages mapping on both core and accessory fragments are reported in Table 2.

While population structure predicted by using MLST data suggested two clusters (K=2), analysis of 2b-RAD data with the same software (STRUCTURE) showed the maximal probability at K=3 (Figure 2b). Cluster 1 reflected lineage I strains while clusters 2 and 3 both included lineage II strains. Replicated samples were always assigned to the same cluster. Samples appeared to be genetically homogeneous except for L56 and L57, which exhibited a mixed composition. Strain L56 was assigned to cluster 2 and cluster 3 with a probability of 44% and 66%. Conversely, L57 was assigned to cluster 2 with a probability of 58% and to cluster 3 with a probability 42%. A few additional isolates showed a mixed composition (*e. g.* L60, L63, L20, L39), but they were still assigned to a single cluster with a probability higher than 90%. The MST analysis performed with PHYLOViZ on 2b-RAD data recovered a network where nearly all isolated were represented. In fact, the total number of unique allelic profiles, including TRs, was equal to 59 (*i. e.* the number of filled circles). Samples ATCC19117 TR3, TR4 and TR5 disappeared from the diagram because their profiles were identical to those obtained for ATCC19117_TR1, TR2 or TR3. Likewise L18_TR2 and L70_TR2 were absent because of they were equal to L18_TR1 and L70_TR1, respectively. Moreover, among the isolates

belonging to the ST121 group, L42 was not reported because it was identical to L41. Finally L67 and L69 (both ST18) were identical to L70_TR3. Figure 3b shows the complete MST divided by groups and allowing only links between two nodes of distances equal or less than 6.

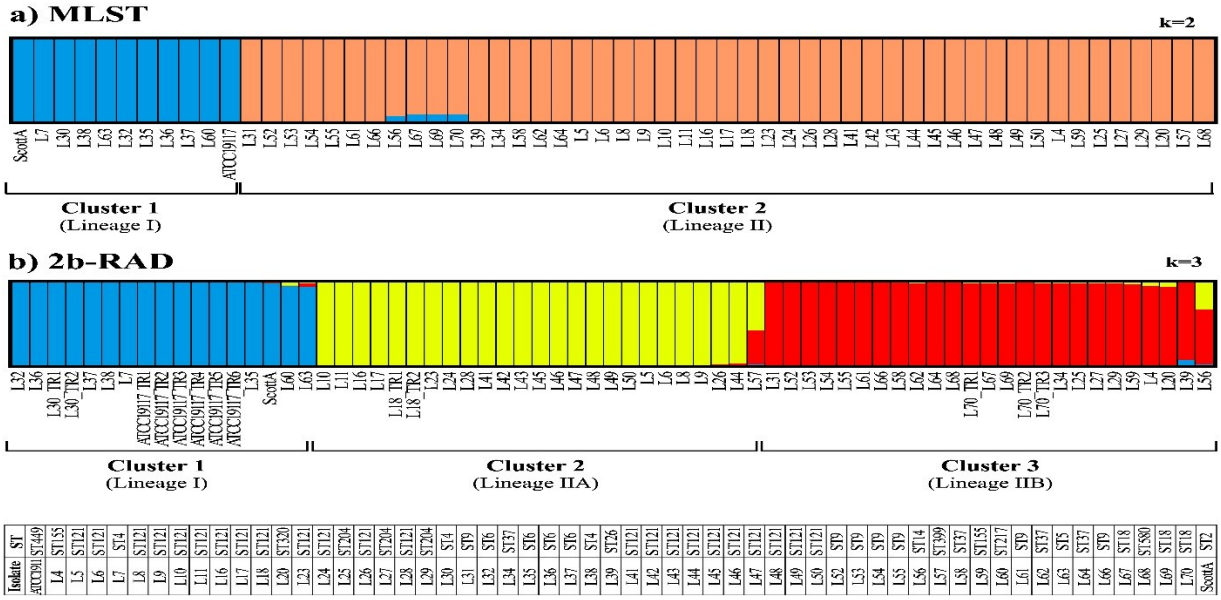


Figure 2. STRUCTURE bar-plot - laboratory analysis. (a) Estimated population structure generated from MLST data: proportions of ancestry from ancestral lineage I (blue) and ancestral lineage II (orange) as inferred by STRUCTURE assuming K = 2 ancestral subpopulations. **(b)** Estimated population structure generated from 2b-RAD data: proportions of ancestry from ancestral lineage I (blue), ancestral lineage IIA (yellow) and IIB (red) as inferred by STRUCTURE assuming K = 3 ancestral subpopulations. Each vertical bar represents an individual sample or technical replicate (TR). The table at the bottom of the figure specifies the ST number for each isolate.

This means that in Figure 3b links are drawn only if the number of differences between two samples is equal or inferior to 6. This threshold level has been conservatively chosen because it represents a value three times higher than the one at which TRs are still linked (level=2). Basically, samples linked and forming a "group" (a cluster of samples) were considered to have a degree of internal differences that is similar to that expected between technical replicates. Conversely, samples of different groups were characterized by a number of differences higher than that arising from 2b-RAD variability across TRs. Moreover, samples that are not linked to any other sample, each considered as a single group, were characterized by a unique profile with more than six different alleles compared to all remaining samples. In Figure 3b, each ST appears as a single cluster of samples or as a single sample (when a ST was represented by a unique sample). In addition, in the case of ST6, ST9, ST37, ST121 and ST155 (highlighted by dashed lines in Figure 3b) the MST generated more than a single group. Overall, 2b-RAD typing recognized a total of 23 distinct groups, thus outperforming MLST, which

identified 17 single STs. Specifically, the MST highlighted the separation of L35 (from the ST6 group), L31 (ST9), L34 (ST37), L44 (ST121) and L4 (ST155) from their closest strains with the same ST type. The majority of polymorphic loci discriminating each of the aforementioned isolates in PHYLOViZ mapped against *L. monocytogenes* accessory fragments.

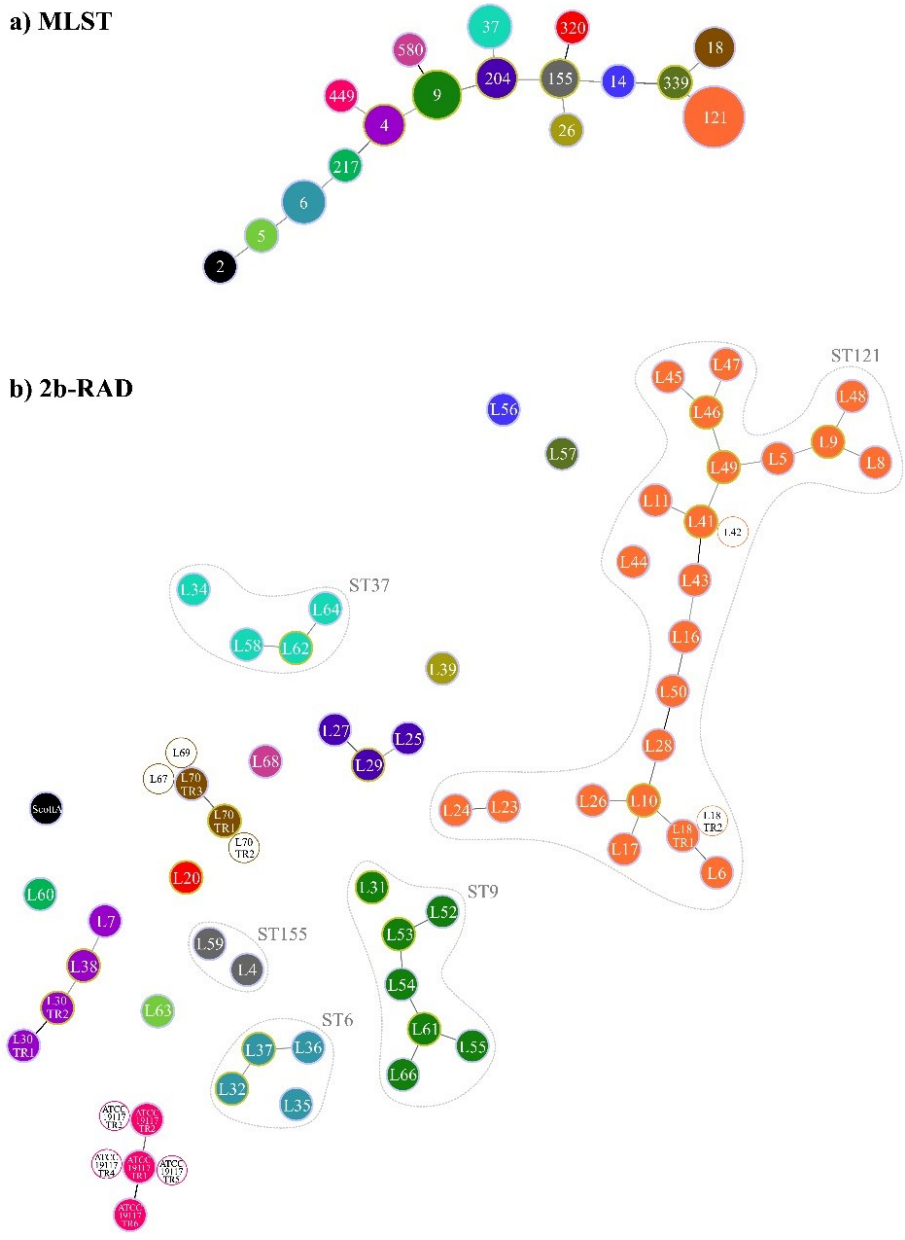


Figure 3. Minimum spanning tree (MST) - laboratory analysis. (a) MST generated from MLST data at a threshold level of 7. Each circle identifies a single ST and circle sizes are proportional to the number of strains in the dataset belonging to that ST. **(b)** MST generated from 2b-RAD data. Strains are identified by the IDs introduced in Table S1. The pattern of colours is the same as in figure 3a and each colour reflects a single ST. Empty circles identify strains/TRs having an allelic profile identical to the closer filled circle. Dashed lines define ST generating more than a single group.

Discussion

The overall aim of the present study was to verify the feasibility of a 2b-RAD typing as a streamlined alternative to the well-established MLST to investigate microbial population structure using *L. monocytogenes* as a case study.

PFGE, the reference typing method for *L. monocytogenes*, is in fact a valuable tool for recognizing common source outbreaks, but it suffers from a few disadvantages such as being time-consuming and requiring stringent standardization for inter-laboratory data comparison (*Chenal-Francisque et al. 2013*). Furthermore, PFGE is unsuitable for inferring genetic relationship between isolates, which is an important limitation to understand the evolution of key phenotypic traits. On the contrary, MLST is an appropriate method studying broader genetic relationships based on core genome sequences, but it is neither rapid nor cheap and it showed a limited discriminatory power within *L. monocytogenes* (*Chenal-Francisque et al. 2013*).

The present study demonstrated that 2b-RAD has a discriminatory power within *L. monocytogenes* strains that is greater than that reported for MLST. Allelic profiles generated by *in silico* MLST and 2b-RAD predicted three genetically different ancestral populations (*i. e.* three *L. monocytogenes* lineages), showing a substantial similarity between the two typing approaches. Conversely, the network analysis conducted by means of PHYLOViZ suggested greater resolution between strains with 2b-RAD (Figure 1). In fact, the MST constructed from 2b-RAD data (Figure 1b) yielded 25 unique allelic profiles, three more than the total number of unique MLST-based allelic profiles (Figure 1a). Notably, 2b-RAD genotyping provided a higher discriminatory power being able to identify a specific genotype also for strains g04 (J1 220) g07 (Clip80459), g08 (L312) and g24 (N53 1), which were otherwise identical to other strains when analysed by MLST. On the opposite, although strains g01 (F2365) and g10 (SLCC2378) belong to distinct ST groups, they showed an identical 2b-RAD-based allelic profile. This is the specific case of ST1 and ST73 (g01 and g10, respectively), which are different only on the basis of the MLST *ldh* allele (*i. e.* a single nucleotide polymorphism). This result can be explained in the context of *ldh* gene mutations. In fact, Cantinelli and colleagues have recently suggested that in *L. monocytogenes* this gene is characterized by atypical variations, which could have been selected during long-term storage in the laboratory (*Cantinelli et al. 2013*). The improved typing ability of 2b-RAD compared to MLST was confirmed by the experimental analysis conducted on 58 different isolates (Table S1).

Analysis with STRUCTURE conducted on MLST experimental data predicted two main clusters, each corresponding to one *L. monocytogenes* lineage (Figure 2a), while the same analysis on 2b-RAD data identified three main genetically homogenous ancestral populations, discovering hidden genetic structure within lineage II isolates (Figure 2b). In fact, ST121 isolates formed a genetically

homogeneous cluster (cluster 2 in Figure 2b) separated from cluster 3, which included ST9, ST14, ST18, ST26, and ST37. ST121, frequently isolated from food samples in several European countries, has been already reported to carry an insertion typical of *L. innocua* strains (2.2 kb) within the Stress Survival Islet 1 (SSI1, 9.7 Kb) (*Hein et al. 2011*). The presence of a region possibly originating from *L. innocua* in the genome of ST121 isolates might explain the clustering of this ST as a separate lineage in STRUCTURE. Using PCR assays, we demonstrated that all the ST121 isolates of the experimental population contained the 2.2 kb insertion in the location of SSI1, unlike the remaining isolates that contained the *L. monocytogens* classical SSI1 (9.6 Kb) or were negative for the SSI1 (1.1 Kb) (data not shown).

The great potential of 2b-RAD for bacterial strains typing was further evidenced by PHYLOViZ analysis. Each MLST-based ST appears as a single cluster of samples or as a single sample in the 2b-RAD-based network, thus confirming the consistency between NGS results and the MLST. Noteworthy, given the large amount of data currently available in MLST databases, maintaining the backwards compatibility with MLST schemes appears to be crucial for any new bacterial typing method, so the results can be interpreted in their proper historical context.

In addition, PHYLOViZ analysis demonstrates that 2b-RAD has greater discrimination power than MLST, recognizing 23 distinct groups (Figure 3b).

Mapping 2b-RAD tags against core and accessory genomes obtained with *Panseq* showed unbiased representation of these two genomic compartments (Table 2). Despite *Panseq* software has been already demonstrated to efficiently find and extract *L. monocytogenes* novel accessory as well as core genomic information for a group of genomic sequence (*Laing et al. 2010*), here a small percentage of both *in silico* and experimental 2b-RAD tags did not mapped neither against core nor accessory genome. Most likely, the *Panseq* assembly algorithm, producing a series of fixed and user-defined length fragments, prevents mapping of tags matching at 5' and 3' terminal portions of fragments and spanning two different, but consecutive fragments. Interestingly, the majority of polymorphic 2b-RAD loci discriminating strains L4, L31, L34, L35 and L44 from their closest isolates, which were classified as the same ST, are located in accessory *L. monocytogenes* fragments.

Equal representation of core and accessory genome fragments in 2b-RAD analysis is extremely relevant as demonstrated by a recently published study, which highlighted the importance of the accessory genome for characterizing and typing nosocomial strains of *Acinetobacter baumannii* (*Turton et al. 2011*). Likewise, multiple evidence in *L. monocytogenes* suggests that the accessory genome is involved in pathogenicity and stress resistance (*Mellin & Cossart 2012*; *Fuchs et al. 2012*; *Behrens et al. 2014*), thus being of large interest (*den Bakker et al. 2013; Kuenne et al. 2013*). To date, several studies have proposed alternative bacterial typing approaches combining MLST

schemes and NGS methods (*e. g. Larsen et al. 2012; Zhao et al. 2012*) or targeting a set of SNPs located in the MLST genes (*Eusebio et al. 2013a; Eusebio et al. 2013b; Trembizki et al. 2014*) but all these methods rely on a specific number of polymorphic alleles located in house-keeping genes or core genome fragments, losing the key information contained in the accessory genome.

The 2b-RAD typing approach proposed here was compared with MLST and WGS method focusing on output, computational, and laboratory requirements, costs and turn-around time (Table 3). Notably, WGS provides all or nearly all genetic information of bacterial isolates and can theoretically distinguish strains differing even for a single nucleotide, but it requires

high computing power and is still laborious, time-consuming and expensive (115 euro per sample, see Table 3 for specifications on how prices were estimated). MLST is well established, with large databases already available, is cheaper than WGS (about 45 euro per sample) and applicable in laboratories without expertise in genomic analysis, but it provides limited information (variable and depending on the number of loci investigated), and is labour-intensive, especially for large numbers of samples. 2b-RAD typing allows a reliable, time- and cost-effective approach (about 19 euro per sample) to discover a subsample of genome-wide variations across hundreds of strains in a single run. Table 3 clearly demonstrates that 2b-RAD, compared to WGS and MLST, provides a good compromise between information content, practical applicability, and costs.

Although the method proposed here appears to be a cost-effective and promising molecular approach, an important issue is the development of a database for *L. monocytogenes* 2b-RAD data.

Table 3. *L. monocytogenes* genotyping methods comparison. The table shows outputs, costs and time calculated on the basis of the starting samples (*n*). As far as 2b-RAD and WGS, to allow an effective methods comparison, *n* has been defined as the maximum number of samples which can be reliably processed achieving a sufficient coverage in an Illumina Miseq, the instrument employed in the present study.

[1]*include costs of kit and other laboratory reagents and supplies*
[2]*refer to the current mean prices offered by external providers*

	MLST (*n*=96)	**2b-RAD (*n*=192)**	**WGS (*n*=20)**
Sample prep	as in MLST Pasteur Web Site	as in *Wang et al. (2012)*	Nextera® XT Kit
Sequencing platform	Sanger	Illumina Miseq 50 SE v2 kit	Illumina Miseq 150 PE v2 kit
Max sequencing output	200 Mb	850Mb (15 M reads)	4.5 Gb (30 M reads)
Hardware requirements	Desktop PC	Desktop PC (32G RAM)	Server (64G RAM)
Software requirements	Window/Unix	Unix	Unix
Theoretical marker loci	7 sequences (≈500 bp each)	≈1000 tags (≈34 bp each)	Whole-genome
Max coverage	2X	≈80X	≈75X
Sample prep costs (€/sample) [1]	5	12	31
Sequencing costs (€/sample) [2]	40	7	84
Total costs (€/sample)	45	19	115
Estimated days for DNA isolation-library prep	7	7	2
Estimated days for Bioinformatic analysis	30	4	14
Turn-around time (days)	37	11	16

Since a publicly available data repository and the establishment of a common approach are important prerequisites for the development of new methods in microbial ecology, a *L. monocytogenes* 2b-RAD analysis workflow was devised (Figure S2), which might help in the definition of a common nomenclature for the identification of isolates. The suggested workflow implies the use of a catalog of 2b-RAD loci (the first version of which was built in the present study and deposited in Dryad database) and provides the possibility of comparing new strains with previously analysed ones. The pipeline automatically updates the catalog by including eventual new tags and provides a STACKS tabular file containing the genotypes of all evaluated isolates at all discovered loci (*i. e.* a database). The approach suggested here is relatively simple and do not rely heavily on human input, yet it would be still advisable the intervention of a curator, to organize catalog releases, to supervise haplotype updates, and possibly to publish such information in a specific internet site.

The availability of different type IIB restriction enzymes and the possibility of adjusting the number of 2b-RAD tags using selective adaptors (*Wang et al. 2012a*), make 2b-RAD genotyping extremely versatile. In addition, a restriction enzyme can be employed to type different bacterial species without

a priori sequence information, opposite to what happens with MLST, where species-specific PCR primers need to be designed on target gene sequences. This provides a great potential for customization and extension to less characterized bacterial species. Versatility of 2b-RAD has been recently demonstrated in *Drosophila* species (*Seetharam & Stuart 2013*). To our knowledge, however, neither 2b-RAD nor other RAD-like approaches have been implemented in prokaryotes. In the present study, the applicability of 2b-RAD for bacterial typing has been tested for the first time. In addition to significant information content and discrimination power, as discussed above, analysis of technical replicates showed high genotyping accuracy and reproducibility, with few missing tags between replicates. Routine analysis of technical replicates is, however, highly recommended, as already suggested for 2b-RAD and other RAD-like technologies used in animals and plants (*Mastretta-Yanes et al. 2015*). This does not represent a limitation for 2b-RAD since up to 384 individual barcodes are available. Moreover, the multiplexing strategy, coupled with the use of higher throughput sequencing platforms (*e. g.* Illumina HiSeq 2500, NexSeq500), ensures the possibility to genotyping hundreds of isolates/strains, including technical replicates, at reasonable costs.

In conclusion, the present work represents just a first example of application of RAD-like technologies in microbial genetics and further studies are required to fully assess the potential use of such methods in prokaryotes. On the other hand, the obtained results already suggest that 2b-RAD might prove an extremely useful tool for molecular epidemiology (*e. g.* disease transmission, evolution of virulence) and public health (*e. g.* monitor vaccination programs), as well as in other areas such as phylogenetics, taxonomy, speciation, population genetics, and biosafety. It would be highly desirable a similar comparative study on a larger set of samples including the PFGE reference method in order to verify the reliability of 2b-RAD genotyping in the context of large-scale bacterial traceability.

Materials and Methods

In silico experiments

A total of 30 *L. monocytogenes* genomes (strains information and accession numbers in Table 1) were used for *in silico* comparative analysis of two different genotyping methods: MLST and 2b-RAD. **i) MLST:** the MLST scheme is based on sequence analysis of a portion of the following seven housekeeping genes: *acbZ* (ABC transporter), *bglA* (beta-glucosidase), *cat* (catalase), *dapE* (succinyl diaminopimelate desuccinylase), *dat* (D-amino acid aminotransferase), *ldh* (lactate deshydrogenase), and *lhkA* (histidine kinase) as proposed by Ragon et colleagues (*Ragon et al. 2008*). These gene partial sequences were extracted from the 30 genomes and each allele was compared to those available in

the MLST *L. monocytogenes* database (http://www.pasteur.fr) by using blast similarity searches. Strains

Table 1. Dataset for the *in silico* analysis. The table reports the list of the 30 strains and their genome accession numbers (when public). For the sack of clarity each strain has been labelled with a single ID (g01 to g30).

ID	Strain	Accession number
g01	*Listeria monocytogenes F2365*	NC_002973.6
g02	*Listeria monocytogenes LL195*	NC_019556.1
g03	*Listeria monocytogenes ATCC 19117*	NC_018584.1
g04	*Listeria monocytogenes J1 220*	NC_021830.1
g05	*Listeria monocytogenes SLCC2482*	NC_018591.1
g06	*Listeria monocytogenes 07PF0776*	NC_017728.1
g07	*Listeria monocytogenes Clip80459*	NC_012488.1
g08	*Listeria monocytogenes L312*	NC_018642.1
g09	*Listeria monocytogenes SLCC2755*	NC_018587.1
g10	*Listeria monocytogenes SLCC2378*	NC_018585.1
g11	*Listeria monocytogenes SLCC2540*	NC_018586.1
g12	*Listeria monocytogenes FSL R2 561*	NC_017546.1
g13	*Listeria monocytogenes SLCC2479*	NC_018589.1
g14	*Listeria monocytogenes J0161*	NC_017545.1
g15	*Listeria monocytogenes EGD*	NC_022568.1
g16	*Listeria monocytogenes SLCC5850*	NC_018592.1
g17	*Listeria monocytogenes L70*	LDJD00000000
g18	*Listeria monocytogenes EGD e*	NC_003210.1
g19	*Listeria monocytogenes L64*	LDJC00000000
g20	*Listeria monocytogenes 10403S*	NC_017544.1
g21	*Listeria monocytogenes SLCC7179*	NC_018593.1
g22	*Listeria monocytogenes 08 5923*	NC_013768.1
g23	*Listeria monocytogenes La111*	NC_020557.1
g24	*Listeria monocytogenes N53 1*	NC_020558.1
g25	*Listeria monocytogenes SLCC2372*	NC_018588.1
g26	*Listeria monocytogenes Finland 1998*	NC_017547.1
g27	*Listeria monocytogenes 08 5578*	NC_013766.2
g28	*Listeria monocytogenes SLCC2376*	NC_018590.1
g29	*Listeria monocytogenes M7*	NC_017537.1
g30	*Listeria monocytogenes L99*	NC_017529.1

with identical allelic profiles were identified with the same ST. **ii) 2b-RAD:** *in silico* digestion with *AlfI* and *CspCI* type IIB restriction enzymes and identification of restriction fragments were separately performed by using the bioinformatic scripts available at http://people.oregonstate.edu/~meyere/tools.html. Genotype analysis for individual sample was carried out by running the *denovo_map.pl* pipeline implemented in STACKS (*Catchen et al. 2013*), setting the following parameters: m=3, n=2, M=2. This pipeline aims at building loci, creating a catalog of loci, and matching samples against such a catalog, followed by genotype assignment, through the identification of SNPs located on *in silico* digested fragments. To evaluate whether polymorphic loci were localized mainly in accessory or core genes, tag sequences (34 bp) were mapped against *L. monocytogenes* core and accessory genome fragments, which were assembled as described below. Mapping analysis was carried out by means of *CLC Genomic Workbench 7.5* (length fraction = 0.8; similarity fraction = 0.7; non-specific match handling: ignore). **iii) Pangenome**:

putative core and accessory *L. monocytogenes* genomic regions were identified by using the *Panseq* web-based software (*Laing et al. 2010*). Panseq "pan" analysis was performed by setting a minimum fragmentation size of 2000 bp, a minimum novel region size of 50 bp and the maximum core genome threshold (*i. e.* the value corresponding to the number of analysed strains). Core and accessory genome fragments were deposited into the Dryad Data Repository (DOI:10.5061/dryad.m1bm5).

Laboratory experiments

A total of 56 field *L. monocytogenes* strains was analysed. All strains were collected from various food products or food-processing plants from different Italian regions (detailed information in Table S1). In addition, two *L. monocytogenes* reference strains (ScottA and ATCC19117) were obtained and included in the study. Three out of the whole set of analysed strains were also evaluated during the *in silico* experiment: L64, L70 and ATCC19117. L64 and L70 (g19 and g17 respectively) are two field strains belonging to the BCA Department Collection, for which a draft genome was sequenced and deposited in Genbank (LDJC00000000, LDJD00000000). ATCC19117 (g03) is a reference strain that was selected for the laboratory experiment being its genome sequence already available in Genbank (NC_018584.1). Strains were stored at −80 °C in TSB (Tryptic Soy Broth, Oxoid, Thermo Fisher Scientific, Waltham, Massachusetts, USA) with 50% v/v glycerol (Sigma–Aldrich, St. Louis, Missouri, USA). A single pure colony of each strain was isolated from ALOA (Agar Listeria, Biolife, Milan, Italy) and inoculated in 5 ml TSB for 18h at 37°C. Genomic DNA was extracted by using the commercial kit Insvisorb® Spin Tissue Mini Kit (Invitek, STRATEC Biomedical, Birkenfeld, Germany) according to the manufacturer's instructions. Samples concentration and quality were assessed by using both a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and a Qubit® 2.0 Fluorometer (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA). The same genomic DNA was used as template for the MLST and 2b-RAD genotyping techniques. **i) MLST:** the target genes were the same as those described in the *in silico* MLST analysis (see section 2.1). PCRs were performed as described in (*Ragon et al. 2008*) by using an Applied Biosystems 2720 Thermal Cycler (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The templates were sent to Macrogen Inc. (Amsterdam, Netherlands) for direct Sanger sequencing with the respective primer pairs used for PCR amplification as sense and antisense sequencing primers. The entire sequences were checked for quality, edited, and aligned to obtain a sequence of the correct length for each locus. The visualisation, analysis and editing of the chromatograms obtained for the 7 genes were performed with *FinchTV 1.4.0* software (Geospiza). The sequences were confirmed by the alignment of both forward and reverse sequences using *ClustalW* (http://www.ebi.ac.uk). The sequence of each allele was compared to those available in the MLST *L. monocytogenes* database (http://www.pasteur.fr) by using blast similarities searches.

Strains with identical allelic profiles were identified with the same ST. **ii) 2b-RAD:** a total of 67 2b-RAD libraries were constructed by following the protocol reported by *Wang et al. 2012a* with some modifications. L18 and L30 were processed twice independently to produce two technical replicates (TRs), for L70 three TRs were obtained, while for ATCC19117 six TRs were prepared. DNA (300 ng) was digested in 6-µL reaction volume using 1.5 U *CspCI* (NEB New England Biolabs, Ipswich, Massachusetts, USA) or 1 U *AlfI* (Thermo Fisher Scientific, Waltham, Massachusetts, USA) at 37 °C for 1 h.

Once the digestion was completed, enzyme heat inactivation was conducted at 65 °C for 20 min. The ligation reaction was performed by combining 5 µL of digested DNA with 12 µL of a ligation master mix containing 0.4 µM each of two library-specific adaptors (with fully degenerate cohesive ends (5′-NN-3′), 0.2 mM ATP (NEB), and 1000 U T4 DNA ligase (CABRU, Arcore, Italy). Ligation was carried out at 16°C for 3 hours with subsequent heat inactivation for 10 minutes at 65°C. Sample-specific barcodes were designed through a Barcode Generator program (http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator) and introduced by PCR with platform-specific barcode-bearing primers. Each 50-µl PCR reaction contained 12 µl of ligated DNA product, 0.2 µM of each primer, 0.3 mM dNTP, 1× Phusion HF buffer and 1 U Phusion high-fidelity DNA polymerase (NEB). The PCR amplification was conducted under specific thermal cycling conditions: 12 cycles of 95 °C for 5 s, 60 °C for 20 s and 72 °C for 5 s. Adaptor and primer sequences were those reported in (*Wang et al. 2012a*). PCR products were purified by using the *SPRIselect purification kit* (Beckman Coulter, Pasadena, California, USA). After the final step of beads purification, samples were quantified through a Qubit® 2.0 Fluorometer (Invitrogen). The quality of the amplicon libraries was tested by running them on an Agilent 2100 Bioanalyzer. Finally, libraries were pooled and sequenced at the BMR Genomics facilities (Padova, Italy) by following a 96-plex 50SE strategy on a Miseq instrument (Illumina, San Diego, California, USA). Quality and adapters trimming of the sequenced reads was performed by running a customized script (File S1), thus obtaining 34-bp fragments ready to be evaluated for SNPs presence in STACKS (*Catchen et al. 2013*). As previously reported for *in silico* 2b-RAD analysis, individual genotypes were constructed using the *denovo_map.pl* STACKS pipeline (see section 2.1) by setting the following parameters: m=10, n=2, M=2.

To evaluate if polymorphic loci were located in accessory or core genes, tag sequences showing at least one SNP were mapped against the *L. monocytogenes* core and accessory genomic regions constructed as previously described. Since the sample dataset did not include lineage III isolates, STACKS catalog tags were mapped against core and accessory fragments assembled by using only

the 27 lineage I and II strains (Table 1). Details on genomes assembly and mapping parameters were reported in section 2.1.

Population structure analysis

SNP data (from MLST and 2b-RAD, both for *in silico* and laboratory experiments) were analysed with STRUCTURE and PHYLOViZ. All input files used in STRUCTURE and PHYLOViZ were deposited into the Dryad Data Repository (DOI:10.5061/dryad.m1bm5).

Cluster analyses were performed by using STRUCTURE v.2.3.4 (*Pritchard et al. 2000*). An admixture model and correlated allele frequency model were used to analyse the dataset without prior population information. At first, 20 runs of STRUCTURE were performed for each number of possible clusters (K = 1 to 10). The burn-in time was set to 50 000 samples, and the number of Monte Carlo Markov chains (MCMC) repeats after burn-in was set to 100 000 samples in each run. This procedure assigns a probability of ancestry for each polymorphic locus for a given number of groups, K, and also estimates q, the combined probability of ancestry from each of the K groups for each individual isolate. The most probable number of clusters was assessed using the online program Structure Harvester (*Earl & vonHoldt 2012*) according to both the L(K) and delta(K) methods (*Pritchard et al. 2000; Evanno et al. 2005*). STRUCTURE results were displayed with the software Clumpak (http://clumpak.tau.ac.il/).

Minimum Spanning Tree (MST) analysis was performed by using PHYLOViZ, a platform-independent JAVA software (*Francisco et al. 2012*), which allows the analysis of sequence-based typing methods and generates allelic profiles. As additional information, the presence/absence of tags was specified in the input file by adding a fifth state coded by "0". PHYLOViZ uses the goeBURST algorithm, a refinement of eBURST algorithm proposed by Feil and co-workers (*Feil et al. 2004*) and its expansion to generate a complete MST displaying the relationships between closely-related isolate and to identify potential clonal complexes and founders.

Data accessibility

Reads obtained by the Illumina sequencing of *AlfI*-digested fragments were deposited into the NCBI's Sequence Read Archive under the study accession number SRP058349. STRUCTURE/PHYLOViZ input files, STACKS output files and files required for the integrated analysis described in Figure S2 were deposited into the Dryad Data Repository (DOI:10.5061/dryad.m1bm5)

2.6 References

Baird NA, Etter PD, Atwood TS, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376.

Behrens S, Widder S, Mannala GK, et al. (2014) Ultra deep sequencing of *Listeria monocytogenes* sRNA transcriptome revealed new antisense RNAs. *PLoS One* 9:e83979.

Boers SA, van der Reijden WA, Jansen R (2012). High-throughput multilocus sequence typing: bringing molecular typing to the next level. *PLoS One* 7:e39630.

Cantinelli T, Chenal-Francisque V, Diancourt L, et al. (2013) "Epidemic clones" of *Listeria monocytogenes* are widespread and ancient clonal groups. *Journal of Clinical Microbiology* 51:3770-9.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22:3124-40.

Chen Y, Zhang W, Knabel ST (2007) Multi-virulence-locus sequence typing identifies single nucleotide polymorphisms which differentiate epidemic clones and outbreaks trains of *Listeria monocytogenes*. *Journal of Clinical Microbiology* 45:835-46.

Chenal-Francisque V, Diancourt L, Cantinelli T, et al. (2013) Optimized Multilocus variable-number tandem-repeat analysis assay and its complementarity with pulsed-field gel electrophoresis and multilocus sequence typing for *Listeria monocytogenes* clone identification and surveillance. *Journal of Clinical Microbiology* 51:1868-80.

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, et al. (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics* 12:4.

Cromie GA, Hyma KE, Ludlow CL, et al. (2103) Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *Genes, Genomes, Genetics* 3:2163-71.

Davey JW, Hohenlohe PA, Etter PD, et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12:499-510.

den Bakker HC, Desjardins CA, Griggs AD, et al. (2013) Evolutionary dynamics of the accessory genome of *Listeria monocytogenes*. *PLoS One* 8:e67511.

den Bakker HC, Didelot X, Fortes ED, Nightingale KK, Wiedmann M (2008) Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Developmental Biology* 8:277.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359-61.

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards) (2014) Scientific opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: part 2 (surveillance and data management activities). *EFSA Journal* 12:3784.

Emerson KJ, Merz CR, Catchen JM, et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *PNAS* 107:16196-200.

Eusebio N, Coutinho CP, Sá-Correia I, Araujo R (2013a) SNaPBcen: a novel and practical tool for genotyping *Burkholderia cenocepacia*. *Journal of Clinical Microbiology* 51:2646-53.

Eusebio N, Pinheiro T, Amorim AA, et al. (2013b) SNaPaer: a practical single nucleotide polymorphism multiplex assay for genotyping of *Pseudomonas aeruginosa*. *PLoS One* 8:e66083.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-20.

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology* 186:1518-30.

Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ (2014) *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *Journal of Food Protection* 77:150-70.

Francisco AP, Vaz C, Monteiro PT, et al. (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 13:87.

Fuchs TM, Eisenreich W, Kern T, Dandekar T (2012) Toward a Systemic Understanding of *Listeria monocytogenes* Metabolism during Infection. *Frontiers in Microbiology* 3:23.

Guo Y, Yuan H, Fang D, et al. (2014) An improved 2b-RAD approach (I2b-RAD) offering genotyping tested by a rice (*Oryza sativa* L.) F2 population. *BMC Genomics* 15:956.

Haase JK, Didelot X, Lecuit M, et al. (2014) The ubiquitous nature of *Listeria monocytogenes* clones: a large-scale Multilocus Sequence Typing study. *Environmental Microbiology 16:405-16.*

Hain T, Ghai R, Billion A, et al. (2012) Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*. *BMC Genomics* 13:144.

Hein I, Klinger S, Dooms M, et al. (2011) Stress survival islet 1 (SSI-1) survey in *Listeria monocytogenes* reveals an insert common to *Listeria innocua* in sequence type 121 *L. monocytogenes* strains. *Applied Environmental Microbiology* 77:2169-73.

Hohenlohe PA, Bassham S, Etter PD, et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6:e1000862.

Jadhav S, Bhave M, Palombo EA (2012) Methods used for the detection and subtyping of *Listeria monocytogenes*. *Journal of Microbiological Methods* 88:327-41.

Jiang L, Chen J, Xu J, et al. (2008) Virulence characterization and genotypic analyses of *Listeria monocytogenes* isolates from food and processing environments in eastern China. *International Journal of Food Microbiology* 121:53-9.

Kuenne C, Billion A, Mraheil MA, et al. (2013) Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* 14:47.

Laing C, Buchanan C, Taboada EN, et al. (2010) Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* 11:461.

Larsen MV, Cosentino S, Rasmussen S, et al. (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of Clinical Microbiology* 50:1355-61.

Loman NJ, Constantinidou C, Chan JZ, et al. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology* 10:599-606.

Lomonaco S, Knabel SJ, Dalmasso A, Civera T, Bottero MT (2011) Novel multiplex single nucleotide polymorphism-based method for identifying epidemic clones of *Listeria monocytogenes*. *Applied Environmental Microbiology* 77:6290-4.

Lomonaco S, Nucera D, Filipello V (2015) The evolution and epidemiology of *Listeria monocytogenes* in Europe and the United States. *Infection, Genetics and Evolution* 35:172-183.

Maiden MC (2006) Multilocus sequence typing of bacteria. *Annual Review of Microbiology* 60:561-88.

Mastretta-Yanes A, Arrigo N, Alvarez N, et al. (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* 15:28-41.

Mellin JR, Cossart P (2012) The non-coding RNA world of the bacterial pathogen *Listeria monocytogenes*. *RNA Biology* 9:372-8.

Nightingale KK, Windham K, Wiedmann M (2005) Evolution and molecular phylogeny of *Listeria monocytogenes* isolated from human and animal listeriosis cases and foods. *Journal of Bacteriology* 187:5537-51.

Orsi RH, den Bakker HC, Wiedmann M (2011) *Listeria monocytogenes* lineages: Genomics, evolution, ecology, and phenotypic characteristics. *International Journal of Medical Microbiology* 301:79-96.

Parisi A, Latorre L, Normanno G, et al. (2010) Amplified Fragment Length Polymorphism and Multi-Locus Sequence Typing for high-resolution genotyping of *Listeria monocytogenes* from foods and the environment. *Food Microbiology* 27:101-8.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics 1*55:945-59.

Pujolar JM, Jacobsen MW, Frydenberg J, et al. (2013) A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources* 13:706-14.

Ragon M, Wirth T, Hollandt F, et al. (2008) A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathogen* 4:e1000146.

Rückerl I, Muhterem-Uyar M, Muri-Klinger S, et al. (2014) *L. monocytogenes* in a cheese processing facility: Learning from contamination scenarios over three years of sampling. *International Journal of Food Microbiology* 189:98-105.

Salcedo C, Arreaza L, Alcalá B, de la Fuente L, Vázquez JA (2003) Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *Journal of Clinical Microbiology* 41:757-62.

Sauders BD, Overdevest J, Fortes E, et al. (2012). Diversity of Listeria species in urban and natural environments. *Applied Environmental Microbiology* 78:4420-33.

Seetharam AS, Stuart GW (2013) Whole genome phylogeny for 21 Drosophila species using predicted 2b-RAD fragments. *PeerJ* 1:e226.

Sesto N, Touchon M, Andrade JM, et al. (2014) A PNPase dependent CRISPR System in Listeria. *PLoS Genetics* 10:e1004065.

Singh P, Foley SL, Nayak R, Kwon YM (2012) Multilocus sequence typing of Salmonella strains by high-throughput sequencing of selectively amplified target genes. *Journal of Microbiological Methods* 88:127-33.

Trembizki E, Smith H, Lahra MM, et al. (2014) High-throughput informative single nucleotide polymorphism-based typing of *Neisseria gonorrhoeae* using the Sequenom MassARRAY iPLEX platform. *Journal of Antimicrobial Chemotherapy* 69:1526-32.

Tsai YH, Maron SB, McGann P, et al. (2011) Recombinationand positive selection contributed to the evolution of *Listeria monocytogenes* lineages III and IV, two distinct and well supported uncommon L. monocytogenes lineages. *Infection, Genetics and Evolution* 11:1881-90.

Turton JF, Baddal B, Perry C (2011) Use of the accessory genome for characterization and typing of *Acinetobacter baumannii*. *Journal of Clinical Microbiology* 49:1260-6.

Wang S, Meyer E, McKay JK, Matz MV (2012a) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods 9:808-10.

Wang Y, Zhao A, Zhu R, et al. (2012b) Genetic diversity and molecular typing of *Listeria monocytogenes* in China. *BMC Microbiology* 12:119.

Ward TJ, Ducey TF, Usgaard T, Dunn KA, Bielawski JP (2008) Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates. *Applied Environmental Microbiology* 74:7629-42.

Zhang W, Jayarao BM, Knabel SJ (2004) Multi-virulence-locus sequence typing of *Listeria monocytogenes*. *Applied Environmental Microbiology* 70:913-20.

Zhao F, Chaloner G, Darby A, et al. (2012) Optimization of *Bartonella henselae* multilocus sequence typing scheme using single-nucleotide polymorphism analysis of SOLiD sequence data. *Chinese Medical Journal* 125:2284-8.

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary Figure S1.

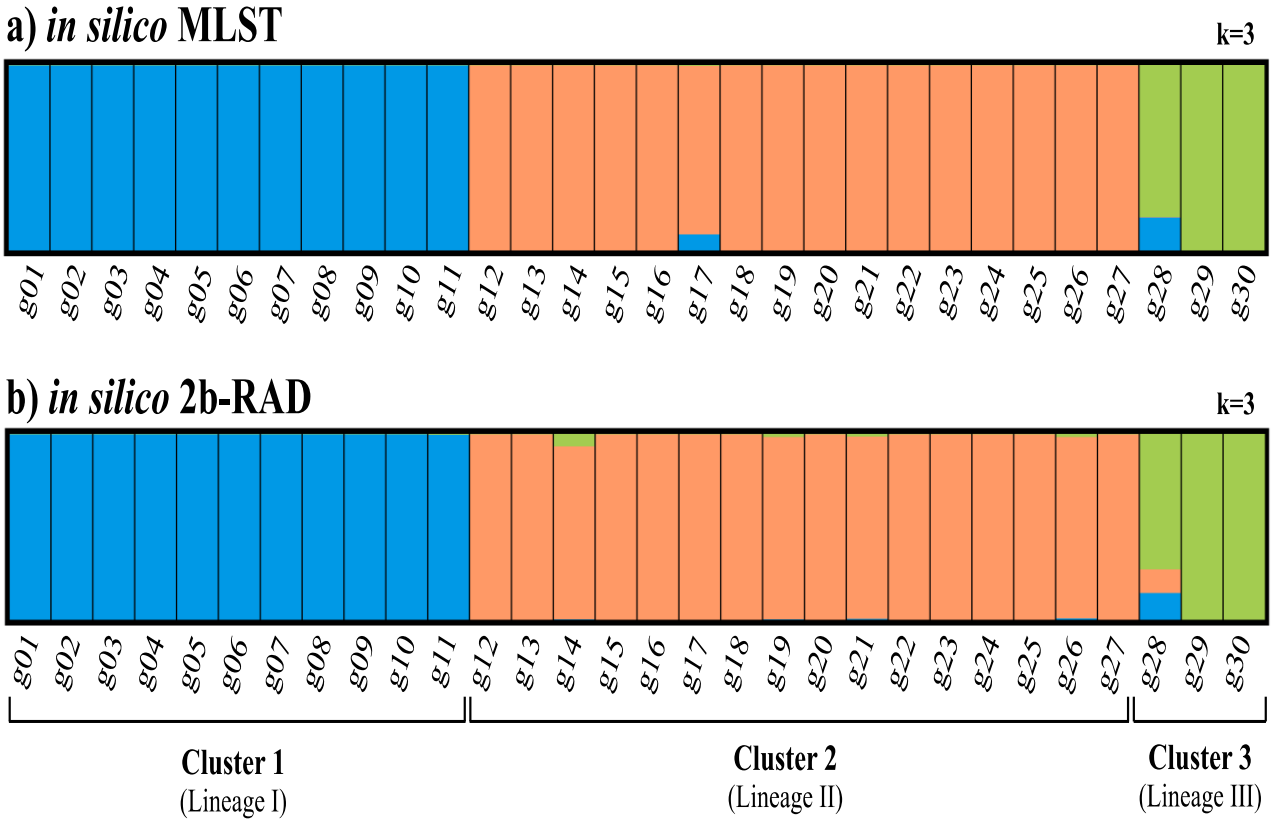


Figure S1. STRUCTURE bar-plot - *in silico* analysis. Estimated population structure generated from MLST **(a)** and 2b-RAD data **(b)**. Proportions of ancestry from ancestral lineage I (blue), ancestral lineage II (orange), ancestral lineage III (green) as inferred by STRUCTURE assuming K = 3 ancestral subpopulations. Each vertical bar represents an individual strain identified by the ID introduced in Table 1.

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary Figure S2.

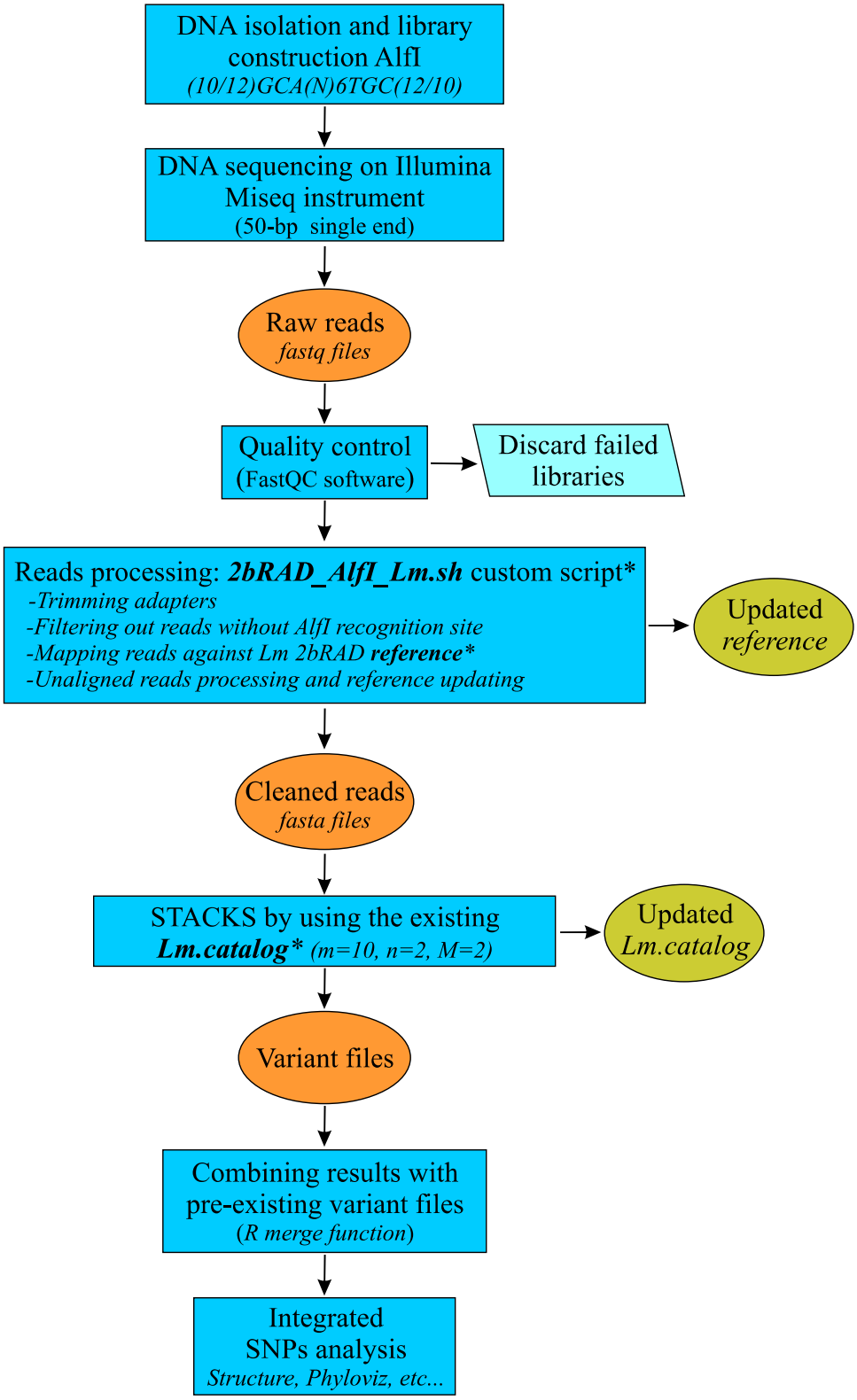


Figure S2. *L. monocytogenes* integrated analysis workflow. Schematic representation of the *L. monocytogenes* 2b-RAD protocol from DNA isolation to SNPs analysis.
**available at the Dryad database under the accession number 10.5061/dryad.m1bm5.*

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary Table S1.

Table S1. Origin of the 56 field strains.							
isolate	ST	taxonomic_designation	isolation_year	world_region	country	source_type	source_description
L4	155	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L5	121	Listeria monocytogenes	2010	Europe	Italy	Food product	cheese
L6	121	Listeria monocytogenes	2010	Europe	Italy	Food product	cheese
L7	4	Listeria monocytogenes	2010	Europe	Italy	Food product	poultry
L8	121	Listeria monocytogenes	2010	Europe	Italy	Food product	ready-to-eat product
L9	121	Listeria monocytogenes	2010	Europe	Italy	Food product	food
L10	121	Listeria monocytogenes	2010	Europe	Italy	Food product	food
L11	121	Listeria monocytogenes	2010	Europe	Italy	Food product	meat product
L16	121	Listeria monocytogenes	2010	Europe	Italy	Food product	bakery product
L17	121	Listeria monocytogenes	2010	Europe	Italy	Food product	cheese
L18	121	Listeria monocytogenes	2010	Europe	Italy	Food product	cheese
L20	320	Listeria monocytogenes	2010	Europe	Italy	Production environment	food processing plant
L23	121	Listeria monocytogenes	2010	Europe	Italy	Production environment	fish processing plant
L24	121	Listeria monocytogenes	2010	Europe	Italy	Production environment	fish processing plant
L25	204	Listeria monocytogenes	2010	Europe	Italy	Production environment	fish processing plant
L26	121	Listeria monocytogenes	2010	Europe	Italy	Production environment	meat processing plant
L27	204	Listeria monocytogenes	2010	Europe	Italy	Production environment	meat processing plant
L28	121	Listeria monocytogenes	2010	Europe	Italy	Production environment	fish processing plant
L29	204	Listeria monocytogenes	2010	Europe	Italy	Production environment	fish processing plant
L30	4	Listeria monocytogenes	2010	Europe	Italy	Production environment	fish processing plant
L31	9	Listeria monocytogenes	2010	Europe	Italy	Production environment	meat processing plant
L32	6	Listeria monocytogenes	2010	Europe	Italy	Food product	milk
L34	37	Listeria monocytogenes	2010	Europe	Italy	Food product	dairy product
L35	6	Listeria monocytogenes	2010	Europe	Italy	Food product	dairy product
L36	6	Listeria monocytogenes	2010	Europe	Italy	Food product	dairy product
L37	6	Listeria monocytogenes	2010	Europe	Italy	Food product	dairy product
L38	4	Listeria monocytogenes	2010	Europe	Italy	Food product	dairy product
L39	26	Listeria monocytogenes	2010	Europe	Italy	Food product	dairy product
L41	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L42	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L43	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L44	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L45	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L46	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L47	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L48	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L49	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L50	121	Listeria monocytogenes	2010	Europe	Italy	Food product	seafood
L52	9	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L53	9	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L54	9	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L55	9	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L56	14	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L57	399	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L58	37	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L59	155	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L60	217	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L61	9	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L62	37	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L63	5	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L64	37	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L66	9	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L67	18	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L68	580	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L69	18	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product
L70	18	Listeria monocytogenes	2012	Europe	Italy	Food product	meat product

Table S1. Origin of the 56 field strains.

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary Table S2.

In silico MLST analysis									
ID	*ST*	*abc*	*bgl*	*cat*	*dap*	*dat*	*ldh*	*lhk*	Lineage
g01	1	3	1	1	1	3	1	3	I
g02	1	3	1	1	1	3	1	3	I
g03*	2	1	1	11	11	2	1	5	I
g04	2	1	1	11	11	2	1	5	I
g05	3	4	4	4	3	2	1	5	I
g06	4	1	2	12	3	2	5	3	I
g07	4	1	2	12	3	2	5	3	I
g08	4	1	2	12	3	2	5	3	I
g09	66	4	4	4	3	2	28	5	I
g10	73	3	1	1	1	3	11	3	I
g11	617	2	9	12	14	3	247	36	I
g12	9	6	5	6	4	1	4	1	II
g13	9	6	5	6	4	1	4	1	II
g14	11	7	6	10	6	1	2	1	II
g15	12	5	8	5	7	6	22	1	II
g16	12	5	8	5	7	6	22	1	II
g17	18	7	6	15	18	12	6	1	II
g18	35	6	5	6	20	1	4	1	II
g19	37	5	7	3	5	1	8	6	II
g20	85	5	8	5	7	6	38	1	II
g21	91	7	6	15	6	5	2	1	II
g22	120	5	6	2	29	5	3	1	II
g23	121	7	6	8	8	6	37	1	II
g24	121	7	6	8	8	6	37	1	II
g25	122	6	5	6	4	1	62	1	II
g26	155	7	10	16	7	5	2	1	II
g27	292	57	6	2	29	5	3	1	II
g28	71	18	11	21	24	17	35	13	III
g29	201	19	17	22	25	15	79	12	III
g30	201	19	17	22	25	15	79	12	III
Laboratory MLST analysis									
ID	*ST*	*abc*	*bgl*	*cat*	*dap*	*dat*	*ldh*	*lhk*	Lineage
ATCC19117	449	3	1	1	1	18	28	3	I
L4	155	7	10	16	7	5	2	1	II
L5	121	7	6	8	8	6	37	1	II
L6	121	7	6	8	8	6	37	1	II
L7	4	1	2	12	3	2	5	3	I
L8	121	7	6	8	8	6	37	1	II
L9	121	7	6	8	8	6	37	1	II
L10	121	7	6	8	8	6	37	1	II
L11	121	7	6	8	8	6	37	1	II
L16	121	7	6	8	8	6	37	1	II
L17	121	7	6	8	8	6	37	1	II
L18	121	7	6	8	8	6	37	1	II
L20	320	7	10	16	7	35	2	1	II
L23	121	7	6	8	8	6	37	1	II
L24	121	7	6	8	8	6	37	1	II
L25	204	5	7	6	4	5	4	1	II
L26	121	7	6	8	8	6	37	1	II
L27	204	5	7	6	4	5	4	1	II
L28	121	7	6	8	8	6	37	1	II
L29	204	5	7	6	4	5	4	1	II
L30	4	1	2	12	3	2	5	3	I
L31	9	6	5	6	4	1	4	1	II
L32	6	3	9	9	3	3	1	5	I
L34	37	5	7	3	5	1	8	6	II
L35	6	3	9	9	3	3	1	5	I
L36	6	3	9	9	3	3	1	5	I
L37	6	3	9	9	3	3	1	5	I
L38	4	1	2	12	3	2	5	3	I
L39	26	5	10	8	21	6	2	1	II
L41	121	7	6	8	8	6	37	1	II
L42	121	7	6	8	8	6	37	1	II
L43	121	7	6	8	8	6	37	1	II
L44	121	7	6	8	8	6	37	1	II
L45	121	7	6	8	8	6	37	1	II
L46	121	7	6	8	8	6	37	1	II
L47	121	7	6	8	8	6	37	1	II
L48	121	7	6	8	8	6	37	1	II
L49	121	7	6	8	8	6	37	1	II
L50	121	7	6	8	8	6	37	1	II
L52	9	6	5	6	4	1	4	1	II
L53	9	6	5	6	4	1	4	1	II
L54	9	6	5	6	4	1	4	1	II
L55	9	6	5	6	4	1	4	1	II
L56	14	8	6	13	6	5	2	1	II
L57	399	8	6	15	6	6	2	1	II
L58	37	5	7	3	5	1	8	6	II
L59	155	7	10	16	7	5	2	1	II
L60	217	1	9	12	3	7	5	3	I
L61	9	6	5	6	4	1	4	1	II
L62	37	5	7	3	5	1	8	6	II
L63	5	2	1	11	3	3	1	7	I
L64	37	5	7	3	5	1	8	6	II
L66	9	6	5	6	4	1	4	1	II
L67	18	7	6	15	18	12	6	1	II
L68	580	6	37	6	4	1	4	1	II
L69	18	7	6	15	18	12	6	1	II
L70	18	7	6	15	18	12	6	1	II
ScottA	2	1	1	11	11	2	1	5	I

Table S2. MLST. Allelic profiles and STs achieved by both in silico and laboratory analyses. Each strain is labelled with a unique ID. Information about lineage (retrieved from MLST database) is also provided.

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary Table S3.

Strain	N. of predicted *AlfI* fragments	N. of predicted *CspCI* fragments
g01	372	295
g02	372	294
g03	370	297
g04	377	301
g05	375	293
g06	368	295
g07	372	293
g08	371	290
g09	378	302
g10	373	298
g11	379	306
g12	405	287
g13	401	288
g14	397	309
g15	397	302
g16	398	302
g17	398	289
g18	404	286
g19	418	297
g20	395	296
g21	386	291
g22	413	294
g23	379	274
g24	377	272
g25	401	289
g26	391	293
g27	418	296
g28	364	278
g29	397	294
g30	399	294

Table S3. In silico digestion. Number of predicted AlfI and CspCI fragments obtained through the in silico digestion of the 30 target genomes.

Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary Table S4.

Strain	Unique fragments	Structure cluster
ATCC19117_TR1	337	1
ATCC19117_TR2	345	1
ATCC19117_TR3	335	1
ATCC19117_TR4	334	1
ATCC19117_TR5	329	1
ATCC19117_TR6	331	1
L4	376	3
L5	378	2
L6	390	2
L7	334	1
L8	347	2
L9	360	2
L10	389	2
L11	375	2
L16	378	2
L17	384	2
L18_TR1	394	2
L18_TR2	382	2
L20	365	3
L23	377	2
L24	368	2
L25	359	3
L26	388	2
L27	375	3
L28	388	2
L29	366	3
L30_TR1	354	1
L30_TR2	351	1
L31	384	3
L32	335	1
L34	333	3
L35	339	1
L36	345	1
L37	350	1
L38	331	1
L39	338	3
L41	370	2
L42	367	2
L43	370	2
L44	372	2
L45	357	2
L46	366	2
L47	372	2
L48	357	2
L49	363	2
L50	381	2
L52	370	3
L53	368	3
L54	372	3
L55	387	3
L56	382	3
L57	359	2
L58	376	3
L59	350	3
L60	333	1
L61	379	3
L62	377	3
L63	374	1
L64	374	3
L66	383	3
L67	378	3
L68	391	3
L69	376	3
L70_TR1	369	3
L70_TR2	370	3
L70_TR3	377	3
ScottA	332	1

Table S4. 2b-RAD unique stacks. Number of unique fragments obtained through AlfI digestion. For each strain, also the Structure cluster membership was reported.
Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing (MLST) and 2b-RAD to investigate *Listeria monocytogenes* genetic structure.
Supplementary File S1.

```
#!/bin/bash
##############################################################################################################################################################
# This script is intended to process 2b-RAD raw reads obtained by AlfI DNA digestion. Fastq format is required as input. The final output are fasta files  ready for STACKS
# pipeline. This script require the following free tools: "TruncateFastq.pl" & "2b_Extract.pl" (http://people.oregonstate.edu/~meyere/tools.html), "fnafile" (NEWBLERtools),"SHRiMP
# 2.2.3" (http://research-pub.gene.com/gmap/) ,"revcompl.pl" (https://github.com/KorfLab/Perl_utils) and CD-HIT-EST(http://weizhongli-lab.org/cd-hit/) .
##############################################################################################################################################################
a=elab
b=Trimmed
c=stacks_input
d=gmapper_output
e=Reference
#f=cdhit
g=Ref_Samples
c1=1                        #var -c first cd-hit
c2=0.9                      #var -n first cd-hit
n1=8                    #var -c second cd-hit
n2=8                    #var -n second cd-hit
site=".{11}GCA.{6}TGC.{11}" #recognize site "ALfI"
test -d $a || mkdir -p $a
test -d $b || mkdir -p $b
test -d $c || mkdir -p $c
test -d $d || mkdir -p $d
test -d $e || mkdir -p $e
#test -d $f || mkdir -p $f
test -d $g || mkdir -p $g
for file in *.fastq;
do
#Truncating all reads at nucleotide 36.
cat $file | TruncateFastq.pl $file 1 36 ./elab/$file && cat ./elab/$file | perl -lane '$i++; if($i%4==1){s/^@/>/; print;} elsif($i%4==2){print;}'> ./elab/T_$file.fasta
done
rm ./elab/*.fastq
rename fastq.fasta fasta ./elab/*fastq.fasta       #RedHat
#rename 's/\.fastq//' ./elab/*fastq.fasta      #ubuntu
for file in ./elab/*.fasta;
do
cat $file | sed 's/\s.*//'| 2b_Extract_mod.pl $file $site $file.trimm && cp $file.trimm ./Trimmed;
done
rm ./elab/*.*
rename fasta.trimm fasta ./Trimmed/*fasta.trimm           #RedHat
#rename 's/\.trimm//' ./Trimmed/*fasta.trimm              #Ubuntu
cd ./Trimmed/
for file in *.fasta;
do
cat $file | cd-hit-est -i $file -c $c1 -n $n1 -T 0 -M 0 -d 0 -o res;
mv res ../Reference/res_$file
rm res.clstr
done
cd ..;cp ./Reference/*.* ./Ref_Samples/;cat ./Reference/*.* > ./Reference/Ref_singl.fasta;rm ./Reference/res_* *.clstr;
file=./Reference/Ref_singl.fasta;
cat $file | cd-hit-est -i $file -o ./Reference/reference.fasta -c $c2 -n $n2 -T 0 -M 0 -d 0;
#gmapper mapping
for file in ./Trimmed/*fasta;
do
cat $file | gmapper $file ./Reference/reference.fasta -N 20 --all-contigs --single-best-mapping -E > $file.out && cat $file.out | grep -e "*" | sed -e 's/^\(.\{50\}\).*$/\1/' | grep -P "\t0" |
sed 's/\s.*//' > $file.plus && fnafile -i $file.plus $file && rename Reads.fna $file.seqplus Reads.fna && cat $file.out | grep -e "*" | sed -e 's/^\(.\{50\}\).*$/\1/' | grep -P "\t16" |
sed 's/\s.*//'> $file.minus && fnafile -i $file.minus $file && rename Reads.fna $file.seqminus Reads.fna && revcompl.pl $file.seqminus $file.seqminusrev && cat $file.seqminusrev $file.seqplus > $file.stacks
done
mv ./Trimmed/*.stacks ./stacks_input
mv ./Trimmed/*.out ./gmapper_output
rm ./Trimmed/*.minus
rm ./Trimmed/*.seqminusrev;rm ./Trimmed/*.plus;rm ./Trimmed/*.seqminus;rm ./Trimmed/*.seqplus
rm -R ./elab
rm -R gmapper_output
rename fasta.stacks fasta ./stacks_input/*fasta.stacks;
exit
```

File S1. Script for 2b-RAD raw reads processing.

Publication III

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Massimiliano Babbucci[1*], Serena Ferraresso[1], Marianna Pauletto[1], Rafaella Franch[1], Chiara Papetti[2], Tomaso Patarnello[1], Paolo Carnier[1], Luca Bargelloni[1].

Abstract

Skeletal anomalies in farmed fish are a relevant issue affecting animal welfare and health and causing significant economic losses. Here, a high-density genetic map of European seabass for QTL mapping of jaw deformity was constructed and a genome-wide association study (GWAS) was carried out on a total of 298 juveniles, 148 of which belonged to four full-sib families. Out of 298 fish, 107 were affected by mandibular prognathism (MP). Three significant QTLs and two candidate SNPs associated with MP were identified. The two GWAS candidate markers were located on ChrX and Chr17, both in close proximity with the peaks of the two most significant QTLs. Notably, the SNP marker on Chr17 was positioned within the *Sobp* gene coding region, which plays a pivotal role in craniofacial development. The analysis of differentially expressed genes in jaw-deformed animals highlighted the "nervous system development" as a crucial pathway in MP. In particular, *Zic2*, a key gene for craniofacial morphogenesis in model species, was significantly down-regulated in MP-affected animals. Gene expression data revealed also a significant down-regulation of *Sobp* in deformed larvae. Our analyses, integrating transcriptomic and GWA methods, provide evidence for putative mechanisms underlying seabass jaw deformity.

Introduction

The European seabass (*Dicentrarchus labrax*) is one of the first non-salmonid species in marine aquaculture, with an annual production greater than 155,000 tons in 2014 [1]. Despite great progress, several problems still limit the performance of sea bass farming, such as infectious diseases and larval and juvenile mortalities, often linked to anomalies in development. In the present paper, the focus is on a developmental malformation, mandibular prognathism (MP), in which the lower jaw is longer than the upper one providing a very distinctive underbite facial phenotype. In intensively reared European seabass, MP is a sporadic event and despite an incidence of 11% was reported in literature [2], to our best knowledge seabass farms rarely experience incidences higher than 2%. In general, skeletal anomalies in farmed fish are a relevant problem entailing economic as well as animal health and welfare issues. For instance, in species that are mainly marketed as whole-fish, such as seabass, an anomalous external morphology could substantially impact the consumers' overall perception of the product. Thus, fish need to be checked frequently and the deformed ones manually removed, even at large-scale production sites. Such defective fish are generally downgraded with consequent loss of profit.

Several external variables (*e. g.* nutritional unbalances, toxicants in water, mechanical stress, high water temperature) have been proposed as causative factors for skull and spinal malformations in

cultured fish [3–5]. In European seabass, for example, it was shown that both temperature and different water-current speed had a significant effect on the incidence of lordosis, a severe vertebral deformity [6]. It is seemingly established that if a genetic base for skeletal anomalies is observed, this phenotype is expressed only when exceptional environmental conditions occur (i.e intensive farming) [4].

With regard to jaw anomalies, it has been suggested that they develop mainly during the early larval stages [3,4,7]. In farmed seabass, skull and skeletal malformations were shown to be associated with excess of vitamin A [8] and with low phospholipids [8,9] in the diet, although MP was not reported having higher frequency under such dietary conditions. Genetic factors also appear to be involved in fish skeletal deformities, as quantitative genetic studies in different species documented additive genetic variation underlying skeletal anomalies [10] (and references therein). In the European seabass, the genetic basis of skeletal malformations was evaluated across different on-growing sites using the same genetic material [11]. Moderate heritability was observed for spine anomalies with varying incidence at different locations, although with negligible GxE interactions. Skull deformities were not reported and no data is available on genetic factors influencing jaw deformities in seabass. On the other hand, a genetic component for lower jaw anomalies has been proposed for other vertebrate species. In humans, prognathism typically shows familial aggregation and it is considered a multifactorial and polygenic trait likely explained by a threshold model [12]. Notably, it has been recently demonstrated that mutations at specific loci might be associated with MP. For instance, genes encoding matrilin 1 (MATN1) and fibroblast growth factor (FGF23) were proposed as candidate loci for MP [13,14]. In other mammalian species (dog, cattle, ass, horse) as well, lower jaw protrusion is considered a genetic disorder (Online Mendelian Inheritance in Animals database, http://omia.angis.org.au/home/), although only recently a genome-wide association study (GWAS) on prognathous horses has identified a candidate genomic region on equine chromosome 13 (ECA13, [15]).

	Dam	
Sire	R41	R26
R45	97	-
R25	34	-
R12	17	-
R17	-	24
Total	148	24

Table 1. Family structure with number of offspring per half-sib family and per full-sib family.

The present study was prompted by the observation of two batches of seabass larvae showing high incidence of prognathism (> 40%) at a commercial hatchery. As such batches experienced identical environmental conditions as several other batches of similar age that were minimally affected by the same malformation (< 2%), a genetic predisposition in prognathous fish was suspected. In fact, each batch originated from fertilized eggs from single mass-spawning events, which are known to be dominated by a single dam and few sires [16].

Samples for gene expression analysis of deformed and normal larvae from the affected batches were collected as soon as the malformation was evident (38 days post hatching, dph). The same sampling was repeated on early juveniles at 58 dph. Finally, at 79 - 83 dph individual fish were sampled, photographed, and fin-clips collected for genetic analysis. Here, we report on the results of the integrated analysis of prognathism in seabass based on transcriptome profiling and GWAS of the collected fish.

Despite the incidence of MP in farmed seabass is relatively low compared to other skeletal malformations in teleost fish [2,4], this case-study represents an example of how an integrated genomic approach might identify the molecular mechanisms underlying developmental anomalies and provide genetic tools to potentially mitigate the damage.

Results and discussion

Family-based genetic analysis

Microsatellite-based parentage assignment identified four main medium-sized full-sib (FS) families originated from four sires and two dams (Table 1) and a large number of smaller ones. For the four FS families, 2bRAD sequencing was carried out also on parental DNAs, in order to construct a high-density linkage map and to search for QTLs associated with jaw deformity. The sequencing data obtained in our study was deposited in the NCBI-short read archive (SRA) database under the accession number SRP076258. A catalogue containing 7,390 loci was constructed with only data from parents, and used as reference for SNP discovery and genotyping to map families. A total of 5,304 SNP markers were identified and genotyped in more than 80% of the progeny. Genotyped SNPs were used to build a linkage map, after removing markers with distorted segregation. The number of informative SNPs in the mapping panel was 3,266. These markers were distributed over 24 Linkage Groups (LGs) in a sex-averaged linkage map, using a LOD = 9 as threshold for mapping data (Fig. 1 and Table 2). The total genetic length of the map was 2,787 cM. The genetic length of individual LGs ranged from 96 cM for LG19, containing 116 markers, to 147 cM for LG02, containing 171 markers, with an average of 116.12 cM. SNP markers of each LG were mapped

against the genome of European seabass. As reported in Table 2, all LGs had a perfect match on a single chromosome of the seabass genome, except for LG21, which matched against two distinct chromosomes (Chr3 and Chr14). This discrepancy may be due to a not well-resolved LG21 or to a misassembly issue involving Chr03/Chr14. Notably, a total of 312 markers ranging from 3 (LG17) to 38 (LG6) mapped against ChrUN, which includes contigs not yet confidently placed on a specific chromosome.

Recently, a RAD-based linkage map using a mapping panel of 175 offspring that originated from a factorial cross between two dams and four sires from a single full-sib family was reported for the European seabass [17], with 6,706 SNPs clustered in 24 LGs, and a total length of 4,816 cM. Both maps represent a substantial improvement over the previous linkage maps, which were based on microsatellite, AFLPs, and a limited number of SNP loci. Future integration of the two sets could help refine the genome assembly and provide ordered markers for genetic studies in seabass. In the present study, the construction of a linkage map was instrumental to QTL mapping of jaw deformity. The maternal half-sib regression analysis identified a total of 18 QTLs significant either at genome-wide or chromosome-wide level (Table 3). Three QTLs, located on three different LGs, were genome-wide significant. The most significant one belongs to LG18 ($P<0.01$), which corresponds to seabass Chr17 and explained 13.21% of total phenotypic variation.

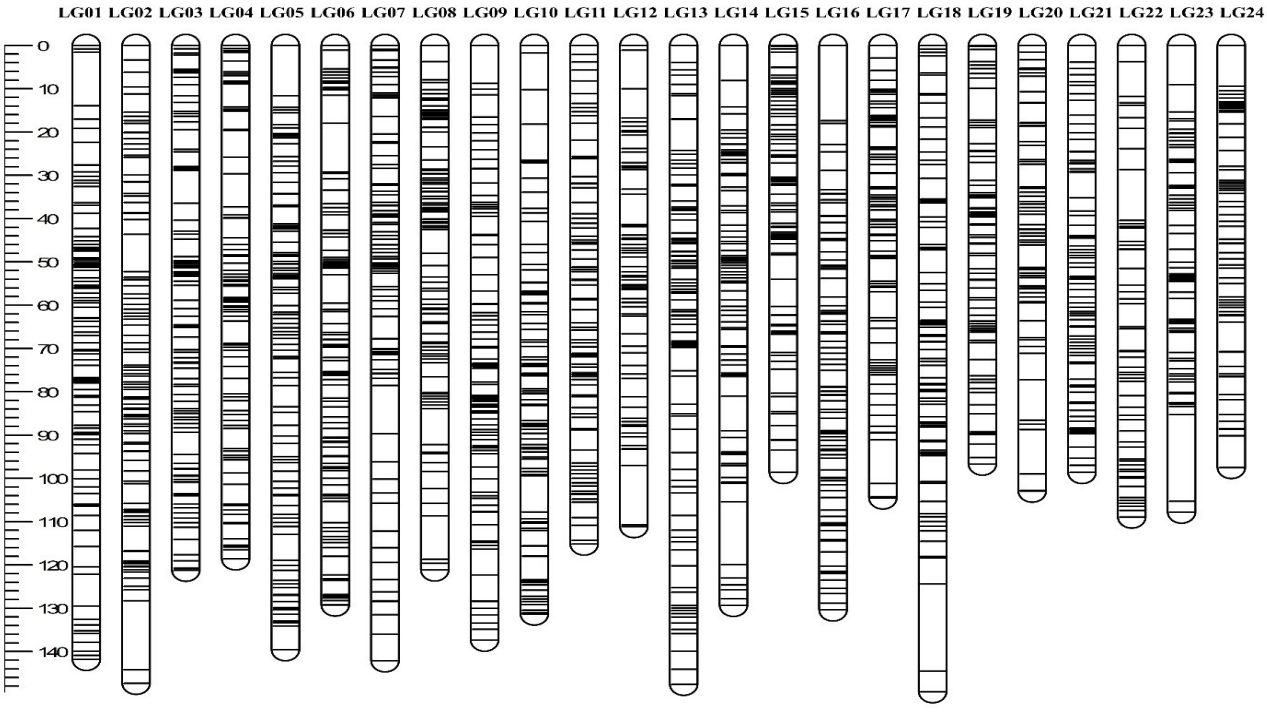


Figure 1. Genetic lengths and marker distribution of 24 linkage groups (LGs) in the sex-averaged linkage map of the European seabass.

LG	Sex-averaged map			
	Mapped Markers	Genetic length (cM)	Marker interval (cM)	Physical position
1	171	141	0.83	Chr16
2	171	147	0.86	Chr13
3	158	121	1.32	Chr05
4	148	118	0.80	Chr20
5	157	139	0.89	Chr06
6	162	129	0.80	Chr1B
7	147	142	0.97	Chr1A
8	137	121	0.89	Chr04
9	147	137	0.94	Chr02
10	145	131	0.91	Chr08
11	138	115	0.84	Chr10
12	131	111	0.85	Chr07
13	147	147	1.01	Chr15
14	140	129	0.93	Chr14
15	128	98	0.77	Chr (22-25)
16	134	130	0.98	Chr09
17	125	104	0.84	Chr11
18	130	149	1.15	Chr17
19	116	96	0.84	Chr19
20	108	102	0.96	Chr12
21	126	98	1.60	Chr03/Chr14
22	101	108	1.08	ChrX
23	107	107	1.01	Chr (18-21)
24	92	97	1.07	Chr24

Table 2. Summary statistics of the sex-averaged genetic map of European seabass. Physical position of the linkage groups is referred to the European seabass genome

A second significant QTL was identified on LG22 ($P<0.05$), matching ChrX, which explained 11.54% of total phenotypic variation. The third significant QTL was located on LG20 ($P<0.05$), corresponding to Chr12 and accounting for 11.33% of the total phenotypic variation.

The same analysis, performed for trait "total length" (TL) highlighted eight significant QTLs (Table S1), four of them at 5% genome-wide level of significance. None of them overlapped with the MP-associated QTLs, thus putatively excluding, at genetic level, a correlation between both traits.

In addition, no significant coefficient of correlation was revealed between MP and TL at phenotypic level ($r_{pb} = -0.0573$, p-value = 0.33).

QTL	LG	Position (cM)	Chr	F	Expl. Variation
MHS-01****	18	84	17	23.30	13.21%
MHS-02***	20	44	12	19.57	11.33%
MHS-03***	22	74	X	19.96	11.54%
MHS-04**	24	23	24	16.68	3.24%
MHS-05**	19	58	19	16.60	6.65%
MHS-06**	14	71	14	15.94	3.60%
MHS-07**	8	38	4	15.25	3.54%
MHS-08**	16	6	9	15.22	5.61%
MHS-09**	15	32	(22-25)	14.56	3.48%
MHS-10*	6	106	1B	14.37	2.47%
MHS-11**	17	12	11	13.36	3.38%
MHS-12**	5	79	6	12.90	3.33%
MHS-13*	4	111	20	11.71	2.21%
MHS-14*	23	22	(18-21)	11.07	3.14%
MHS-15*	21	28	03	11.05	3.14%
MHS-16*	7	71	1A	11.20	1.15%
MHS-17*	9	135	2	10.66	3.10%
MHS-18*	11	62	10	10.64	3.09%

Table 3. Summary statistics of the significant QTL for prognathism in European seabass.
MHS= maternal half-sib, LG = Linkage Group, cM = centimorgan, Chr = Chromosome, F= F-statistic.
****Genome-wide significant QTL (P<0.01)
***Genome-wide significant QTL (P<0.05)
**Chromosome-wide significant QTL (P<0.01)
*Chromosome-wide significant QTL (P<0.05)

GWAS for Mandibular Prognathism

A larger set of samples, including all the 298 juveniles collected from all families, was used for GWAS accounting for family structure. A total of 9,250 SNP markers was genotyped in more than 80% of the experimental population. After filtering for MAF < 0.05, 7,362 loci were retained for further analysis. All these SNPs mapped onto the seabass genome. The case-control allelic association analysis for MP was carried out with a mixed linear model and applied to 107 cases and 191 controls. Two SNPs, respectively on ChrX and Chr17, reached Bonferroni-corrected genome-wide significance level (P < 0.05) (Fig. 2, Table 4). These two candidate markers were in close proximity with the peaks of the two most significant QTLs on LG22 and LG18. GWAS results are based on a larger test panel than QTL analysis and include unrelated individuals. This suggests that the two identified regions could be involved in the determination of lower jaw deformity beyond a single family.

The most significant association with prognathism was found for SNP marker L_39743 (P < 0.01 after Bonferroni correction), which was located at position 3,443,465 on ChrX within a putative gene showing an *in silico* predicted transcript (DLAgn_00206060) without any significant sequence similarity. The nearest genotyped SNP markers were positioned approximately 100 kilo base pairs

(kbp) upstream (L_39738, ChrX:3,348,951) and downstream (L_39753, ChrX:3,540,808), defining a genomic region where four protein-coding genes are present (DLAgn_00206040, DLAgn_00206050, DLAgn_00206070, DLAgn_00206080).

SNP	Seabass chromosome	Position (bp)	Minor allele frequency	Harbouring gene	Nearest gene	p-value
L_39743	ChrX	3,443,463	G(0.11)/T	-NA-	PHLDA1	1.32E-6*
L_12903	Chr17	1,566,309	G(0.42)/A	SOBP	ROCK2	1.1E-5*

Table 4. SNPs associated with mandibular prognathism using a case/control mixed linear model based association analysis. Significance after Bonferroni correction was highlighted by an asterisk. NA= not annotated

DLAgn_00206040 encodes a protein that contains a Fibronectin type 3 domain, a calcium-binding EGF-like domain, and a Zona pellucida-like domain and appears to be well-conserved within Acanthomorphs, but without a clearly identified ortholog in other vertebrates. DLAgn_00206050 codes for cAMP-regulated D2 protein-like (cAMP-D2), which has a putative orthologue in zebrafish (ZDB-GENE-060503-450, ENSDARG00000058492). Based on evidence from Ensembl Compara, ENSDARG00000058492 shows only one additional ortholog in cavefish. Paralogs originating from an ancestral gene duplication are present in several teleosts while no homolog is identified outside ray-finned fish. DLAgn_00206070 is the closest protein-coding gene to SNP L_39743 (< 35 kbp downstream) and shows significant homology with *pleckstrin homology-like domain family a member 1* (*Phlda1*). In humans, *Phlda1* is associated with autosomal recessive intermediate osteoporosis where MP was reported in affected patients [18]. DLAgn_00206080 encodes a protein homologous to nucleosome assembly protein 1-like 1 (NAP1L1), which is involved in chromatin re-assembly. Remarkably, a genomic region of approximately 100 kbp without any known protein-coding gene spans between DLAgn_00206050 and DLAgn_00206070 and includes marker L_39743. Such a region, however, appears to be well conserved across several teleost genomes, with the presence of non-coding sequence elements showing high similarity especially within Acanthomorphs, the most derived teleost species (Fig. 3). Conserved non-coding elements (CNEs) have been found in comparisons of genomic regions at different taxonomic levels. Multiple lines of evidence suggest that CNEs have an important role in regulating gene expression, often encoding enhancers that act on nearby genes as well as distant ones and in several cases exerting their action on genetic loci involved in development [19]. It is therefore possible that the observed CNEs located between DLAgn_00206050 and DLAgn_00206070 are involved in transcriptional regulation. In turn, the genetic variant linked to L_39743 might contribute to lower jaw deformity altering patterns of gene expression rather than directly affecting the sequence of a protein coding gene. In fact, a recent

study, using a combination of genomic technologies in mice, has shown the role of distant-acting enhancers in craniofacial development [20]. Although such evidence comes from a mammalian species, it is well recognized that despite the large evolutionary distance, the fundamental signaling pathways and cellular events that shape the craniofacial skeleton appear to be highly conserved from fish to human [21].

The second significant marker ($P < 0.05$ after Bonferroni correction), L_12903, is positioned on Chr17 within the coding region of the gene *sine oculis-binding protein homolog* (*Sobp*). The closest upstream and downstream SNP markers were respectively L_51826 (Chr17:15,653,554) and L_12906 (Chr17:15,697,886) defining a relatively small region of approximately 45 kb, where four putative genes are located (DLAgn_00071270, DLAgn_00071280, DLAgn_00071290, DLAgn_00071300). DLAgn_00071270 encodes rho-associated protein kinase 2-like (ROCK2). ROCK2 is a protein kinase, which is a key regulator of actin cytoskeleton and cell polarity. The Rho kinase is involved in several biological processes including migration of neural cell precursors [22] and survival of neural crest cells (NCC) [23]. NCC give rise to a great proportion of the tissues forming the vertebrate head and face. DLAgn_00071280 harbours, as already mentioned, marker L_12903 and codes for SOBP, a nuclear zinc finger protein whose molecular functions are only partially understood. Recessive mutations at the *Sobp* locus in mice cause defective patterning of sensory epithelium and deformed organ of Corti in the inner ear, while a homozygous missense mutation in *Sobp* has been associated with a syndrome causing mental retardation, anterior maxillary protrusion and strabismus [24,25]. SOBP may act as a critical transcription factor for neural development [26]. In fact, *Sobp* is included in the list of genes specifically expressed in NCC from the cranial mesenchyme, which play a key role in craniofacial development [27]. Two additional genes that characterize cranial mesenchyme NCC are located nearby marker L_12903, DLAgn_00071250 (Chr17:15,618,828-15,622,452) encoding leucine-rich alpha-2-glycoprotein, and DLAgn_00071210 (Chr17: 15,550,765-15,570,987) coding for apolipoprotein b-100.

Two additional genes are comprised between markers L_51826 and L_12906. DLAgn_00071290 seems to be expressed, but encodes a short protein sequence with a positive match of limited similarity with hypothetical proteins just in a few teleost species. DLAgn_00071300 encodes reticulon-4-interacting protein 1 mitochondrial-like (RTN4IP1), which has not been associated, until present, with any developmental process or malformation.

At locus L_39743, 55 (51%) prognathous fish were heterozygous for the minor allele and 52 (49%) were homozygous for the major allele. In control animals, 172 (90%) were homozygous for the major allele and 19 (10%) were heterozygous for the minor allele. At locus L_12903, 42 (39.25%) MP affected fish were heterozygous, 8 (7.5%) were homozygous for the minor allele while 57 (53.25%)

were homozygous for the major allele. Among normal fish, 105 (55%) were heterozygotes, 45 (23.5%) were homozygotes for the minor allele, and 41 (21.5%) were homozygotes for the major allele (Fig. 4). The SNP effect calculated by GCTA was 0.23 for locus L_39743 and -0.16 for locus L_12903. A positive value means that the minor allele increases risk relative to phenotype, conversely a negative value represents a reduced risk (a protective effect of the minor allele). On the basis of the kinship matrix derived from the SNP data, the MP trait heritability was estimated using GCTA [28]. A moderate and significant value of 0.26 ± 0.08 (P-value= 9.6E-15) was found. Notably, this value is in agreement with estimates previously reported in the European seabass for skeletal deformities [11].

A GWAS analysis carried out on growth trait pointed out two loci (not significant after Bonferroni correction) both located on European seabass Chr8 (Table S2). None of the significant GWAS loci associated to the MP were found on Chr8, excluding a putative correlation between both traits.

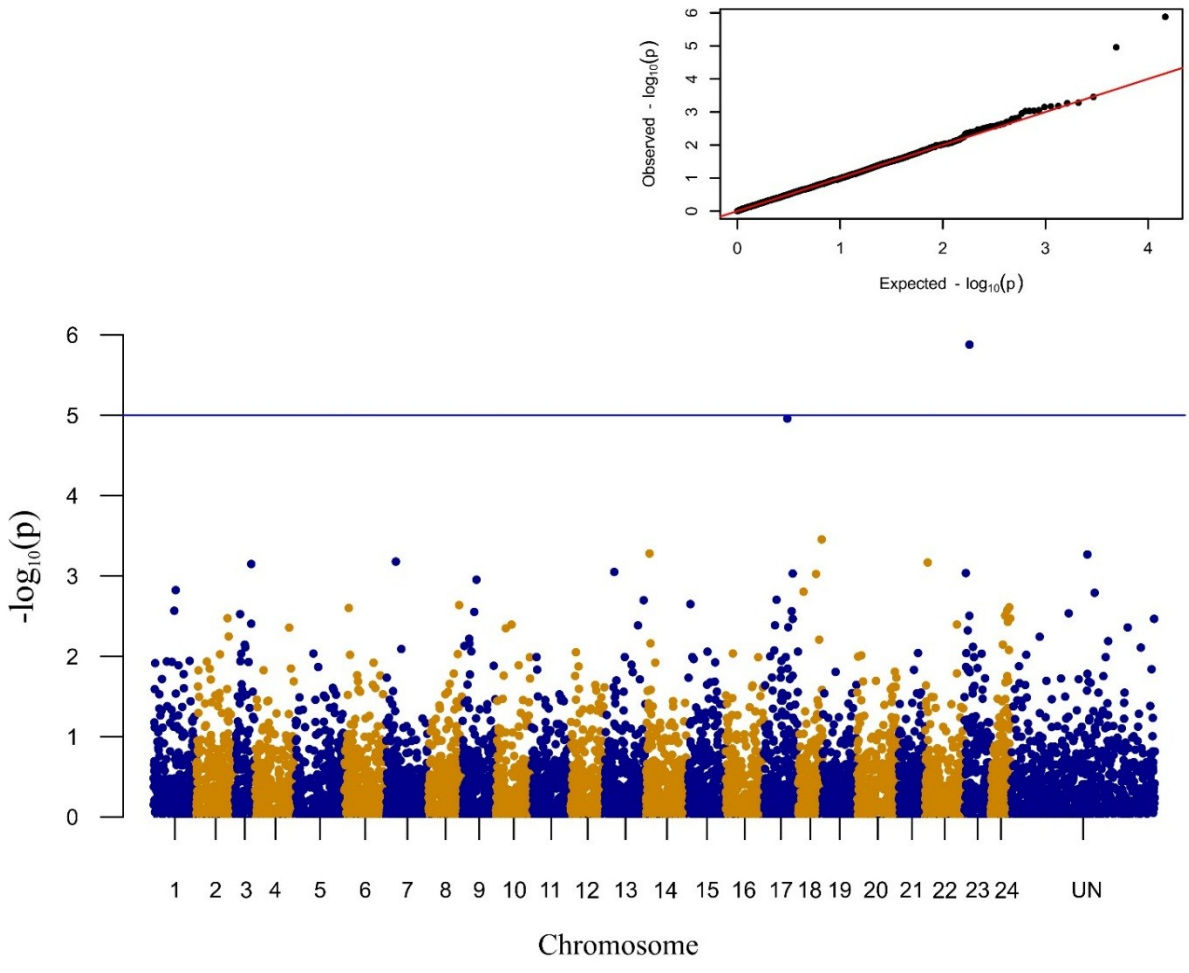


Figure 2. Manhattan plot for mandibular prognathism. A genome-wide case-control study showed a significant association of the phenotype prognathism on ChrX (labelled as Chr23) and a marginal significant association on Chr17. The blue line indicates the threshold level (-log10(1e-05)). The inset shows a quantile-quantile (qq) plot with the observed plotted against the expected p-values. The remaining unanchored scaffolds/contigs, those that could not be localized to a chromosome were concatenated into the virtual chromosome "UN" with 100 bp gaps between scaffolds.

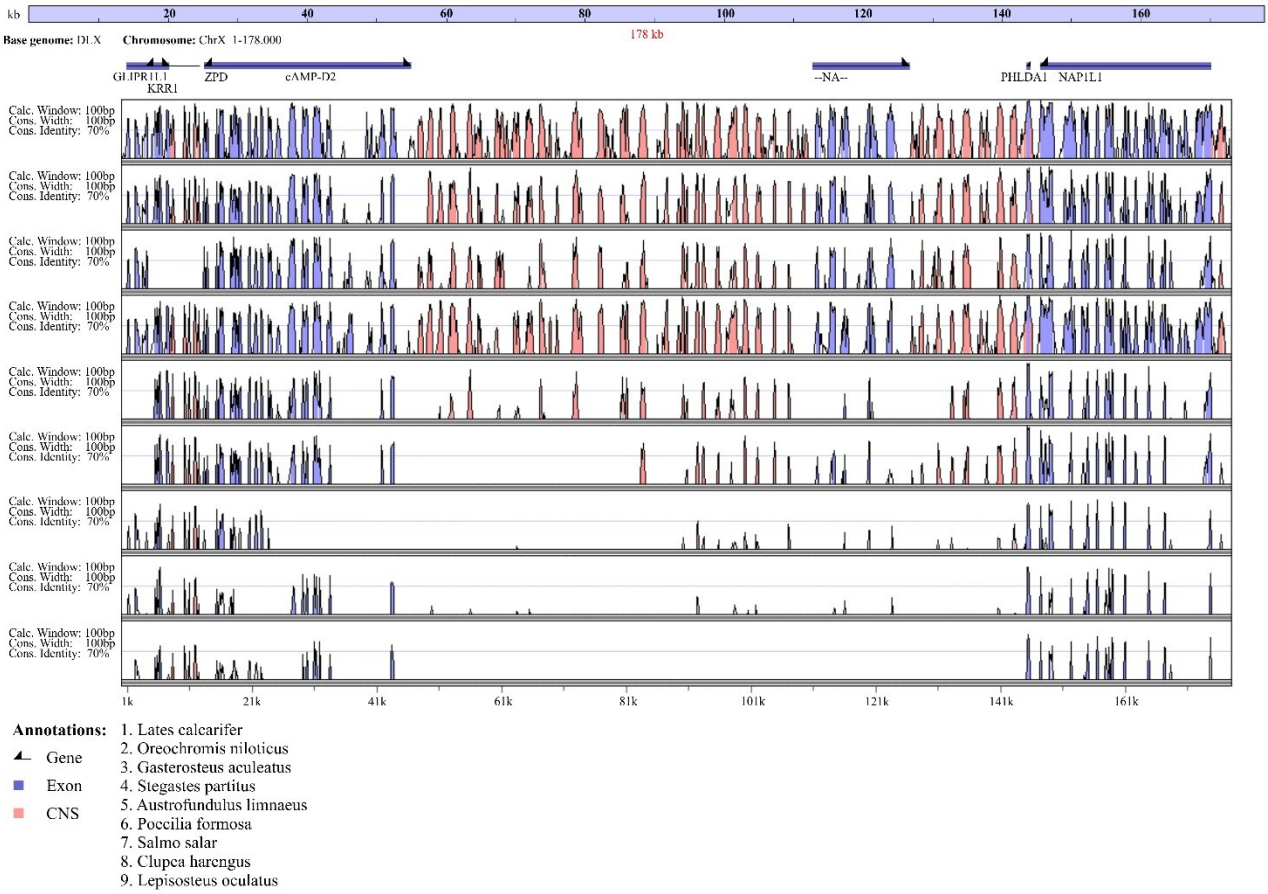


Figure 3. Comparison at different taxonomic levels of the ChrX genomic region flanking the significant locus L_39743.

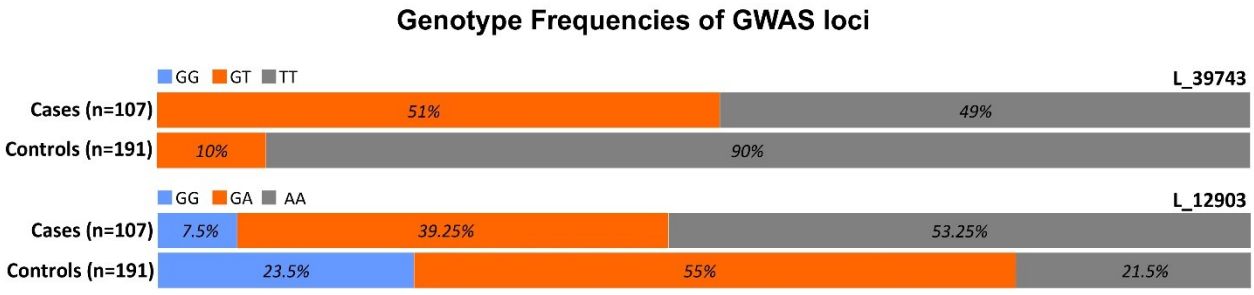


Figure 4. Genotype frequencies of the two best associated SNPs L_39743 and L_12903.

Gene expression profiling of prognathous larvae and juveniles

Partial microarray data on gene expression profiles of the lower jaw at 58-days post hatching were already reported by [29]. Here, additional samples for the same developmental stage and novel data on larval heads at 38 dph are presented after the implementation of novel statistical tools to correct for array batch effects. Raw and normalized fluorescence data, including the newly added microarray experiments, were deposited in the GEO database under accession number GSE85056.

A two-class unpaired SAM analysis was performed to identify differentially expressed genes between normal and jaw deformed larvae (whole-head; 38 dph) and juveniles (dissected jaw; 58 dph). In the first case, 92 probes were significant at FDR < 0.05. Of these, 72 were more expressed in the whole head of prognathous larvae and 62 had a putative ortholog in zebrafish (Supplementary Table S3). In the case of dissected lower jaw, 708 probes, corresponding to 429 unique transcripts, were found significant, 21 (18 unique transcripts) being up-regulated and 687 (411 unique transcripts) being down-regulated in jaw deformed juveniles (Supplementary Table S4). The larger number of differentially expressed genes (DEGs) in the latter case suggests that dissecting the affected anatomical region substantially increased the discriminatory power of gene expression profiling. It might also be possible that, as fish grow, the defect becomes more pronounced, not only macroscopically, as already observed, but also in terms of affected genes/gene pathways.

To obtain a more comprehensive interpretation of the set of genes differentially expressed, a functional enrichment analysis was performed using the tool DAVID (see Methods). Six GO Biological Process (GO_BP) terms were significantly enriched, with marginal statistical support (Supplementary Table S5). The most numerous term was “metabolic process” with 26 genes. Of these, quite relevant was the putative ortholog of zebrafish ENSDARG00000074481, which encodes unc-51 like autophagy activating kinase 1 (ULK1b). ULK1 belongs to a kinase triad (AMPK, mTORC1, ULK1) that controls cell growth depending on energy and nutrient status [30]. More specifically, ULK1 is a key player in inducing autophagy, a process that is known to be essential for vertebrate development [31,32].

A total of 172 putative zebrafish orthologs matched DEGs in the juvenile lower jaws. Using zebrafish functional annotations as a proxy for seabass, 38 GO_BP terms were found to be significantly enriched (Supplementary Table S5). Several terms involved microtubule-related processes and contained, in particular, four different genes encoding stathmin-like proteins (STMN1B, STMN2A, STMN2B, STMN4), which were all significantly down-regulated in deformed lower jaw. Stathmins are tubulin-interacting proteins involved in several cell functions [33].

A potential link might also exist between stathmins and one of the genetic loci (DLAgn_00071270 encoding ROCK2) located in the candidate region of chromosome 17, since STMN and ROCK were

shown to interact to control cell migration [34]. Stathmin mediates neuroblastoma metastasis in a tubulin-independent manner via RhoA/ROCK signaling and enhanced transendothelial migration [34]. Stathmins were also reported to be essential for neural cell differentiation[35]. The microtubule destabilizing protein stathmin controls the transition from dividing neuronal precursors to postmitotic neurons during adult hippocampal neurogenesis [35].

In fact, stathmins are also found in another significantly enriched GO_BP term, "nervous system development" (Supplementary Table S5), further confirming the role of neural cells in the development of jaw deformity. In total, 18 DEGs are included in this GO term. Of these, the most relevant gene is possibly DLAgn_00053460, the putative ortholog of zebrafish *Zic2a*. ZIC2 is a zinc-finger transcription factor, which is a key player in craniofacial development. In humans it is strongly linked to holoprosencephaly, the most common cranial malformation. In zebrafish, it was shown to play a dual role during craniofacial development contributing to two different aspects of craniofacial morphogenesis: i) neural crest induction and migration, and ii) early patterning of tissues adjacent to craniofacial chondrogenic condensations [36].

Significant differential expression of a putative *Zic2* ortholog in the deformed jaw provides evidence that altered functioning of NCC likely drives MP in seabass.

Integrating gene expression profiles and genetic data

To further explore the potential links between transcriptomic characterization of jaw deformity and GWAS for MP, a dedicated analysis was performed to assess whether there was any perturbation of expression for genes located in the genomic regions nearby the two GWAS-significant genetic markers. Using a conservative approach, a relatively large target region was considered and a two-class SAM significant test was carried out on all probes mapping in the regions spanning ± 500kb respectively around loci L_39743 and L_12903 (Supplementary Table S6).

At variance with evidence on a whole-transcriptome scale, fewer significant transcripts were observed in gene expression profiling of the lower jaw at 58 days than in whole-heads (38-days larvae), with three genes within region "L_39743" and one gene for region "L_12903". Of these, two distinct probes indicated significant down-regulation of DLAgn_00206050, one of the four genes in the narrow region on ChrX between markers L_39738 and L_39753, encoding cAMP-D2, immediately upstream several CNEs (Fig. 3).

Transcriptome profiling of whole-heads identified a much larger set of DEGs, with five genes located on the ChrX region and eight on the Chr17 one. One gene (DLAgn_00206010) showed substantial up-regulation in deformed larvae with fold-change > 6. DLAgn_00206010 codes for CMP-Neu5Ac hydroxylase (CMAH), an enzyme involved in the synthesis of sialic acid. CMAH enzyme activity was lost in humans, although *Cmah* expression was reported as (mesenchimal) stem cell marker [37].

In lower vertebrates (rainbow trout and *Xenopus laevis*) *Cmah* was found to be expressed in the ovary [38], while *Cmah*-null mice showed hearing loss and several other abnormalities [39].

Among the several DEGs in whole-heads located in the target region on Chr17, *Sobp* was found to be significantly down-regulated in deformed larvae, with concordant evidence from two probes (Supplementary Table S6). Such evidence further supports *Sobp* as a candidate gene contributing to mandibular prognathism.

In conclusion, in the present study, integration of transcriptomic analysis with GWAS provided evidence for the potential mechanisms underlying jaw deformity in the European seabass. QTL mapping and GWA analysis allowed the identification of two regions implied in the determination of lower jaw deformity and pointed out a candidate gene, *Sobp*, likely contributing to seabass MP. Moreover, the presence of a cluster of CNEs around to the most significant SNP on ChrX is suggestive since such elements might act as distant enhancers in craniofacial development.

Finally, as previously reported in model species, differential regulation of several genes involved in neural development, such as putative *Zic2a* and stathmins, confirms the importance of this biological process to develop craniofacial deformities.

The present work might be considered as a case-study proving the feasibility of an integrated genomic approach as a compelling strategy to unravel the molecular bases of skeletal anomalies in fish aquaculture. As a future perspective, these integrated methods could be pivotal to the development of genetic tools intended to be applied in breeding selection.

Methods

Ethics statement. No specific permits were required for the work described here. Individuals included in the present study were bought from a commercial hatchery and they were not subjected to any experimental manipulation. The study was performed in accordance with the EU directive 2010/63/EU and Italian DL 2014/26. The experiments, as well as the euthanasia procedure, were monitored and carried out by authorized staff to minimise animals' suffering.

Samples collection, phenotypes description and parental assignment. A total of 298 juveniles (79-83 days old, average standard length 4 cm) and 48 broodstocks were collected and analysed. All samples were provided by the fish farm "Cà Zuliani" (Pila di Porto Tolle, Italy). Jaw deformed phenotype was assigned to individual juvenile affected fish by two operators independently. Each fish was also photographed, weighed, and its length measured. Out of 298 seabass juveniles, 107 were affected by MP (191 unaffected). For all subsequent analyses (i.e. QTL and GWAS) the presence/absence of prognathism was coded as 1/0, respectively. Microsatellites analysis was performed on both adults and juveniles by using a set of 9 loci according to [40] (see Supplementary

Table S7). Briefly, alleles scoring was achieved by means of Genotyper v3.7 (Applied Biosystems) and the parental assignment test was assessed with the software Cervus (http://fieldgenetics.com) using default settings.

2b-RAD libraries preparation and sequencing. Genomic DNA (gDNA) was extracted from approximately 20 mg of tissue (fin clip) using the commercial kit Invisorb® Spin Tissue Mini Kit (Invitek, STRATEC Biomedical, Germany) following the manufacturer's recommendations. Genomic DNA concentration and purity were quantified by using both a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and a Qubit 2.0 Fluorimeter (Invitrogen, ThermoFisher Scientific, MA, USA). This procedure ensured comparable concentrations of high quality gDNA as required for 2b-RAD library preparation.

A 2b-RAD library was constructed for each individual following a modification of the protocol from [41]. To assess the robustness of the method, two libraries were replicated (Technical replicates, TRs) for two individuals [42]. A total of 200 ng of gDNA from each sample were digested with 2 U *CspC*I (New England Biolabs, NEB, Ipswich, Massachusetts, USA) for 1h at 37 °C. Sample-specific barcodes were designed with Barcode generator (http://comailab.genomecenter.ucdavies.edu) and introduced by PCR with platform-specific barcode-bearing primers. The concentration of each purified individual library was quantified using Qubit®ds DNA BR Assay Kit (Invitrogen, ThermoFisher Scientific, MA, USA) and Mx3000P qPCR instrument, while the libraries quality was checked on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). Individual libraries were pooled into equimolar amounts. Pooled libraries were sequenced on an Illumina HiSeq2500 platform with a 50 bp single-read module at the Genomix4Life S.r.l. facility (Baronissi, Salerno, Italy), which also performed data demultiplexing.

SNP discovery and genotyping. Demultiplexed reads returned by the sequencing facility were quality-checked with the software FastQC (http://bioinformatics.babraham.ac.uk). Adapter trimming was performed by running a customized script (available upon request), thus obtaining 32-bp fragments ready to be evaluated for SNPs presence in STACKS. Filtered reads were then mapped against the European seabass genome [43] with CLC Genomics Workbench version 8.5 (http://qiagenbioinformatics.com) mapping module. The following parameters were applied: *length fraction* = 1.0 and *similarity fraction* = 0.9 (all remaining parameters as default), retaining only uniquely mapped reads. Mapping results were exported in SAM format and used as input for *refmap_map.pl* in STACKS v. 1.36, a program for SNP discovery and genotyping [44,45]. Different settings were tested on the TRs dataset to fine tune the STACKS pipeline parameters and to assess the consistency of the results as in [42]. To construct stacks and catalogue loci, a minimum coverage of 10X was used for parental samples [45]. A maximum of two mismatches between stacks were allowed

for catalogue construction. For the offspring, stacks were assembled with a minimum coverage of 5X sequencing reads. One SNP per locus and maximum 2 alleles allowed were chosen as parameters for the genotype data calling. Loci with more than 30% of missing data and with more than one SNP/locus were excluded from further analyses.

Linkage map construction. An averaged-sex linkage map was constructed using Lep-MAP2 [46], a software suite handling thousands of markers and multiple families. All the four full-sib families were used as input, and SNPs showing significant segregation distortion with P-value < 0.01 and a minor allele frequency MAF < 0.05 were excluded. The remaining markers were used for the linkage group assignment with LOD score thresholds of 6-12. Singular markers were then added to the LGs found using the *JoinSingles* module with a LOD score limit of 6. The order of markers was calculated with *OrderMarkers* module, with a recombination rate parameter of 0.40 and "useKosambi=1" and maxDistance=50" options. The ordering task was performed 6 times and the order with the best likelihood was chosen following the author guidelines of LepMAP2 (https://sourceforge.net/p/lepmap2/wiki/Home/). The map was drawn using MapChart v 2.3 [47].

QTL mapping for MP. The analysis was carried out using a half-sib regression on the MHS family, implemented in the online tool GridQTL (http://gridqtl.org.uk) and described by [48]. This software implements a multi-marker approach of interval mapping in half-sib families; notably, this method of QTL analysis does not assume the parents had fixed QTL alleles therefore relaxing this parameter [48]. The *F* statistic is given for each LG for the most likely position as well for the chromosome-wide and the genome-wide thresholds used to determine the significance of the detected QTL. Each threshold was obtained by 1000 permutations for the trait. A 95% confidence interval for each significant QTL (CI95) was determined using 10,000 bootstraps with resampling. A genome scan for QTLs for growth trait was also performed with GridQTL, using the same parameters described above. The estimated proportion of the phenotypic variance explained by each QTL (i.e., the estimated heritability due to the QTL) was calculated with the formula $1 - 10^{(-\frac{2}{n}LOD)}$, where *n* is the sample size [49].

Genome wide association study. The entire dataset of 298 juvenile seabass was used for GWA analysis. The case/control (107 cases and 191 controls) GWAS for MP was performed using a mixed linear model (suitable for binary traits as prognathism) implemented in GCTA [28] and considering the relatedness of juveniles. The advantages of mixed-linear-model association (MLMA) method include the prevention of false positive associations due to population or relatedness structure and an increase in power obtained through the application of a correction that is specific to this structure [50]. Briefly, first three files (.bed, .bim and .fam) were generated for the GWAS genotypes using PLINK [51]; then a *–make-grm* option was used to generate grm.gz and grm.id files; a phenotype file for prognathism

trait was prepared and the *–mlma* option was used to perform the MLMA association analysis. We considered genome-wide significance where P-values were below the 5% corrected threshold for 7,362 independent test. The adjusted P-values were determined with the function *p.adjusted* implemented in R-CRAN (https://cran.r-project.org). Variance explained by all SNPs was estimated with the software GCTA with a user-specified disease prevalence = 0.1 [28]. First, a genetic relationship matrix (GRM) between pairs of individuals was calculated via the *-make-grm* function in GCTA, then a restricted maximum likelihood (REML) analysis was performed to estimate the phenotypic variance explained by the SNPs [52,53]. A Manhattan and Q-Q plot for GWAS data were created with the R-CRAN package *qqman* v. 0.1.2 [54].

Similarly, GWAS analysis for growth trait was carried out using GCTA with the same parameters as described above.

Conserved non-coding elements (CNEs) were investigated through mVISTA, a tool for comparative genomics analysis available online (http://genome.lbl.gov/vista/index.shtml). A point biserial correlation coefficient (r_{pb}) was calculated to test the association, at phenotypic level, between prognathism and growth using *ltm* v. 1.0 package [55]. The correlation coefficient r_{pb} calculated is a measure of the strength of association between a continuous-level variable and a binary data.

Microarray Gene Expression Analysis. Two seabass developmental stages: i) larvae (38 days-old, average length 12 mm), and ii) juveniles (58 days-old, average standard length 16 mm) were collected at the fish farm "Ca' Zuliani" (Pila di Porto Tolle, Italy) and sacrificed using an excess of anesthetic, as previously described. For 38 dph larvae, the cranial regions were dissected under a stereomicroscope from15 normal and 15 jaw-deformed individuals and pooled (5 heads/pool) to obtain 3 independent pools per condition. For 58 dph juveniles, the lower jaws were dissected and pooled (5 jaws/pool) from a total of 50 individuals, 25 normal and 25 affected by jaw-deformity, thus providing 5 independent pools per condition. Gene expression experiments were performed using a single dye (Cy3) labelling scheme on the Agilent-019810 *D. labrax* oligo microarray (GEO accession: GPL9663) containing 19,035 unique transcripts [29], each represented by two non-overlapping probes. Normalization procedures were performed using R statistical software. Microarray data were cyclic lowess (CL) normalized across all arrays and CL-normalized data were further adjusted for the known between-experiments batch effects by implementing parametric Combat correction in R [56]. A two-class comparison (FDR < 5%, fold-change ≥ 2) was carried out on Significance Analysis of Microarrays (SAM) software in order to identify differentially expressed genes between normal and jaw-deformed groups [57]. A functional interpretation of DEGs was obtained through enrichment analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) software version 6.8 beta [58] with default parameters. Since DAVID database

contains functional annotation data for a limited number of species, it was necessary to use the zebrafish feature as identifiers. To retrieve zebrafish IDs, a BLASTX search (cut off e-value <1.0 e-5) was carried out against the high quality *D. rerio* draft genome stored on ENSEMBL database by using all seabass transcripts as query. These identifiers were used to define a "gene list" and a "background", corresponding to the list of differentially transcribed seabass genes and to all the transcripts that were represented on the array, respectively.

Additional Information

Accession codes: sequencing data was deposited in the NCBI-short read archive (SRA) database under the accession number SRP076258. Gene expression data are available on Gene Expression Omnibus (GEO) repository under the accession number GSE85056.

References

1. FAO Fisheries & Aquaculture - cultured aquatic species information programme - Dicentrarchus labrax (Linnaeus, 1758). www.fao.org/fishery/culturedspecies/Dicentrarchus_labrax/en. (Accessed: 11th May 2016)
2. Zouiten, D., Ben Khemis, I., Slaheddin Masmoudi, A., Huelvan, C. & Cahu, C. Comparison of growth, digestive system maturation and skeletal development in sea bass larvae reared in an intensive or a mesocosm system. *Aquac. Res.* **42,** 1723–1736 (2011).
3. Boglione, C. *et al.* Skeletal anomalies in reared European fish larvae and juveniles. Part 1: normal and anomalous skeletogenic processes. *Rev. Aquac.* **5,** S99–S120 (2013).
4. Boglione, C. *et al.* Skeletal anomalies in reared European fish larvae and juveniles. Part 2: main typologies, occurrences and causative factors. *Rev. Aquac.* **5,** S121–S167 (2013).
5. Boglione, C. & Costa, C. in *Sparidae* (eds. Pavlidis, M. A. & Mylonas, C. C.) 233–294 (Wiley-Blackwell, 2011).
6. Sfakianakis, D. G. *et al.* Environmental determinants of haemal lordosis in European sea bass, Dicentrarchus labrax (Linnaeus, 1758). *Aquaculture* **254,** 54–64 (2006).
7. Koumoundouros, G. Morpho-anatomical abnormalities in Mediterranean marine aquaculture. *Recent Adv. Aquac. Res.* (2010).
8. Mazurais, D. *et al.* Optimal levels of dietary vitamin A for reduced deformity incidence during development of European sea bass larvae (Dicentrarchus labrax) depend on malformation type. *Aquaculture* **294,** 262–270 (2009).
9. Cahu, C., Zambonino Infante, J. & Takeuchi, T. Nutritional components affecting skeletal development in fish larvae. *Aquaculture* **227,** 245–258 (2003).
10. Nguyen, N. H., Whatmore, P., Miller, A. & Knibb, W. Quantitative genetic properties of four measures of deformity in yellowtail kingfish Seriola lalandi Valenciennes, 1833. *J. Fish Dis.* **39,** 217–228 (2016).
11. Karahan, B. *et al.* Heritabilities and correlations of deformities and growth-related traits in the European sea bass (Dicentrarchus labrax, L) in four different sites. *Aquac. Res.* **44,** 289–299 (2013).
12. Cruz, R. M. *et al.* Major gene and multifactorial inheritance of mandibular prognathism. *Am. J. Med. Genet. A.* **146A,** 71–77 (2008).
13. Chen, F. et al. Identification of a mutation in FGF23 Involved in mandibular prognathism. *Sci. Rep.* **5,** 11250 (2015).

14. Jang, J. Y. et al. Polymorphisms in the matrilin-1 gene and risk of mandibular prognathism in koreans. *J. Dent. Res.* **89,** 1203–1207 (2010).
15. Signer-Hasler, H. et al. A chromosomal region on ECA13 is associated with maxillary prognathism in horses. PLoS ONE **9**, e86607 (2014).
16. Volckaert, F. A. et al. Heritability of cortisol response to confinement stress in European sea bass Dicentrarchus labrax. *Genet. Sel. Evol.* **44,** 15 (2012).
17. Palaiokostas, C. *et al.* A new SNP-based vision of the genetics of sex determination in European sea bass (Dicentrarchus labrax). *Genet. Sel. Evol.* **47,** 1–10 (2015).
18. Neville, B. W., Damm, D. D., Chi, A. C. & Allen, C. M. *Oral and Maxillofacial Pathology*. (Elsevier Health Sciences, 2015).
19. Douglas, A. T. & Hill, R. E. Variation in vertebrate cis-regulatory elements in evolution and disease. Transcription **5**, (2014).
20. Attanasio, C. et al. Fine tuning of craniofacial morphology by distant-acting enhancers. Science **342**, 1241006 (2013).
21. Mork, L. & Crump, G. Zebrafish craniofacial development: A window into early patterning. *Curr. Top. Dev. Biol.* **115,** 235–269 (2015).
22. Leong, S. Y., Faux, C. H., Turbic, A., Dixon, K. J. & Turnley, A. M. The Rho kinase pathway regulates mouse adult neural precursor cell migration. *Stem Cells Dayt. Ohio* **29,** 332–343 (2011).
23. Phillips, H. M. et al. Neural crest cell survival is dependent on rho kinase and is required for development of the mid face in mouse embryos. PLOS ONE **7**, e37685 (2012).
24. Basel-Vanagaite, L. *et al.* Autosomal recessive mental retardation syndrome with anterior maxillary protrusion and strabismus: MRAMS syndrome. *Am. J. Med. Genet. A.* **143A,** 1687–1691 (2007).
25. Birk, E. et al. SOBP Is mutated in syndromic and nonsyndromic intellectual disability and is highly expressed in the brain limbic system. *Am. J. Hum. Genet.* **87,** 694–700 (2010).
26. Sun, H.& Hunter, T. Poly-small ubiquitin-like modifier (PolySUMO)-binding proteins identified through a string Search. *J. Biol. Chem.* **287,** 42071–42083 (2012).
27. Brunskill, E. W. *et al.* A gene expression atlas of early craniofacial development. *Dev. Biol.* **391,** 133–146 (2014).
28. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

29. Ferraresso, S. *et al.* Development of an oligo DNA microarray for the European sea bass and its application to expression profiling of jaw deformity. *BMC Genomics* **11,** 1–17 (2010).
30. Dunlop, E. A. & Tee, A. R. The kinase triad, AMPK, mTORC1 and ULK1, maintains energy and nutrient homoeostasis. *Biochem. Soc. Trans.* **41,** 939–943 (2013).
31. Lee, E. *et al.* Autophagy is essential for cardiac morphogenesis during vertebrate development. *Autophagy* **10,** 572–587 (2014).
32. Mizushima, N. & Levine, B. Autophagy in mammalian development and differentiation. *Nat. Cell Biol.* **12,** 823–830 (2010).
33. Cassimeris, L. The oncoprotein 18/stathmin family of microtubule destabilizers. *Curr. Opin. Cell Biol.* **14,** 18–24 (2002).
34. Fife, C. M. *et al.* Stathmin mediates neuroblastoma metastasis in a tubulin-independent manner via RhoA/ROCK signaling and enhanced transendothelial migration. *Oncogene* (2016). 10.1038/onc.2016.220
35. Boekhoorn, K. *et al.* The microtubule destabilizing protein stathmin controls the transition from dividing neuronal precursors to postmitotic neurons during adult hippocampal neurogenesis. *Dev. Neurobiol.* **74,** 1226–1242 (2014).
36. TeSlaa, J. J., Keller, A. N., Nyholm, M. K. & Grinblat, Y. Zebrafish Zic2a and Zic2b regulate neural crest and craniofacial development. *Dev. Biol.* **380,** 73–86 (2013).
37. Nystedt, J. *et al.* Human CMP-N-acetylneuraminic acid hydroxylase is a novel stem cell marker linked to stem cell-specific mechanisms. *Stem Cells Dayt. Ohio* **28,** 258–267 (2010).
38. Gohin, M., Bobe, J. & Chesnel, F. Comparative transcriptomic analysis of follicle-enclosed oocyte maturational and developmental competence acquisition in two non-mammalian vertebrates. *BMC Genomics* **11,** 18 (2010).
39. Kwon, D.-N., Park, W.-J., Choi, Y.-J., Gurunathan, S. & Kim, J.-H. Oxidative stress and ROS metabolism via down-regulation of sirtuin 3 expression in Cmah-null mice affect hearing loss. *Aging* **7,** 579–594 (2015).
40. Souche, E. L. *et al.* Range-wide population structure of European sea bass *Dicentrarchus labrax*. *Biol. J. Linn. Soc.* **116,** 86–105 (2015).
41. Wang, S., Meyer, E., McKay, J. K. & Matz, M. V. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* **9,** 808–810 (2012).
42. Pecoraro, C. *et al.* Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (Thunnus albacares). *Mar. Genomics* **25,** 43–48 (2016).

43. Tine, M. *et al.* European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* **5,** 5770 (2014).
44. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22,** 3124–3140 (2013).
45. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes Genomes Genet*. **1,** 171–182 (2011).
46. Rastas, P., Calboli, F. C. F., Guo, B., Shikano, T. & Merilä, J. Construction of ultradense linkage maps with Lep-MAP2: stickleback F 2 recombinant crosses as an example. *Genome Biol. Evol.* **8,** 78–93 (2016).
47. Voorrips, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93,** 77–78 (2002).
48. Knott, S. A., Elsen, J. M. & Haley, C. S. Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **93,** 71–80 (1996).
49. Broman, K. W. & Sen, S. A Guide to QTL Mapping with R/qtl. (Springer New York, 2009).
50. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46,** 100–106 (2014).
51. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
52. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. 88, 294–305 (2011).
53. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42,** 565–569 (2010).
54. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *biorXiv.* 10.1101/005165 (2014).
55. Rizopoulos, D. ltm: An R Package for Latent Variable Modeling and Item Response Analysis. *J. Stat. Softw.* **17,** 1–25 (2006).
56. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostat. Oxf. Engl.* **8,** 118–127 (2007).
57. Li, J. & Tibshirani, R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Stat. Methods Med. Res.* **22,** 519–536 (2013).

58. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4,** (2009).

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).
Supplementary Table S1.
Supplementary Table S2.

QTL	LG	Position (cM)	Chr	F	Expl. Variation
MHS-01***	5	123	6	21.37	3,95%
MHS-02***	19	7	19	21.12	3.94%
MHS-03***	9	72	2	20.43	3.90%
MHS-04***	23	86	(18-21)	19.81	3.86%
MHS-05**	10	87	8	16.17	3.61%
MHS-06**	1	89	16	14.13	3.45%
MHS-07**	2	118	13	12.39	3.34%
MHS-08*	24	83	24	9.26	2.91%

Supplementary Table S1: Summary statistics of the significant QTL for growth in European seabass. MHS= maternal half-sib, LG = Linkage Group, cM = centimorgan, Chr = Chromosome, F= F-statistic.

****Genome-wide significant QTL (P<0.05)*
***Chromosome-wide significant QTL (P<0.01)*
**Chromosome-wide significant QTL (P<0.05)*

SNP	Seabass chromosome	Position (bp)	Minor allele frequency	Harbouring gene	Nearest gene	p-value
L_37058	Chr8	5,769,773	G(0.13)/A	-NA-	WFIKKN2	2.5E-4^{ns}
L_37059	Chr8	5,775,389	G(0.22)/A	TOB1	WFIKKN2	3.2E-4^{ns}

Supplementary Table S2: SNPs associated with growth using a mixed linear model based association analysis. NA= not annotated.

^{ns} *= not significant after Bonferroni correction.*

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Supplementary Table S3

DLPD Probe	Fold Change	Chromosome	Start position	End Position	DicLab v1.0c gene matching	Gene Name	ENSEMBL Gene Danio rerio
DLPD00017_2	1,590597645	Chr15	20970199	20970575	DLAgn_00054240	NMI	ENSDARG00000100275
DLPD00197_1	2,324705036	Chr20	23369026	23369298			
DLPD00426_2	1,78668966	UN	76403499	76404027			
DLPD00651_1	1,91927223	Chr10	22371058	22371658			
DLPD00651_2	1,947980282	Chr10	22371058	22371658			
DLPD01321_2	1,918533715	Chr6	11829196	11829609			ENSDARG00000102583
DLPD01731_1	4,299209946	Chr3	5755570	5756142	DLAgn_00142300		ENSDARG00000030161
DLPD02205_2	2,068933037	Chr16	7287494	7288574	DLAgn_00058420	dedd1	
DLPD02354_2	1,417158074	Chr10	6522374	6523364	DLAgn_00002350	ppap2b	
DLPD02981_1	1,813450546	Chr1A	25729795	25730699	DLAgn_00097070	dram2b	ENSDARG00000003144
DLPD03693_2	1,391768133	Chr11	24525401	24526001	DLAgn_00017840	wbp1lb	ENSDARG00000002758
DLPD03829_1	1,906786594	Chr14	2892256	2892409	DLAgn_00037940	SAMD9L	ENSDARG00000059933
DLPD04253_1	1,483889982	Chr6	13592199	13592418	DLAgn_00169630	ube2q2	ENSDARG00000044241
DLPD04662_1	2,132698299	Chr1A	7638610	7638803	DLAgn_00089170	trnt1	ENSDARG00000029931
DLPD04913_1	3,829278326	Chr19	6891432	6892180	DLAgn_00082010	srek1	ENSDARG00000029751
DLPD04913_2	1,535296473	Chr19	6891432	6892180	DLAgn_00082010	srek1	ENSDARG00000041182
DLPD05746_1	1,404168041	Chr20	18679045	18679450	DLAgn_00122330	smarca2	ENSDARG00000102729
DLPD05802_1	2,221726186	UN	29121194	29121790	DLAgn_00225940	CU915762.1	ENSDARG00000013990
DLPD05999_1	1,373027606	Chr18-21	7145322	7146267	DLAgn_00075750	ube2w	ENSDARG00000075297
DLPD05999_2	1,384731326	Chr18-21	7145322	7146267	DLAgn_00075750	ube2w	ENSDARG00000009764
DLPD06255_2	1,467200047	Chr17	15355516	15355815	DLAgn_00071040	nrbp1	ENSDARG00000009764
DLPD07092_2	1,358449752	Chr7	21209633	21209828	DLAgn_00183040	C9orf156	ENSDARG00000038881
DLPD07125_1	1,43170625	Chr22-25	5235573	5235715	DLAgn_00127720	ccdc187	ENSDARG00000024746
DLPD07249_2	1,364635155	Chr19	2865597	2865868	DLAgn_00080160	MARCH3	ENSDARG00000052300
DLPD07275_1	1,41377275	Chr17	8212950	8213914	DLAgn_00068160	ino80	ENSDARG00000008904
DLPD07479_1	2,553447774	Chr1A	5713794	5714634	DLAgn_00088750	sox13	ENSDARG00000104657
DLPD07625_1	1,324805796	Chr4	16738088	16738701	DLAgn_00150340	mob3c	ENSDARG00000045360
DLPD08338_1	2,690675996	Chr4	8589650	8589959	DLAgn_00147260	FO704622.1	ENSDARG00000045360
DLPD08338_2	1,986881885	Chr4	8589650	8589959	DLAgn_00147260	FO704622.1	ENSDARG00000001578
DLPD08711_2	1,574655167	Chr1B	6851008	6851305	DLAgn_00101660	CR388052.1	ENSDARG00000038954
DLPD09019_1	1,761854994	Chr15	10839122	10839440	DLAgn_00049860	CU984600.2	ENSDARG00000056089
DLPD09384_1	1,670624828	Chr2	18434217	18434897	DLAgn_00112910	cplx2	ENSDARG00000037245
DLPD09755_2	1,44956987	UN	45328706	45329046			ENSDARG00000053857
DLPD09960_1	1,450543644	Chr10	15090721	15091390	DLAgn_00006760	atg9b	ENSDARG00000061738
DLPD09960_2	1,423881464	Chr10	15090721	15091390	DLAgn_00006760	atg9b	ENSDARG00000013274
DLPD10986_2	1,565353556	Chr20	1050454	1050752	DLAgn_00115000	ulk1a	ENSDARG00000070432
DLPD11030_2	1,55896513	Chr12	8803788	8803963			ENSDARG00000030297
DLPD11502_1	1,456142418	Chr4	18103761	18104172	DLAgn_00151020	creb3l3a	ENSDARG00000043705
DLPD11528_1	1,428624766	UN	7646918	7647593			ENSDARG00000089661
DLPD12398_2	1,383191926	Chr17	5236466	5237259	DLAgn_00067290	cep128	ENSDARG00000089661
DLPD12620_1	2,276034339	Chr14	12134826	12135027	DLAgn_00040920	ucp2	ENSDARG00000104919
DLPD12874_1	1,974282675	Chr15	20066117	20066329	DLAgn_00053980	hspbap1	
DLPD12983_2	1,358196394	Chr14	11694405	11694968	DLAgn_00040610	msi2b	ENSDARG00000100213
DLPD13074_2	1,88120781	Chr14	3024869	3025535	DLAgn_00037970	BX537350.1	ENSDARG00000098204
DLPD13386_2	1,806454397	UN	3629434	3629938			
DLPD14034_2	1,598755434	Chr13	10283722	10284162			ENSDARG00000061918
DLPD14249_2	1,785121078	Chr15	10852425	10852824	DLAgn_00049860	CU984600.2	
DLPD14889_1	1,515376977	Chr11	5326123	5326271	DLAgn_00011390	arid4b	ENSDARG00000055292
DLPD15054_1	1,769446175	Chr14	27369276	27369959			ENSDARG00000055292
DLPD15254_2	1,672765514	Chr22-25	619824	620634			ENSDARG00000074481
DLPD15641_2	1,480863044	Chr1B	14700894	14701090	DLAgn_00104810	dhx8	
DLPD15747_2	1,534291515	Chr1A	2922628	2922876	DLAgn_00088020	LAMB3	ENSDARG00000092638
DLPD16029_2	1,468193594	Chr10	1044304	1045024	DLAgn_00000340		ENSDARG00000056226
DLPD16329_2	2,151472657	Chr7	13293836	13294570	DLAgn_00179660		
DLPD16438_2	1,73115576	Chr2	16716479	16716945	DLAgn_00112800	irf2	ENSDARG00000103771
DLPD16493_2	1,487990261	Chr10	19111889	19112789	DLAgn_00008370	ankrd12	
DLPD16796_1	1,446114594	Chr4	6851776	6852427	DLAgn_00146660	ak4	ENSDARG00000091209
DLPD17090_1	1,447875483	Chr17	18690424	18690649	DLAgn_00072400	E2F6	ENSDARG00000033771
DLPD17668_1	1,369833074	Chr19	6907497	6908190	DLAgn_00082020	erbb2ip	ENSDARG00000032614
DLPD17807_1	1,93889864	Chr8	17694012	17694284			ENSDARG00000073789
DLPD17807_2	1,876417155	Chr8	17694012	17694284			
DLPD17839_1	1,617946988	Chr14	22169679	22169945			ENSDARG00000037101
DLPD17931_1	2,192654968	Chr3	4996104	4996277	DLAgn_00142050	BX005336.3	
DLPD17931_2	2,466412009	Chr3	4996104	4996277	DLAgn_00142050	BX005336.3	
DLPD18116_1	1,638746966	Chr19	14741085	14741805	DLAgn_00085040	aak1b	ENSDARG00000098204
DLPD18116_2	1,620604163	Chr19	14741085	14741805	DLAgn_00085040	aak1b	ENSDARG00000090656
DLPD18435_2	1,589106181	UN	35084119	35085004			
DLPD18515_2	1,317693225	Chr12	15953151	15953899	DLAgn_00023990	sp3b	
DLPD18828_1	1,652489418	Chr14	27580825	27580946	DLAgn_00046590	bckdk	ENSDARG00000054707
DLPD18828_2	1,611399319	Chr14	27580825	27580946	DLAgn_00046590	bckdk	ENSDARG00000059369
DLPD00831_2	0,739054299	UN	72517428	72517591	DLAgn_00251780	il4r.1	ENSDARG00000101547
DLPD01215_2	0,693860249	Chr3	6372829	6373413	DLAgn_00142540	ppp1r14bb	ENSDARG00000008100
DLPD01958_1	0,726253184	Chr20	3875089	3875382	DLAgn_00116240	pxmp2	ENSDARG00000089456
DLPD03440_1	0,729871333	Chr8	10595571	10595767	DLAgn_00191370	atp13a1	
DLPD03790_1	0,702900416	Chr6	12211852	12212023	DLAgn_00168650	rpl4	ENSDARG00000040465
DLPD04920_1	0,702083907	Chr20	14199644	14199917	DLAgn_00120530	acaa2	ENSDARG00000052419
DLPD05074_2	0,653591638	Chr17	7856588	7856919	DLAgn_00068070	hsp90aa1.2	ENSDARG00000006546
DLPD05677_1	0,611686921	Chr8	7738218	7738720	DLAgn_00189830	zgc:171489	ENSDARG00000008119
DLPD06138_2	0,735545799	Chrx	16301268	16301468	DLAgn_00209780	ECHDC3	ENSDARG00000011975
DLPD06159_1	0,679450311	Chr22-25	14664693	14664987	DLAgn_00132130	ctnnbip1	ENSDARG00000044281

DLPD07264_1	0,689940353	Chr12	6342290	6342672	DLAgn_00019800	eif4eb	ENSDARG00000052279
DLPD08438_1	0,707160502	Chr2	10663969	10664160	DLAgn_00110480		
DLPD08665_1	0,654064651	Chr3	9666798	9667499			
DLPD09239_1	0,576048048	Chr10	13287342	13287959			
DLPD11346_2	0,71480324	Chr9	4691915	4692073	DLAgn_00198170	ptprua	ENSDARG00000078731
DLPD12419_2	0,551789689	Chr11	4632593	4632829	DLAgn_00011200		ENSDARG00000078731
DLPD13502_1	0,581596109	Chr2	14732527	14733195	DLAgn_00112060	mtmr7a	ENSDARG00000077686
DLPD13780_1	0,634120531	Chr2	3617862	3618729			ENSDARG00000077686
DLPD16049_2	0,689140696	Chr5	3059119	3059738	DLAgn_00154950	slc7a10a	
DLPD16296_1	0,740020438	Chr3	1223826	1224436	DLAgn_00141050	men1	ENSDARG00000007812
DLPD17413_2	0,720266438	Chr12	9462761	9463818	DLAgn_00021070	si:ch1073-209e23.1	ENSDARG00000016904
DLPD17793_2	0,540082495	Chr1A	26827298	26828089	DLAgn_00097630	osgn1	ENSDARG00000016904

Supplementary Table S3: Differentially expressed genes between normal and jaw deformed larvae (whole-head; 38 dph).

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Supplementary Table S4

DLPD Probe	Fold Change	Chromosome	Start position	End Position	DicLab v1.0c gene matching	Gene Name	ENSEMBL Protein Danio rerio	ENSEMBL Gene Danio rerio
DLPD18962_1	2,281112784	Chr20	13746580	13746812				
DLPD17733_2	2,184714137	Chr14	20045557	20046066	DLAgn_00045130	sowahaa	ENSDARP00000100797	ENSDARG00000079125
DLPD16540_1	2,072317977	Chr12	5541866	5542054				
DLPD16070_2	2,059476708	Chr12	11348564	11348833	DLAgn_00022140	plk4	ENSDARP00000043865	ENSDARG00000004576
DLPD15637_1	2,902910977	Chr13	4435706	4436075	DLAgn_00028520	tcnl	ENSDARP00000089045	ENSDARG00000068088
DLPD15100_1	7,039198663	Chr16	5607922	5608371				
DLPD14666_1	2,419169724	Chr14	15806854	15807458	DLAgn_00042990	slc7a1	ENSDARP00000121637	ENSDARG00000016439
DLPD14597_1	4,378108131	Chr16	9554331	9554801	DLAgn_00060240	efna1b	ENSDARP00000012577	ENSDARG00000018787
DLPD11760_2	2,059068889	Chr17	11554689	11555312	DLAgn_00069730	efr3bb	ENSDARP00000112891	ENSDARG00000069318
DLPD11720_2	2,039978629	UN	96683300	96683611				
DLPD11720_1	2,012619229	UN	96683300	96683611				
DLPD11437_2	2,001842806	UN	90715063	90715319	DLAgn_00263060	ceacam1	ENSDARP00000060265	ENSDARG00000041119
DLPD09471_2	2,05325258	Chr5	25965167	25965547				
DLPD09471_1	2,187556261	Chr5	25965167	25965547				
DLPD09199_2	2,075739317	Chr16	22117273	22117400	DLAgn_00065190	csad	ENSDARP00000024113	ENSDARG00000026348
DLPD09199_1	2,129143354	Chr16	22117273	22117400	DLAgn_00065190	csad	ENSDARP00000024113	ENSDARG00000026348
DLPD08935_2	2,057197218	UN	2184623	2184788				
DLPD08902_2	2,05039714	Chr22-25	23622759	23623312	DLAgn_00135170	stk38a	ENSDARP00000003411	ENSDARG00000019973
DLPD08672_1	2,105229073	Chr17	9815850	9816550				
DLPD08545_1	2,036066235	UN	22262182	22262403	DLAgn_00222250	SH3TC1	ENSDARP00000102920	ENSDARG00000077330
DLPD08224_2	2,255001575	Chr4	5122195	5122397				
DLPD19035_2	0,401936848	Chr6	12045889	12046590	DLAgn_00168550	C18H15orf39	ENSDARP00000091231	ENSDARG00000069168
DLPD19009_2	0,48534171	UN	94619171	94619358	DLAgn_00265650			
DLPD18855_1	0,412000516	Chr6	24237997	24238659	DLAgn_00173070	wt1a	ENSDARP00000029174	ENSDARG00000031420
DLPD18593_2	0,177912503	Chr20	4316979	4317583	DLAgn_00116580	ivd	ENSDARP00000062893	ENSDARG00000042853
DLPD18470_1	0,481780535	Chr22-25	12206737	12207636	DLAgn_00130720	RARG	ENSDARP00000049550	ENSDARG00000034117
DLPD17880_2	0,383559303	Chr13	16475140	16475318	DLAgn_00032310	scn3b	ENSDARP00000138558	ENSDARG00000062359
DLPD17842_2	0,26921402	Chr16	9082496	9082799				
DLPD17842_1	0,315290871	Chr16	9082496	9082799				
DLPD17626_2	0,459539863	UN	3760288	3760352				
DLPD17619_2	0,337678543	Chr19	9178610	9178749	DLAgn_00082660	zgc:55582	ENSDARP00000028290	ENSDARG00000024219
DLPD17619_1	0,319097902	Chr19	9178610	9178749	DLAgn_00082660	zgc:55582	ENSDARP00000028290	ENSDARG00000024219
DLPD17317_2	0,169324296	Chr3	9453176	9453585				
DLPD17317_1	0,151306318	Chr3	9453176	9453585				
DLPD17262_2	0,373864835	Chr3	3766240	3766639				
DLPD17262_1	0,444187564	Chr3	3766240	3766639				
DLPD17012_1	0,465540009	Chr15	12630685	12631055	DLAgn_00050600	lrp2a	ENSDARP00000131232	ENSDARG00000102506
DLPD16995_2	0,420933796	Chr1B	6318684	6318863	DLAgn_00101140	dnaaf3	ENSDARP00000073689	ENSDARG00000036304
DLPD16977_1	0,467418489	UN	8066642	8066723				
DLPD16691_2	0,478276351	UN	64051200	64051347				
DLPD16667_2	0,212417235	Chr14	13313502	13313903				
DLPD16667_1	0,208919322	Chr14	13313502	13313903				
DLPD16563_2	0,42315334	Chr2	3144534	3145084	DLAgn_00106980	glra4a	ENSDARP00000030826	ENSDARG00000006865
DLPD16516_2	0,005758356	Chr22-25	8522456	8523862	DLAgn_00129180	rho	ENSDARP00000011562	ENSDARG00000002193
DLPD16516_1	0,006552321	Chr22-25	8522456	8523862	DLAgn_00129180	rho	ENSDARP00000011562	ENSDARG00000002193
DLPD16193_1	0,40417086	Chr9	11819986	11820231	DLAgn_00200470	smarcc1b	ENSDARP00000099015	ENSDARG00000077946
DLPD16163_1	0,455837394	Chr5	8100196	8101000	DLAgn_00156150	GNRHR	ENSDARP00000055566	ENSDARG00000038116
DLPD16121_2	0,482586127	Chr9	10843464	10843798	DLAgn_00199940	npy	ENSDARP00000052619	ENSDARG00000036222
DLPD16121_1	0,455164683	Chr9	10843464	10843798	DLAgn_00199940	npy	ENSDARP00000052619	ENSDARG00000036222
DLPD15977_2	0,499315943	Chr7	8721293	8721768	DLAgn_00177360	MPV17L	ENSDARP00000140607	ENSDARG00000104457
DLPD15956_2	0,295227357	Chr16	6011819	6012000	DLAgn_00057760		ENSDARP00000124439	ENSDARG00000095831
DLPD15956_1	0,111768395	Chr16	6011819	6012000	DLAgn_00057760		ENSDARP00000124439	ENSDARG00000095831
DLPD15545_1	0,463870486	Chr19	8778347	8779181				
DLPD15525_2	0,445190222	Chr18-21	8368763	8369739				
DLPD15523_2	0,326518366	Chr10	1803620	1804672				
DLPD15283_2	0,498735057	Chr16	2844919	2845176	DLAgn_00056450	notchl	ENSDARP00000128794	ENSDARG00000088308
DLPD15179_1	0,493388217	UN	91252812	91253140	DLAgn_00263390	hhipl1	ENSDARP00000084880	ENSDARG00000062501
DLPD14727_1	0,466846307	Chr12	21724612	21724876				
DLPD14707_2	0,490630703	UN	11318261	11318857	DLAgn_00217470	dmtn	ENSDARP00000141261	ENSDARG00000013110
DLPD14332_2	0,138620849	Chr18-21	13288267	13288663				
DLPD14332_1	0,144251921	Chr18-21	13288267	13288663				
DLPD14250_1	0,436282377	Chr8	6165849	6166062	DLAgn_00189040	si:dkey-10p5.7	ENSDARP00000133532	ENSDARG00000098787
DLPD14194_2	0,233141118	Chr24	8188193	8188677				
DLPD14193_2	0,298589068	Chr1A	16208609	16209123	DLAgn_00093140	GRM2	ENSDARP00000079907	ENSDARG00000004150
DLPD14193_1	0,254923873	Chr1A	16208609	16209123	DLAgn_00093140	GRM2	ENSDARP00000079907	ENSDARG00000004150
DLPD14171_2	0,27405313	Chr6	21838234	21838276				
DLPD14171_1	0,278695514	Chr6	21838234	21838276				
DLPD14142_2	0,373604087	Chr10	8903722	8904362	DLAgn_00003990	rab3ab	ENSDARP00000095064	ENSDARG00000043835
DLPD14142_1	0,390298688	Chr10	8903722	8904362	DLAgn_00003990	rab3ab	ENSDARP00000095064	ENSDARG00000043835
DLPD14130_2	0,173344121	Chr9	3241062	3241406				
DLPD14130_1	0,18511904	Chr9	3241062	3241406				
DLPD14082_1	0,460910707	Chr17	10997651	10998308	DLAgn_00069440	si:ch211-153j24	ENSDARP00000058705	ENSDARG00000040135
DLPD14047_2	0,016518861	Chr14	18776752	18776887	DLAgn_00044660	arr3a	ENSDARP00000073453	ENSDARG00000056511
DLPD14047_1	0,016916166	Chr14	18776752	18776887	DLAgn_00044660	arr3a	ENSDARP00000073453	ENSDARG00000056511
DLPD14043_2	0,357593083	UN	61878466	61879129	DLAgn_00244860	cdk5r2a	ENSDARP00000104135	ENSDARG00000071011
DLPD14043_1	0,399888689	UN	61878466	61879129	DLAgn_00244860	cdk5r2a	ENSDARP00000104135	ENSDARG00000071011
DLPD14041_2	0,402826823	Chr6	8890134	8890871	DLAgn_00167290	lrrc4.2	ENSDARP00000005494	ENSDARG00000003020
DLPD14032_2	0,452079905	Chr22-25	15641411	15642251				
DLPD14032_1	0,421220001	Chr22-25	15641411	15642251				
DLPD14027_2	0,226140063	Chr1A	22823142	22823588				

DLPD14027_1	0,376174028	Chr1A	22823142	22823588				
DLPD14006_2	0,455043591	Chr9	7459612	7460258	DLAgn_00198790	si:dkey-261i16.!	ENSDARP00000093751	ENSDARG00000044719
DLPD13972_2	0,047835871	Chr1A	23291255	23291831	DLAgn_00096180			
DLPD13972_1	0,060203849	Chr1A	23291255	23291831	DLAgn_00096180			
DLPD13969_2	0,261147556	Chr4	10300588	10301143				
DLPD13969_1	0,214821924	Chr4	10300588	10301143				
DLPD13941_2	0,445875782	Chr7	5115446	5116178	DLAgn_00176030	rims1b	ENSDARP00000081549	ENSDARG00000078902
DLPD13941_1	0,317166754	Chr7	5115446	5116178	DLAgn_00176030	rims1b	ENSDARP00000081549	ENSDARG00000078902
DLPD13933_2	0,478049096	Chr14	7476985	7477909				
DLPD13910_2	0,373047703	Chr22-25	19102103	19102612				
DLPD13910_1	0,484216356	Chr22-25	19102103	19102612				
DLPD13874_2	0,293657686	Chr7	10356626	10357149				
DLPD13857_2	0,294282213	Chr16	9079318	9079727				
DLPD13857_1	0,292022884	Chr16	9079318	9079727				
DLPD13847_2	0,251178024	Chr6	25261438	25262182	DLAgn_00173440	ndrg4	ENSDARP00000141868	ENSDARG00000103937
DLPD13847_1	0,245250382	Chr6	25261438	25262182	DLAgn_00173440	ndrg4	ENSDARP00000141868	ENSDARG00000103937
DLPD13844_2	0,338853748	Chr6	9021841	9022183				
DLPD13844_1	0,408366001	Chr6	9021841	9022183				
DLPD13831_2	0,490523971	Chr17	11806105	11806798				
DLPD13824_1	0,283776606	Chr24	371377	371972	DLAgn_00136360	IQSEC2	ENSDARP00000141504	ENSDARG00000102125
DLPD13809_2	0,34842542	Chr9	14992827	14994083				
DLPD13774_2	0,263578537	Chr24	11556598	11557036				
DLPD13774_1	0,367102586	Chr24	11556598	11557036				
DLPD13755_2	0,492764217	Chr20	13977229	13977745	DLAgn_00120340	KCNIP3	ENSDARP00000130174	ENSDARG00000017880
DLPD13755_1	0,495020083	Chr20	13977229	13977745	DLAgn_00120340	KCNIP3	ENSDARP00000130174	ENSDARG00000017880
DLPD13748_2	0,449530878	Chr13	23936296	23936837				
DLPD13748_1	0,357516911	Chr13	23936296	23936837				
DLPD13747_2	0,419265304	Chr8	1535632	1536086	DLAgn_00186770	cdr2l	ENSDARP00000105648	ENSDARG00000026834
DLPD13735_2	0,264054823	Chr12	19968589	19969521	DLAgn_00026080	ccdc177	ENSDARP00000106528	ENSDARG00000067841
DLPD13733_2	0,157368575	Chr10	12428146	12428317	DLAgn_00005620	rab6bb	ENSDARP00000042254	ENSDARG00000031343
DLPD13706_2	0,373653313	Chr22-25	4924184	4924414				
DLPD13698_2	0,265215291	Chr15	13321421	13321929	DLAgn_00051010	CHN1	ENSDARP00000130201	ENSDARG00000101735
DLPD13698_1	0,32937609	Chr15	13321421	13321929	DLAgn_00051010	CHN1	ENSDARP00000130201	ENSDARG00000101735
DLPD13687_2	0,286595974	UN	4854807	4855477				
DLPD13687_1	0,308495826	UN	4854807	4855477				
DLPD13655_2	0,473003253	Chrx	8258690	8259017	DLAgn_00207580	LRRN3	ENSDARP00000131085	ENSDARG00000102702
DLPD13642_2	0,153980345	Chr16	11172237	11172730	DLAgn_00061250	stmn2a	ENSDARP00000094591	ENSDARG00000033234
DLPD13642_1	0,134834863	Chr16	11172237	11172730	DLAgn_00061250	stmn2a	ENSDARP00000094591	ENSDARG00000033234
DLPD13614_2	0,371589868	Chr9	912230	912752				
DLPD13612_2	0,235150484	Chr15	10380403	10381096				
DLPD13569_2	0,394839705	Chr14	13713621	13714341				
DLPD13559_2	0,415279082	Chr13	6445008	6445147	DLAgn_00029100	scg2a	ENSDARP00000131009	ENSDARG00000103993
DLPD13559_1	0,208206724	Chr13	6445008	6445147	DLAgn_00029100	scg2a	ENSDARP00000131009	ENSDARG00000103993
DLPD13549_2	0,127527132	Chr2	10274942	10275261				
DLPD13549_1	0,140038669	Chr2	10274942	10275261				
DLPD13487_2	0,253583592	Chr1A	7722918	7723474	DLAgn_00089190	lrrn1	ENSDARP00000078824	ENSDARG00000060115
DLPD13487_1	0,207682864	Chr1A	7722918	7723474	DLAgn_00089190	lrrn1	ENSDARP00000078824	ENSDARG00000060115
DLPD13465_1	0,083976661	Chr5	9378581	9379067				
DLPD13460_2	0,249863683	Chr8	6914310	6915068				
DLPD13460_1	0,257302043	Chr8	6914310	6915068				
DLPD13456_2	0,1616471	Chr2	10274108	10274438				
DLPD13456_1	0,151506148	Chr2	10274108	10274438				
DLPD13437_2	0,210078421	UN	37983699	37983970				
DLPD13418_2	0,432940719	Chr10	17202679	17202960	DLAgn_00007700	RALYL	ENSDARP00000105963	ENSDARG00000090292
DLPD13418_1	0,401951535	Chr10	17202679	17202960	DLAgn_00007700	RALYL	ENSDARP00000105963	ENSDARG00000090292
DLPD13407_2	0,474174135	Chr12	18397205	18398003				
DLPD13403_1	0,374396539	Chr12	12200596	12200772	DLAgn_00022680	cpsf3	ENSDARP00000040291	ENSDARG00000027649
DLPD13390_2	0,439545193	Chr6	13239389	13239602	DLAgn_00169360	gnb5a	ENSDARP00000140306	ENSDARG00000099685
DLPD13390_1	0,37613161	Chr6	13239389	13239602	DLAgn_00169360	gnb5a	ENSDARP00000140306	ENSDARG00000099685
DLPD13384_2	0,441477163	Chr13	16380529	16380999				
DLPD13377_2	0,316517752	Chr11	22919691	22920080				
DLPD13372_2	0,245015876	Chr1A	14332572	14332697	DLAgn_00092020	prph	ENSDARP00000129087	ENSDARG00000028306
DLPD13372_1	0,201921333	Chr1A	14332572	14332697	DLAgn_00092020	prph	ENSDARP00000129087	ENSDARG00000028306
DLPD13368_2	0,222271409	Chr19	10518018	10518674	DLAgn_00083380	mapk10	ENSDARP00000134326	ENSDARG00000102730
DLPD13368_1	0,200962105	Chr19	10518018	10518674	DLAgn_00083380	mapk10	ENSDARP00000134326	ENSDARG00000102730
DLPD13360_1	0,235401791	Chr19	12092984	12093272	DLAgn_00083920	BRINP1	ENSDARP00000100708	ENSDARG00000078302
DLPD13356_2	0,295562056	Chr1B	11617671	11618289				
DLPD13356_1	0,301033629	Chr1B	11617671	11618289				
DLPD13344_2	0,17116365	Chr3	667719	668131				
DLPD13343_2	0,286809688	Chr5	28341282	28341712				
DLPD13343_1	0,275409041	Chr5	28341282	28341712				
DLPD13293_1	0,399017782	Chr20	13950555	13950729	DLAgn_00120340	KCNIP3	ENSDARP00000130174	ENSDARG00000017880
DLPD13292_2	0,393157909	Chr12	18804301	18804740				
DLPD13292_1	0,411742266	Chr12	18804301	18804740				
DLPD13240_2	0,294893817	Chr15	18531614	18531735	DLAgn_00053500	nalcn	ENSDARP00000048468	ENSDARG00000001835
DLPD13240_1	0,31219325	Chr15	18531614	18531735	DLAgn_00053500	nalcn	ENSDARP00000048468	ENSDARG00000001835
DLPD13053_2	0,463981084	UN	22799697	22801438	DLAgn_00222540	myh6	ENSDARP00000108536	ENSDARG00000090637
DLPD13009_2	0,221185182	UN	50762997	50763262				
DLPD13009_1	0,215017486	UN	50762997	50763262				
DLPD12893_2	0,332787533	Chr20	27809791	27810696				
DLPD12889_2	0,101535423	Chr1A	19539269	19539641				

DLPD12889_1	0,071564675	Chr1A	19539269	19539641				
DLPD12789_2	0,323423115	Chrx	13921100	13921469	DLAgn_00209210	ptn	ENSDARP00000140433	ENSDARG00000102340
DLPD12789_1	0,342066328	Chrx	13921100	13921469	DLAgn_00209210	ptn	ENSDARP00000140433	ENSDARG00000102340
DLPD12774_2	0,413732034	Chr5	8562553	8562938				
DLPD12736_2	0,274342253	Chr22-25	14045412	14045643	DLAgn_00131830	gnat2	ENSDARP00000062362	ENSDARG00000042529
DLPD12736_1	0,42085023	Chr22-25	14045412	14045643	DLAgn_00131830	gnat2	ENSDARP00000062362	ENSDARG00000042529
DLPD12730_1	0,495502599	Chr5	9420201	9421182				
DLPD12549_2	0,322245423	Chr22-25	15664449	15664671	DLAgn_00132450	necab3	ENSDARP00000127670	ENSDARG00000074794
DLPD12549_1	0,316029738	Chr22-25	15664449	15664671	DLAgn_00132450	necab3	ENSDARP00000127670	ENSDARG00000074794
DLPD12547_1	0,462481694	Chr18-21	13067039	13067411				
DLPD12473_2	0,187824368	UN	21175712	21176454	DLAgn_00221790			
DLPD12434_1	0,294233493	Chr4	22430065	22430786				
DLPD12238_2	0,48235377	Chr22-25	8064649	8064862	DLAgn_00128890	C9orf89	ENSDARP00000107905	ENSDARG00000091304
DLPD12238_1	0,436688762	Chr22-25	8064649	8064862	DLAgn_00128890	C9orf89	ENSDARP00000107905	ENSDARG00000091304
DLPD11935_2	0,388389038	Chr13	27559412	27560028				
DLPD11935_1	0,399789483	Chr13	27559412	27560028				
DLPD11865_2	0,138797913	Chr19	19613042	19613928	DLAgn_00086650	phf24	ENSDARP00000099941	ENSDARG00000077596
DLPD11712_2	0,335068878	UN	59051181	59051304	DLAgn_00243390	zmat2	ENSDARP00000002108	ENSDARG00000004956
DLPD11576_2	0,304068958	Chr6	13620823	13620992	DLAgn_00169650	nptna	ENSDARP00000064402	ENSDARG00000043864
DLPD11576_1	0,431580301	Chr6	13620823	13620992	DLAgn_00169650	nptna	ENSDARP00000064402	ENSDARG00000043864
DLPD11565_2	0,424233778	Chr2	20720215	20720411	DLAgn_00113400	immt	ENSDARP00000133453	ENSDARG00000102874
DLPD11539_2	0,095209795	Chr13	17575835	17576366	DLAgn_00032770	ppm1na	ENSDARP00000011222	ENSDARG00000010231
DLPD11539_1	0,117260217	Chr13	17575835	17576366	DLAgn_00032770	ppm1na	ENSDARP00000011222	ENSDARG00000010231
DLPD11486_2	0,451368201	Chr4	20779001	20779247				
DLPD11486_1	0,454348641	Chr4	20779001	20779247				
DLPD11145_2	0,46335089	Chr18-21	14392065	14392327				
DLPD10879_2	0,199538592	Chr7	16205440	16205816	DLAgn_00181100			
DLPD10879_1	0,193961763	Chr7	16205440	16205816	DLAgn_00181100			
DLPD10848_1	0,438541917	Chr6	8153487	8154124	DLAgn_00166970	mafa	ENSDARP00000101942	ENSDARG00000076520
DLPD10724_2	0,268345403	Chr14	20423599	20423698	DLAgn_00045170	GABRA2	ENSDARP00000134688	ENSDARG00000068989
DLPD10724_1	0,220347903	Chr14	20423599	20423698	DLAgn_00045170	GABRA2	ENSDARP00000134688	ENSDARG00000068989
DLPD10718_2	0,112936324	Chr10	4306986	4307766	DLAgn_00001200	sepp1b	ENSDARP00000140653	ENSDARG00000079727
DLPD10718_1	0,148962547	Chr10	4306986	4307766	DLAgn_00001200	sepp1b	ENSDARP00000140653	ENSDARG00000079727
DLPD10636_2	0,373839244	Chr19	5189798	5190203	DLAgn_00081270	PLA2G3	ENSDARP00000023708	ENSDARG00000008948
DLPD10636_1	0,351403029	Chr19	5189798	5190203	DLAgn_00081270	PLA2G3	ENSDARP00000023708	ENSDARG00000008948
DLPD10568_2	0,378880251	Chr16	7460598	7460772				
DLPD10568_1	0,42766973	Chr16	7460598	7460772				
DLPD10546_2	0,377895629	no match	no match	no match				
DLPD10546_1	0,275350971	no match	no match	no match				
DLPD10368_2	0,424346621	UN	34570238	34570654	DLAgn_00229490			
DLPD10364_2	0,427823303	Chr4	2789708	2789848	DLAgn_00145690	fetub	ENSDARP00000070561	ENSDARG00000053973
DLPD10342_1	0,484848573	Chr9	20385943	20386088	DLAgn_00204050	slc9a3.2	ENSDARP00000109287	ENSDARG00000058498
DLPD10315_2	0,168396084	Chr13	8023243	8023976	DLAgn_00029710	apoa1a	ENSDARP00000025613	ENSDARG00000012076
DLPD10315_1	0,167402103	Chr13	8023243	8023976	DLAgn_00029710	apoa1a	ENSDARP00000025613	ENSDARG00000012076
DLPD10234_2	0,494704867	no match	no match	no match				
DLPD10106_2	0,320136319	Chr10	8904540	8905150	DLAgn_00003990	rab3ab	ENSDARP00000095064	ENSDARG00000043835
DLPD10106_1	0,317040248	Chr10	8904540	8905150	DLAgn_00003990	rab3ab	ENSDARP00000095064	ENSDARG00000043835
DLPD10001_2	0,226490117	Chr10	11883670	11884066	DLAgn_00005480	zgc:153615	ENSDARP00000137287	ENSDARG00000038981
DLPD10001_1	0,315143043	Chr10	11883670	11884066	DLAgn_00005480	zgc:153615	ENSDARP00000137287	ENSDARG00000038981
DLPD09989_2	0,23329664	Chr1A	12648321	12648516				
DLPD09989_1	0,218263687	Chr1A	12648321	12648516				
DLPD09979_2	0,242946597	Chr22-25	24579112	24579635				
DLPD09979_1	0,244743573	Chr22-25	24579112	24579635				
DLPD09959_2	0,404413864	Chr1A	11551214	11551554	DLAgn_00090760	hmp19	ENSDARP00000140381	ENSDARG00000102975
DLPD09959_1	0,382471952	Chr1A	11551214	11551554	DLAgn_00090760	hmp19	ENSDARP00000140381	ENSDARG00000102975
DLPD09909_2	0,11886338	Chr22-25	4831373	4831799				
DLPD09909_1	0,322979394	Chr22-25	4831373	4831799				
DLPD09893_2	0,31788245	Chr14	12513449	12513699	DLAgn_00041100	nlgn3b	ENSDARP00000134245	ENSDARG00000104786
DLPD09893_1	0,323134986	Chr14	12513449	12513699	DLAgn_00041100	nlgn3b	ENSDARP00000134245	ENSDARG00000104786
DLPD09843_1	0,497437991	Chr5	22774038	22774372	DLAgn_00161190	GRAMD2	ENSDARP00000130256	ENSDARG00000103736
DLPD09786_2	0,201557023	Chr20	7263785	7264031				
DLPD09786_1	0,280595281	Chr20	7263785	7264031				
DLPD09774_2	0,132864924	Chr7	3366451	3367055				
DLPD09774_1	0,129882682	Chr7	3366451	3367055				
DLPD09731_2	0,312913894	Chr1B	12032155	12032321	DLAgn_00103730	myoz1b	ENSDARP00000096413	ENSDARG00000071445
DLPD09731_1	0,281017496	Chr1B	12032155	12032321	DLAgn_00103730	myoz1b	ENSDARP00000096413	ENSDARG00000071445
DLPD09667_2	0,138040788	Chr20	23379358	23379663				
DLPD09667_1	0,139281083	Chr20	23379358	23379663				
DLPD09506_2	0,338745596	Chr11	17599089	17599461	DLAgn_00016070	eef1a1b	ENSDARP00000069823	ENSDARG00000069951
DLPD09506_1	0,348764174	Chr11	17599089	17599461	DLAgn_00016070	eef1a1b	ENSDARP00000069823	ENSDARG00000069951
DLPD09488_2	0,275739167	Chr22-25	19103998	19104310	DLAgn_00133780	gabrd	ENSDARP00000077851	ENSDARG00000059763
DLPD09488_1	0,298325105	Chr22-25	19103998	19104310	DLAgn_00133780	gabrd	ENSDARP00000077851	ENSDARG00000059763
DLPD09425_2	0,401030847	Chr22-25	7723534	7723646				
DLPD09425_1	0,39583736	Chr22-25	7723534	7723646				
DLPD09409_2	0,313765019	UN	48047660	48048494				
DLPD09384_2	0,411391943	Chr2	18434217	18434897	DLAgn_00112910	cplx2	ENSDARP00000137988	ENSDARG00000061918
DLPD09384_1	0,443016407	Chr2	18434217	18434897	DLAgn_00112910	cplx2	ENSDARP00000137988	ENSDARG00000061918
DLPD09376_2	0,176331914	Chr8	21132483	21132897				
DLPD09376_1	0,18177667	Chr8	21132483	21132897				
DLPD09361_2	0,459170932	Chr9	11005223	11005363	DLAgn_00200010	dync1i1	ENSDARP00000138576	ENSDARG00000060948
DLPD09361_1	0,413570934	Chr9	11005223	11005363	DLAgn_00200010	dync1i1	ENSDARP00000138576	ENSDARG00000060948

DLPD09304_2	0,264100879	Chr9	14697474	14697843				
DLPD09304_1	0,287425176	Chr9	14697474	14697843				
DLPD09268_2	0,368856457	Chr14	10566993	10567100	DLAgn_00040450	pcdh2ac	ENSDARP00000130311	ENSDARG00000099783
DLPD09268_1	0,367714199	Chr14	10566993	10567100	DLAgn_00040450	pcdh2ac	ENSDARP00000130311	ENSDARG00000099783
DLPD09256_1	0,440745958	Chr5	20059509	20059802	DLAgn_00160120	wdsub1	ENSDARP00000106094	ENSDARG00000090418
DLPD09244_2	0,116886083	Chr10	5938315	5938443				
DLPD09244_1	0,111865057	Chr10	5938315	5938443				
DLPD09234_2	0,49987136	Chr1B	1574425	1574686	DLAgn_00098610	ADAP1	ENSDARP00000125714	ENSDARG00000038151
DLPD09223_2	0,20456554	no match	no match	no match				
DLPD09223_1	0,192739245	no match	no match	no match				
DLPD09207_2	0,400948752	Chr6	11161798	11162034	DLAgn_00168100	ETFA	ENSDARP00000136146	ENSDARG00000101631
DLPD09207_1	0,358488635	Chr6	11161798	11162034	DLAgn_00168100	ETFA	ENSDARP00000136146	ENSDARG00000101631
DLPD09152_1	0,374610051	Chr3	5738970	5739221	DLAgn_00142280	syt4	ENSDARP00000045478	ENSDARG00000036505
DLPD09137_2	0,229368796	Chr20	17442050	17442288				
DLPD09137_1	0,227922876	Chr20	17442050	17442288				
DLPD09116_2	0,440371393	UN	20066625	20067263				
DLPD09116_1	0,408478844	UN	20066625	20067263				
DLPD09092_2	0,301246494	UN	19312324	19312786				
DLPD09092_1	0,30100905	UN	19312324	19312786				
DLPD09073_2	0,223462101	Chr17	14757852	14758169				
DLPD09073_1	0,248158102	Chr17	14757852	14758169				
DLPD09043_2	0,269253221	Chr16	8640885	8641342	DLAgn_00059550	josd2	ENSDARP00000074792	ENSDARG00000057626
DLPD09043_1	0,298919227	Chr16	8640885	8641342	DLAgn_00059550	josd2	ENSDARP00000074792	ENSDARG00000057626
DLPD08981_2	0,169021765	no match	no match	no match				
DLPD08981_1	0,144492786	no match	no match	no match				
DLPD08964_2	0,187353522	UN	64710880	64711495				
DLPD08964_1	0,221199665	UN	64710880	64711495				
DLPD08957_2	0,309645838	Chr20	7851939	7852178				
DLPD08957_1	0,321573027	Chr20	7851939	7852178				
DLPD08923_2	0,319623778	Chrx	2097606	2097962				
DLPD08923_1	0,31100536	Chrx	2097606	2097962				
DLPD08916_2	0,314865569	Chr1A	17215967	17216122	DLAgn_00093880	cadpsb	ENSDARP00000042682	ENSDARG00000070567
DLPD08916_1	0,421014042	Chr1A	17215967	17216122	DLAgn_00093880	cadpsb	ENSDARP00000042682	ENSDARG00000070567
DLPD08910_2	0,19594883	Chr22-25	5871125	5871263				
DLPD08910_1	0,178496258	Chr22-25	5871125	5871263				
DLPD08904_1	0,498942348	Chr6	11936808	11936937	DLAgn_00168470	islr2	ENSDARP00000068073	ENSDARG00000051875
DLPD08852_1	0,273097808	Chr1A	16009418	16009874				
DLPD08849_2	0,45402011	Chr22-25	7723251	7723533				
DLPD08849_1	0,432200303	Chr22-25	7723251	7723533				
DLPD08826_2	0,392064455	Chr5	15891792	15892253	DLAgn_00158880	calca	ENSDARP00000091761	ENSDARG00000056590
DLPD08826_1	0,397122389	Chr5	15891792	15892253	DLAgn_00158880	calca	ENSDARP00000091761	ENSDARG00000056590
DLPD08781_1	0,267533657	Chr12	11244054	11244332	DLAgn_00022100			
DLPD08736_2	0,339661171	Chr6	13622846	13623019	DLAgn_00169650	nptna	ENSDARP00000064402	ENSDARG00000043864
DLPD08736_1	0,332889586	Chr6	13622846	13623019	DLAgn_00169650	nptna	ENSDARP00000064402	ENSDARG00000043864
DLPD08634_2	0,236907429	Chr17	19179599	19180209	DLAgn_00072580	zgc:101840	ENSDARP00000040556	ENSDARG00000033201
DLPD08634_1	0,252208495	Chr17	19179599	19180209	DLAgn_00072580	zgc:101840	ENSDARP00000040556	ENSDARG00000033201
DLPD08618_2	0,464129095	Chr1A	15315409	15316164	DLAgn_00092620	apof	ENSDARP00000110505	ENSDARG00000090980
DLPD08601_2	0,106178639	Chr11	19962139	19962635				
DLPD08601_1	0,127842345	Chr11	19962139	19962635				
DLPD08570_2	0,42819525	UN	86188205	86188523				
DLPD08543_2	0,268116784	Chr16	19808881	19809304				
DLPD08543_1	0,293593488	Chr16	19808881	19809304				
DLPD08519_2	0,134789401	Chr19	17571895	17572058				
DLPD08519_1	0,15744661	Chr19	17571895	17572058				
DLPD08461_2	0,343945637	Chr15	13312403	13312688				
DLPD08461_1	0,328325037	Chr15	13312403	13312688				
DLPD08450_1	0,457414011	Chr10	12576003	12576231				
DLPD08393_2	0,266040957	Chr20	2173097	2173426				
DLPD08393_1	0,290307785	Chr20	2173097	2173426				
DLPD08356_2	0,480655209	Chr5	22784525	22784950	DLAgn_00161190	GRAMD2	ENSDARP00000130256	ENSDARG00000103736
DLPD08352_2	0,391539551	Chr22-25	19103458	19103686				
DLPD08352_1	0,334386186	Chr22-25	19103458	19103686				
DLPD08323_2	0,259914236	Chr24	8190301	8190519				
DLPD08323_1	0,218872123	Chr24	8190301	8190519				
DLPD08321_2	0,18156266	Chr6	12624448	12625310	DLAgn_00168900	myrf	ENSDARP00000104600	ENSDARG00000078676
DLPD08321_1	0,236256235	Chr6	12624448	12625310	DLAgn_00168900	myrf	ENSDARP00000104600	ENSDARG00000078676
DLPD08265_2	0,370279859	Chr17	18674846	18675726				
DLPD08265_1	0,337956952	Chr17	18674846	18675726				
DLPD08256_2	0,48467244	Chr2	14569119	14569244	DLAgn_00111940	ITM2A	ENSDARP00000020118	ENSDARG00000007098
DLPD08256_1	0,482741214	Chr2	14569119	14569244	DLAgn_00111940	ITM2A	ENSDARP00000020118	ENSDARG00000007098
DLPD08254_2	0,342155202	Chr7	8738313	8738794				
DLPD08254_1	0,314669939	Chr7	8738313	8738794				
DLPD08253_2	0,136601465	Chr19	13547390	13547855	DLAgn_00084340			
DLPD08253_1	0,19810419	Chr19	13547390	13547855	DLAgn_00084340			
DLPD08245_2	0,354004526	Chr15	20517898	20518045				
DLPD08245_1	0,388697935	Chr15	20517898	20518045				
DLPD08225_2	0,246941591	Chr19	9857075	9857626	DLAgn_00083050	zgc:73226	ENSDARP00000037307	ENSDARG00000023759
DLPD08225_1	0,237285473	Chr19	9857075	9857626	DLAgn_00083050	zgc:73226	ENSDARP00000037307	ENSDARG00000023759
DLPD08209_2	0,209166225	Chr6	13234556	13234806	DLAgn_00169360	gnb5a	ENSDARP00000140306	ENSDARG00000099685
DLPD08209_1	0,202644425	Chr6	13234556	13234806	DLAgn_00169360	gnb5a	ENSDARP00000140306	ENSDARG00000099685
DLPD08196_2	0,477500076	Chr5	13485790	13486372	DLAgn_00158130	ryr2b	ENSDARP00000134589	ENSDARG00000003706

DLPD08161_2	0,164907656	Chr9	999186	999447				
DLPD08161_1	0,201550597	Chr9	999186	999447				
DLPD08085_2	0,334677761	Chr17	9372138	9372326				
DLPD08085_1	0,423593299	Chr17	9372138	9372326				
DLPD08074_1	0,296667786	Chr7	5099325	5099888				
DLPD08057_1	0,401661373	Chr13	26014530	26014794				
DLPD08045_2	0,315216229	Chr8	15395647	15396331	DLAgn_00193560	si:dkeyp-94b4.1	ENSDARP00000130945	ENSDARG00000102056
DLPD08045_1	0,259363828	Chr8	15395647	15396331	DLAgn_00193560	si:dkeyp-94b4.1	ENSDARP00000130945	ENSDARG00000102056
DLPD08005_2	0,135895511	Chr3	7473152	7473306				
DLPD08005_1	0,131813017	Chr3	7473152	7473306				
DLPD07877_2	0,495114286	Chr1B	17082164	17082692	DLAgn_00105940	si:ch1073-416j2	ENSDARP00000114966	ENSDARG00000055934
DLPD07865_2	0,20532173	Chr17	14113036	14113741				
DLPD07860_2	0,243321318	Chr24	8189451	8189574				
DLPD07860_1	0,290929523	Chr24	8189451	8189574				
DLPD07857_2	0,225967004	Chr9	14005528	14006336	DLAgn_00201560	stmn2a	ENSDARP00000072761	ENSDARG00000070537
DLPD07857_1	0,219119834	Chr9	14005528	14006336	DLAgn_00201560	stmn2a	ENSDARP00000072761	ENSDARG00000070537
DLPD07827_2	0,478533897	Chr6	2999698	3000272	DLAgn_00165110	kcng4a	ENSDARP00000121290	ENSDARG00000062967
DLPD07764_2	0,235370982	Chr19	21338486	21338869	DLAgn_00087060	stxbp1a	ENSDARP00000012776	ENSDARG00000001994
DLPD07764_1	0,264252997	Chr19	21338486	21338869	DLAgn_00087060	stxbp1a	ENSDARP00000012776	ENSDARG00000001994
DLPD07754_2	0,418146707	Chr1B	12541806	12542118				
DLPD07754_1	0,303492986	Chr1B	12541806	12542118				
DLPD07751_2	0,473181668	Chr22-25	5241039	5241352				
DLPD07751_1	0,452612902	Chr22-25	5241039	5241352				
DLPD07730_2	0,200016808	Chr9	14697892	14698464				
DLPD07730_1	0,268997386	Chr9	14697892	14698464				
DLPD07690_2	0,095606607	Chr1A	22828028	22828559				
DLPD07690_1	0,131928135	Chr1A	22828028	22828559				
DLPD07641_1	0,436181686	Chr2	10331693	10331888	DLAgn_00110250	sept8a	ENSDARP00000102463	ENSDARG00000032606
DLPD07632_2	0,30625932	Chr2	21614461	21615029				
DLPD07632_1	0,314982723	Chr2	21614461	21615029				
DLPD07619_2	0,14834444	UN	64710218	64710568				
DLPD07619_1	0,15727518	UN	64710218	64710568				
DLPD07608_2	0,452709679	Chr4	14477715	14477897	DLAgn_00149380	xpr1a	ENSDARP00000084739	ENSDARG00000062449
DLPD07608_1	0,445716514	Chr4	14477715	14477897	DLAgn_00149380	xpr1a	ENSDARP00000084739	ENSDARG00000062449
DLPD07581_2	0,248874098	Chr5	16985581	16986473				
DLPD07581_1	0,149296091	Chr5	16985581	16986473				
DLPD07572_1	0,498852598	Chr5	13241747	13242226	DLAgn_00157980	gnao1b	ENSDARP00000124476	ENSDARG00000016676
DLPD07525_1	0,204879451	Chr7	16975838	16976050	DLAgn_00181310	crmp1	ENSDARP00000064467	ENSDARG00000056742
DLPD07447_2	0,331088559	Chr1B	6347018	6347571				
DLPD07447_1	0,318528156	Chr1B	6347018	6347571				
DLPD07378_2	0,367612687	Chr20	10213082	10213412				
DLPD07378_1	0,372314269	Chr20	10213082	10213412				
DLPD07322_2	0,202139275	Chr17	9014617	9014890				
DLPD07322_1	0,206032292	Chr17	9014617	9014890				
DLPD07271_2	0,299000062	Chr4	14365713	14366147				
DLPD07247_2	0,308447301	Chr11	9848602	9848934				
DLPD07247_1	0,346110495	Chr11	9848602	9848934				
DLPD07208_2	0,188198104	Chr5	16984881	16985140				
DLPD07208_1	0,154437455	Chr5	16984881	16985140				
DLPD07172_2	0,191709429	Chr14	12770582	12770779				
DLPD07172_1	0,185661017	Chr14	12770582	12770779				
DLPD07149_2	0,410125481	Chr7	15890065	15890279	DLAgn_00180890	serhl	ENSDARP00000134552	ENSDARG00000032340
DLPD07149_1	0,447260592	Chr7	15890065	15890279	DLAgn_00180890	serhl	ENSDARP00000134552	ENSDARG00000032340
DLPD07137_2	0,4401843	Chr5	16536245	16536480				
DLPD07137_1	0,474420557	Chr5	16536245	16536480				
DLPD07129_2	0,433122003	Chr17	3985050	3985358				
DLPD07119_2	0,163671139	Chr14	12399313	12399492	DLAgn_00041080	gria3a	ENSDARP00000142042	ENSDARG00000032737
DLPD07119_1	0,168137663	Chr14	12399313	12399492	DLAgn_00041080	gria3a	ENSDARP00000142042	ENSDARG00000032737
DLPD07101_1	0,323554463	Chr9	17376524	17376693				
DLPD07071_2	0,141543468	Chr9	6684465	6685222				
DLPD07062_2	0,374740529	Chr14	20482262	20482807				
DLPD07062_1	0,421988728	Chr14	20482262	20482807				
DLPD07019_2	0,376716891	Chr4	16844387	16844625				
DLPD07019_1	0,366246655	Chr4	16844387	16844625				
DLPD07015_2	0,344168037	Chr20	16292721	16292971				
DLPD07015_1	0,31938699	Chr20	16292721	16292971				
DLPD07001_2	0,056944587	Chr5	4240445	4240833				
DLPD07001_1	0,058750715	Chr5	4240445	4240833				
DLPD06976_2	0,252146037	Chr14	8797813	8798159				
DLPD06976_1	0,235278279	Chr14	8797813	8798159				
DLPD06945_2	0,478904806	Chr16	8288684	8289042	DLAgn_00059270	mfap5	ENSDARP00000123341	ENSDARG00000090560
DLPD06945_1	0,480699836	Chr16	8288684	8289042	DLAgn_00059270	mfap5	ENSDARP00000123341	ENSDARG00000090560
DLPD06933_2	0,461458122	Chr16	10153759	10154039				
DLPD06926_2	0,362657329	Chr15	9686212	9686361				
DLPD06926_1	0,390665278	Chr15	9686212	9686361				
DLPD06912_1	0,467842453	Chr14	18606836	18607791	DLAgn_00044580	ublcp1	ENSDARP00000071186	ENSDARG00000044492
DLPD06806_2	0,297394582	Chr15	10527725	10527861	DLAgn_00049720	ptprna	ENSDARP00000136466	ENSDARG00000058646
DLPD06806_1	0,303319299	Chr15	10527725	10527861	DLAgn_00049720	ptprna	ENSDARP00000136466	ENSDARG00000058646
DLPD06787_2	0,133669396	Chr7	16975051	16975340				
DLPD06787_1	0,181545211	Chr7	16975051	16975340				
DLPD06690_2	0,229771878	Chr19	8869531	8870514	DLAgn_00082540	map1b	ENSDARP00000079577	ENSDARG00000060434

DLPD06690_1	0,22323186	Chr19	8869531	8870514	DLAgn_00082540	map1b	ENSDARP00000079577	ENSDARG00000060434
DLPD06673_2	0,210068367	Chr1A	29011850	29012083				
DLPD06673_1	0,193241102	Chr1A	29011850	29012083				
DLPD06571_2	0,298273168	Chr2	4923255	4923616				
DLPD06571_1	0,269372556	Chr2	4923255	4923616				
DLPD06530_2	0,308950567	Chr20	23338765	23338873	DLAgn_00123560	add2	ENSDARP00000099313	ENSDARG00000074581
DLPD06530_1	0,268515372	Chr20	23338765	23338873	DLAgn_00123560	add2	ENSDARP00000099313	ENSDARG00000074581
DLPD06520_2	0,200407629	Chr24	356273	356443				
DLPD06520_1	0,185435591	Chr24	356273	356443				
DLPD06508_2	0,120829863	Chr5	4239383	4239784	DLAgn_00155280	snap25a	ENSDARP00000019054	ENSDARG00000020609
DLPD06508_1	0,104585267	Chr5	4239383	4239784	DLAgn_00155280	snap25a	ENSDARP00000019054	ENSDARG00000020609
DLPD06421_2	0,195177432	UN	3072351	3072647				
DLPD06421_1	0,196627552	UN	3072351	3072647				
DLPD06354_2	0,1800933	Chr10	11103924	11105420	DLAgn_00005210	scg2b	ENSDARP00000099613	ENSDARG00000038574
DLPD06354_1	0,15025417	Chr10	11103924	11105420	DLAgn_00005210	scg2b	ENSDARP00000099613	ENSDARG00000038574
DLPD06337_2	0,205783039	UN	55746226	55746549				
DLPD06337_1	0,187113672	UN	55746226	55746549				
DLPD06320_2	0,062554493	Chr5	4240253	4240444				
DLPD06320_1	0,071193401	Chr5	4240253	4240444				
DLPD06317_2	0,296251951	Chr17	17387549	17387922				
DLPD06317_1	0,176072249	Chr17	17387549	17387922				
DLPD06274_2	0,120100025	Chr22-25	3241500	3241649	DLAgn_00126550	tuba1c	ENSDARP00000072006	ENSDARG00000055216
DLPD06274_1	0,129232805	Chr22-25	3241500	3241649	DLAgn_00126550	tuba1c	ENSDARP00000072006	ENSDARG00000055216
DLPD06253_2	0,385651639	Chr4	22372468	22373150				
DLPD06253_1	0,393326727	Chr4	22372468	22373150				
DLPD06237_2	0,085471641	Chr7	12028150	12028389				
DLPD06237_1	0,081470207	Chr7	12028150	12028389				
DLPD06216_2	0,229412305	Chr6	24507145	24507505				
DLPD06216_1	0,186928205	Chr6	24507145	24507505				
DLPD06196_2	0,321916783	Chrx	11440495	11441015	DLAgn_00208390	cracr2ab	ENSDARP00000114486	ENSDARG00000045758
DLPD06196_1	0,30926115	Chrx	11440495	11441015	DLAgn_00208390	cracr2ab	ENSDARP00000114486	ENSDARG00000045758
DLPD06195_2	0,467273981	Chr6	17418084	17418331				
DLPD06195_1	0,464003676	Chr6	17418084	17418331				
DLPD06176_2	0,094100493	Chr17	19195745	19196014	DLAgn_00072590	fabp7a	ENSDARP00000018865	ENSDARG00000007697
DLPD06176_1	0,074249368	Chr17	19195745	19196014	DLAgn_00072590	fabp7a	ENSDARP00000018865	ENSDARG00000007697
DLPD06173_1	0,475523323	Chr14	17657513	17657882				
DLPD06169_2	0,216787772	Chr14	12998198	12998475				
DLPD06169_1	0,207726846	Chr14	12998198	12998475				
DLPD06135_2	0,395644909	Chr13	6283730	6283985				
DLPD06135_1	0,349975489	Chr13	6283730	6283985				
DLPD06110_2	0,360253918	Chr14	5777759	5778463	DLAgn_00038790	fez1	ENSDARP00000109454	ENSDARG00000023174
DLPD06110_1	0,376532297	Chr14	5777759	5778463	DLAgn_00038790	fez1	ENSDARP00000109454	ENSDARG00000023174
DLPD06035_2	0,087925307	Chr5	4504880	4505313				
DLPD06035_1	0,086481279	Chr5	4504880	4505313				
DLPD06032_2	0,446137355	no match	no match	no match				
DLPD06032_1	0,414404267	no match	no match	no match				
DLPD06015_2	0,343471724	Chr1A	19634531	19634866				
DLPD06015_1	0,331606848	Chr1A	19634531	19634866				
DLPD05996_2	0,35393243	Chr14	20339875	20340814				
DLPD05996_1	0,388042634	Chr14	20339875	20340814				
DLPD05942_2	0,324737753	Chr14	12758598	12758988	DLAgn_00041310	NYAP1	ENSDARP00000130957	ENSDARG00000103123
DLPD05942_1	0,377686152	Chr14	12758598	12758988	DLAgn_00041310	NYAP1	ENSDARP00000130957	ENSDARG00000103123
DLPD05914_2	0,390561809	Chr13	16404184	16404844				
DLPD05914_1	0,371417284	Chr13	16404184	16404844				
DLPD05907_2	0,198369609	Chr22-25	453580	453844				
DLPD05907_1	0,24210362	Chr22-25	453580	453844				
DLPD05872_2	0,129524668	Chr7	3367056	3367268				
DLPD05872_1	0,135495257	Chr7	3367056	3367268				
DLPD05858_2	0,380327414	UN	773831	773992				
DLPD05858_1	0,396419038	UN	773831	773992				
DLPD05817_2	0,381798647	Chrx	13916156	13917085	DLAgn_00209200	chrm2a	ENSDARP00000139587	ENSDARG00000098612
DLPD05817_1	0,271935516	Chrx	13916156	13917085	DLAgn_00209200	chrm2a	ENSDARP00000139587	ENSDARG00000098612
DLPD05816_2	0,172165305	Chr14	19133445	19133728				
DLPD05816_1	0,233915693	Chr14	19133445	19133728				
DLPD05783_2	0,06482047	UN	64980903	64981145				
DLPD05783_1	0,086200433	UN	64980903	64981145				
DLPD05756_2	0,159713847	Chr13	7699017	7699267	DLAgn_00029480	itpk1b	ENSDARP00000128561	ENSDARG00000070583
DLPD05756_1	0,158837968	Chr13	7699017	7699267	DLAgn_00029480	itpk1b	ENSDARP00000128561	ENSDARG00000070583
DLPD05752_2	0,168035507	no match	no match	no match				
DLPD05752_1	0,132555356	no match	no match	no match				
DLPD05727_2	0,220207377	no match	no match	no match				
DLPD05727_1	0,358117471	no match	no match	no match				
DLPD05690_2	0,290497122	Chr6	13626758	13627566				
DLPD05690_1	0,270550547	Chr6	13626758	13627566				
DLPD05684_2	0,285571244	Chr13	16394111	16394253	DLAgn_00032280	hepacama	ENSDARP00000124635	ENSDARG00000056934
DLPD05684_1	0,310626019	Chr13	16394111	16394253	DLAgn_00032280	hepacama	ENSDARP00000124635	ENSDARG00000056934
DLPD05675_2	0,197403542	Chr12	18811530	18811766	DLAgn_00025350	vsnl1a	ENSDARP00000038830	ENSDARG00000023228
DLPD05675_1	0,168805814	Chr12	18811530	18811766	DLAgn_00025350	vsnl1a	ENSDARP00000038830	ENSDARG00000023228
DLPD05584_1	0,419282279	Chr5	8989425	8989813				
DLPD05555_2	0,241491128	Chr8	21138731	21138844	DLAgn_00196000	nsfa	ENSDARP00000133697	ENSDARG00000007654
DLPD05555_1	0,127796493	Chr8	21138731	21138844	DLAgn_00196000	nsfa	ENSDARP00000133697	ENSDARG00000007654

DLPD05527_2	0,392094777	Chr10	10281292	10281872				
DLPD05527_1	0,345438876	Chr10	10281292	10281872				
DLPD05466_2	0,417443946	Chr7	15889386	15890064	DLAgn_00180890	serhl	ENSDARP00000134552	ENSDARG00000032340
DLPD05466_1	0,418811243	Chr7	15889386	15890064	DLAgn_00180890	serhl	ENSDARP00000134552	ENSDARG00000032340
DLPD05465_2	0,194848081	Chr12	15782797	15783091	DLAgn_00023880	fabp7a	ENSDARP00000018865	ENSDARG00000007697
DLPD05465_1	0,197673863	Chr12	15782797	15783091	DLAgn_00023880	fabp7a	ENSDARP00000018865	ENSDARG00000007697
DLPD05315_2	0,163287362	Chr5	8337540	8337738				
DLPD05315_1	0,175534301	Chr5	8337540	8337738				
DLPD05309_2	0,461079165	Chr20	4458558	4459136	DLAgn_00116720	NEFH	ENSDARP00000074706	ENSDARG00000021351
DLPD05309_1	0,494202149	Chr20	4458558	4459136	DLAgn_00116720	NEFH	ENSDARP00000074706	ENSDARG00000021351
DLPD05308_2	0,361688833	Chr11	14361374	14361611	DLAgn_00015090	rgs7a	ENSDARP00000044992	ENSDARG00000016584
DLPD05300_2	0,300249516	Chr2	4130195	4130411	DLAgn_00107150	atp1b4	ENSDARP00000069762	ENSDARG00000053262
DLPD05300_1	0,307687354	Chr2	4130195	4130411	DLAgn_00107150	atp1b4	ENSDARP00000069762	ENSDARG00000053262
DLPD05277_2	0,066730328	Chr5	4239881	4240252				
DLPD05277_1	0,059107309	Chr5	4239881	4240252				
DLPD05271_1	0,429589363	Chr17	21294755	21294975	DLAgn_00073300	serinc1	ENSDARP00000027703	ENSDARG00000009106
DLPD05159_1	0,108855283	Chr9	14880027	14880418	DLAgn_00201920	pabpc1b	ENSDARP00000034822	ENSDARG00000021140
DLPD05050_2	0,438857872	Chr16	23211362	23211519				
DLPD05050_1	0,460221301	Chr16	23211362	23211519				
DLPD05031_2	0,160043635	Chr8	1686397	1686891	DLAgn_00186830	nptx1l	ENSDARP00000091100	ENSDARG00000074671
DLPD05031_1	0,122541526	Chr8	1686397	1686891	DLAgn_00186830	nptx1l	ENSDARP00000091100	ENSDARG00000074671
DLPD05000_2	0,339281642	Chr10	7947876	7948195	DLAgn_00003370	atp1b3a	ENSDARP00000013993	ENSDARG00000015790
DLPD05000_1	0,33907664	Chr10	7947876	7948195	DLAgn_00003370	atp1b3a	ENSDARP00000013993	ENSDARG00000015790
DLPD04903_2	0,226522115	Chr16	1618190	1618579				
DLPD04903_1	0,265555692	Chr16	1618190	1618579				
DLPD04899_2	0,144623468	Chr2	12538411	12539015				
DLPD04899_1	0,147908954	Chr2	12538411	12539015				
DLPD04894_2	0,128199372	Chr24	1051821	1052405	DLAgn_00136740			
DLPD04894_1	0,146378354	Chr24	1051821	1052405	DLAgn_00136740			
DLPD04890_2	0,123506955	Chr22-25	16798683	16798964	DLAgn_00132990	kif5aa	ENSDARP00000139080	ENSDARG00000098936
DLPD04890_1	0,124996989	Chr22-25	16798683	16798964	DLAgn_00132990	kif5aa	ENSDARP00000139080	ENSDARG00000098936
DLPD04870_1	0,221531281	Chr14	26113514	26113954				
DLPD04802_1	0,270963871	no match	no match	no match				
DLPD04746_1	0,47689354	Chr7	26467214	26467569				
DLPD04695_2	0,413785683	Chr5	22785453	22786160	DLAgn_00161190	GRAMD2	ENSDARP00000130256	ENSDARG00000103736
DLPD04679_2	0,255741346	Chr17	4303593	4303830				
DLPD04679_1	0,230576039	Chr17	4303593	4303830				
DLPD04629_2	0,322189443	Chr1A	17421032	17421370				
DLPD04629_1	0,363909509	Chr1A	17421032	17421370				
DLPD04598_2	0,475226697	Chr18-21	11108272	11108686	DLAgn_00077420	ADCYAP1	ENSDARP00000008265	ENSDARG00000004015
DLPD04598_1	0,307956948	Chr18-21	11108272	11108686	DLAgn_00077420	ADCYAP1	ENSDARP00000008265	ENSDARG00000004015
DLPD04539_2	0,265591802	Chr1A	5807834	5808100				
DLPD04537_2	0,466836808	Chr1A	19633979	19634176				
DLPD04409_1	0,478677608	Chr14	18077440	18077576	DLAgn_00044230	sept4a	ENSDARP00000003666	ENSDARG00000010385
DLPD04332_1	0,358907988	Chr20	11062820	11063480				
DLPD04303_2	0,419784388	Chr16	3046024	3046206				
DLPD04303_1	0,443447982	Chr16	3046024	3046206				
DLPD04302_2	0,177298837	Chr18-21	9458181	9458554				
DLPD04302_1	0,146149278	Chr18-21	9458181	9458554				
DLPD04295_2	0,22029312	UN	68159709	68159971				
DLPD04295_1	0,199973395	UN	68159709	68159971				
DLPD04262_2	0,22068656	Chr2	21617595	21617885				
DLPD04256_2	0,472808106	UN	11321086	11321985	DLAgn_00217470	dmtn	ENSDARP00000141261	ENSDARG00000013110
DLPD04256_1	0,409233402	UN	11321086	11321985	DLAgn_00217470	dmtn	ENSDARP00000141261	ENSDARG00000013110
DLPD04234_2	0,261815538	Chr4	23295310	23295633	DLAgn_00153620	ttc39c	ENSDARP00000135033	ENSDARG00000102308
DLPD04234_1	0,247065734	Chr4	23295310	23295633	DLAgn_00153620	ttc39c	ENSDARP00000135033	ENSDARG00000102308
DLPD04202_1	0,194251633	Chr17	17386573	17387542				
DLPD04199_2	0,171716901	Chr12	19968352	19968607	DLAgn_00026070	plekhd1	ENSDARP00000110527	ENSDARG00000091349
DLPD04199_1	0,157512049	Chr12	19968352	19968607	DLAgn_00026070	plekhd1	ENSDARP00000110527	ENSDARG00000091349
DLPD04181_2	0,124905575	UN	43466378	43466510				
DLPD04181_1	0,155934929	UN	43466378	43466510				
DLPD04173_2	0,114917836	Chr7	21168846	21169134				
DLPD04173_1	0,118165928	Chr7	21168846	21169134				
DLPD04167_2	0,306692085	Chr13	965583	965785	DLAgn_00027460	clip3	ENSDARP00000071119	ENSDARG00000054456
DLPD04167_1	0,307513035	Chr13	965583	965785	DLAgn_00027460	clip3	ENSDARP00000071119	ENSDARG00000054456
DLPD04139_2	0,415263294	Chr22-25	212766	213079				
DLPD04139_1	0,413842211	Chr22-25	212766	213079				
DLPD04099_2	0,173978653	Chr16	10581344	10581627	DLAgn_00060970	FRRS1L	ENSDARP00000134057	ENSDARG00000103940
DLPD04099_1	0,096809373	Chr16	10581344	10581627	DLAgn_00060970	FRRS1L	ENSDARP00000134057	ENSDARG00000103940
DLPD03996_2	0,416546254	Chr10	7213341	7213669				
DLPD03996_1	0,32729247	Chr10	7213341	7213669				
DLPD03975_2	0,347382872	Chr1A	18876480	18876965				
DLPD03975_1	0,274718251	Chr1A	18876480	18876965				
DLPD03953_2	0,214696273	Chr17	3856301	3857486	DLAgn_00066900	stmn4	ENSDARP00000113219	ENSDARG00000030106
DLPD03953_1	0,20178881	Chr17	3856301	3857486	DLAgn_00066900	stmn4	ENSDARP00000113219	ENSDARG00000030106
DLPD03940_2	0,468506992	Chr22-25	12472880	12473179	DLAgn_00130860	arf3b	ENSDARP00000137953	ENSDARG00000036998
DLPD03912_2	0,244265603	Chr24	2187205	2187734				
DLPD03909_2	0,420569913	Chr20	12794736	12795002	DLAgn_00119830	oxt	ENSDARP00000062884	ENSDARG00000042845
DLPD03909_1	0,43748319	Chr20	12794736	12795002	DLAgn_00119830	oxt	ENSDARP00000062884	ENSDARG00000042845
DLPD03907_2	0,371696654	Chr15	4302131	4302728				
DLPD03907_1	0,41071229	Chr15	4302131	4302728				

DLPD03898_2	0,265041869	Chr10	20515107	20515753				
DLPD03898_1	0,28462058	Chr10	20515107	20515753				
DLPD03841_1	0,337947765	Chr9	4919172	4919391	DLAgn_00198180	stmn1b	ENSDARP00000041006	ENSDARG00000033655
DLPD03840_2	0,322211953	Chr9	4918684	4919056	DLAgn_00198180	stmn1b	ENSDARP00000041006	ENSDARG00000033655
DLPD03840_1	0,303874368	Chr9	4918684	4919056	DLAgn_00198180	stmn1b	ENSDARP00000041006	ENSDARG00000033655
DLPD03816_2	0,120232151	Chr2	23827076	23827721				
DLPD03816_1	0,143136946	Chr2	23827076	23827721				
DLPD03806_2	0,321895713	Chr20	2173987	2174521				
DLPD03782_2	0,483363264	Chr4	23048275	23048632	DLAgn_00153540	rab6ba	ENSDARP00000049176	ENSDARG00000034522
DLPD03740_2	0,137752678	Chr17	3990963	3991172				
DLPD03740_1	0,142285017	Chr17	3990963	3991172				
DLPD03660_2	0,458111704	Chr5	6805432	6805844				
DLPD03615_2	0,164715399	Chr5	13241417	13241576	DLAgn_00157980	gnao1b	ENSDARP00000124476	ENSDARG00000016676
DLPD03615_1	0,16637625	Chr5	13241417	13241576	DLAgn_00157980	gnao1b	ENSDARP00000124476	ENSDARG00000016676
DLPD03550_1	0,266341737	Chr14	14222145	14222533	DLAgn_00042180	map7d2a	ENSDARP00000080404	ENSDARG00000068480
DLPD03540_2	0,338780247	Chr3	9626082	9626922	DLAgn_00143110	fabp2	ENSDARP00000021241	ENSDARG00000006427
DLPD03540_1	0,360190276	Chr3	9626082	9626922	DLAgn_00143110	fabp2	ENSDARP00000021241	ENSDARG00000006427
DLPD03493_2	0,45445368	Chr14	12522565	12523167	DLAgn_00041110	plp1b	ENSDARP00000122278	ENSDARG00000011929
DLPD03493_1	0,419153978	Chr14	12522565	12523167	DLAgn_00041110	plp1b	ENSDARP00000122278	ENSDARG00000011929
DLPD03487_2	0,345886057	Chr15	18384401	18384918	DLAgn_00053460	zic2a	ENSDARP00000026131	ENSDARG00000015554
DLPD03487_1	0,340272927	Chr15	18384401	18384918	DLAgn_00053460	zic2a	ENSDARP00000026131	ENSDARG00000015554
DLPD03483_2	0,376143065	Chr7	23219652	23219925	DLAgn_00183600	uchl1	ENSDARP00000034536	ENSDARG00000026871
DLPD03483_1	0,330222963	Chr7	23219652	23219925	DLAgn_00183600	uchl1	ENSDARP00000034536	ENSDARG00000026871
DLPD03432_2	0,305447291	Chr1B	12154728	12155520				
DLPD03394_2	0,081197497	Chr7	16974744	16975050				
DLPD03394_1	0,084121297	Chr7	16974744	16975050				
DLPD03391_1	0,412494059	Chr9	9425199	9425820				
DLPD03364_2	0,232128249	Chr12	17386087	17386201				
DLPD03364_1	0,221831142	Chr12	17386087	17386201				
DLPD03333_2	0,21443996	Chr2	21616631	21617460				
DLPD03333_1	0,385690769	Chr2	21616631	21617460				
DLPD03289_2	0,267363211	Chr2	4965993	4966246				
DLPD03289_1	0,200493921	Chr2	4965993	4966246				
DLPD03274_2	0,338969709	Chr10	10442352	10442882				
DLPD03274_1	0,33365632	Chr10	10442352	10442882				
DLPD03209_2	0,370652402	Chr4	3790938	3791258	DLAgn_00145850	cfhl4	ENSDARP00000138093	ENSDARG00000102456
DLPD03183_2	0,433114248	Chr17	12733932	12735193	DLAgn_00070170	scg5	ENSDARP00000127419	ENSDARG00000032126
DLPD03183_1	0,434538527	Chr17	12733932	12735193	DLAgn_00070170	scg5	ENSDARP00000127419	ENSDARG00000032126
DLPD03174_1	0,227467446	Chr19	8453655	8454115				
DLPD03085_2	0,384407968	Chr10	23233610	23234166				
DLPD03085_1	0,374702257	Chr10	23233610	23234166				
DLPD03052_2	0,169360714	Chr13	7699532	7699896	DLAgn_00029480	itpk1b	ENSDARP00000128561	ENSDARG00000070583
DLPD03052_1	0,181680252	Chr13	7699532	7699896	DLAgn_00029480	itpk1b	ENSDARP00000128561	ENSDARG00000070583
DLPD03041_2	0,219421031	Chr24	8189879	8190169				
DLPD03041_1	0,233596088	Chr24	8189879	8190169				
DLPD03023_2	0,342521761	Chr10	14763325	14764103				
DLPD03023_1	0,307411365	Chr10	14763325	14764103				
DLPD03019_2	0,16052365	Chr13	7699268	7699477	DLAgn_00029480	itpk1b	ENSDARP00000128561	ENSDARG00000070583
DLPD03019_1	0,218630632	Chr13	7699268	7699477	DLAgn_00029480	itpk1b	ENSDARP00000128561	ENSDARG00000070583
DLPD03014_2	0,172404962	Chr22-25	5871264	5871462	DLAgn_00128050	BX323087.1	ENSDARP00000099543	ENSDARG00000075779
DLPD03014_1	0,172612699	Chr22-25	5871264	5871462	DLAgn_00128050	BX323087.1	ENSDARP00000099543	ENSDARG00000075779
DLPD03006_2	0,213474385	Chr9	11848484	11849322	DLAgn_00200480	cspg5b	ENSDARP00000136195	ENSDARG00000099793
DLPD03006_1	0,186085519	Chr9	11848484	11849322	DLAgn_00200480	cspg5b	ENSDARP00000136195	ENSDARG00000099793
DLPD02961_2	0,283076715	Chr16	7689033	7689470	DLAgn_00058720	clstn3	ENSDARP00000128441	ENSDARG00000073883
DLPD02961_1	0,470392101	Chr16	7689033	7689470	DLAgn_00058720	clstn3	ENSDARP00000128441	ENSDARG00000073883
DLPD02937_2	0,408407142	Chr13	26246321	26246440				
DLPD02937_1	0,356434225	Chr13	26246321	26246440				
DLPD02928_2	0,28751393	Chr17	7079304	7079750				
DLPD02928_1	0,386648575	Chr17	7079304	7079750				
DLPD02923_2	0,329205816	Chr10	23563596	23564014	DLAgn_00009760			
DLPD02923_1	0,334428201	Chr10	23563596	23564014	DLAgn_00009760			
DLPD02879_1	0,329663845	Chr22-25	2046398	2046979	DLAgn_00125880	hnf4a	ENSDARP00000029754	ENSDARG00000021494
DLPD02778_2	0,451870719	Chr19	9741829	9741947	DLAgn_00082960	st8sia5	ENSDARP00000093315	ENSDARG00000036584
DLPD02778_1	0,458967701	Chr19	9741829	9741947	DLAgn_00082960	st8sia5	ENSDARP00000093315	ENSDARG00000036584
DLPD02733_2	0,294800526	Chr10	20514898	20515089				
DLPD02733_1	0,293563397	Chr10	20514898	20515089				
DLPD02699_2	0,183821274	Chr3	9453781	9454393				
DLPD02699_1	0,361108904	Chr3	9453781	9454393				
DLPD02698_2	0,254575025	Chr16	19561740	19561980				
DLPD02698_1	0,251361421	Chr16	19561740	19561980				
DLPD02657_2	0,364841842	Chr16	8614319	8614559				
DLPD02657_1	0,36383734	Chr16	8614319	8614559				
DLPD02650_2	0,184944857	Chr17	3990274	3990744				
DLPD02650_1	0,256475266	Chr17	3990274	3990744				
DLPD02595_1	0,425192121	Chr13	14652412	14652860				
DLPD02513_2	0,116514892	Chr22-25	4830766	4831374				
DLPD02513_1	0,235680107	Chr22-25	4830766	4831374				
DLPD02505_2	0,436205717	Chr2	24304735	24304845				
DLPD02505_1	0,408464146	Chr2	24304735	24304845				
DLPD02425_2	0,169074919	Chrx	9164975	9165389	DLAgn_00207750	cntn1b	ENSDARP00000067165	ENSDARG00000045685
DLPD02425_1	0,256353508	Chrx	9164975	9165389	DLAgn_00207750	cntn1b	ENSDARP00000067165	ENSDARG00000045685

DLPD02311_2	0,315161017	Chr2	4134347	4134677	DLAgn_00107150	atp1b4	ENSDARP00000069762	ENSDARG00000053262
DLPD02311_1	0,341189982	Chr2	4134347	4134677	DLAgn_00107150	atp1b4	ENSDARP00000069762	ENSDARG00000053262
DLPD02301_2	0,308516391	Chr14	12069343	12069529				
DLPD02301_1	0,306653957	Chr14	12069343	12069529				
DLPD02282_1	0,233805318	Chr17	7400456	7400960	DLAgn_00067920	naa30	ENSDARP00000104949	ENSDARG00000079686
DLPD02242_2	0,24823342	Chr22-25	13320732	13322065				
DLPD02242_1	0,260121563	Chr22-25	13320732	13322065				
DLPD02201_2	0,446431184	Chr6	16968682	16969507	DLAgn_00170870	SLC6A13	ENSDARP00000088027	ENSDARG00000067567
DLPD02201_1	0,202316414	Chr6	16968682	16969507	DLAgn_00170870	SLC6A13	ENSDARP00000088027	ENSDARG00000067567
DLPD02168_2	0,093883514	Chr1B	11859498	11859726				
DLPD02168_1	0,091646913	Chr1B	11859498	11859726				
DLPD02151_2	0,136438968	Chr24	2980382	2980705	DLAgn_00137200	map2	ENSDARP00000104237	ENSDARG00000055052
DLPD02151_1	0,144904937	Chr24	2980382	2980705	DLAgn_00137200	map2	ENSDARP00000104237	ENSDARG00000055052
DLPD02121_2	0,114054204	Chr7	3367325	3367424				
DLPD02121_1	0,122171335	Chr7	3367325	3367424				
DLPD02113_2	0,187578385	Chr20	5015736	5016078	DLAgn_00117070	f2r	ENSDARP00000078547	ENSDARG00000060012
DLPD02113_1	0,197024167	Chr20	5015736	5016078	DLAgn_00117070	f2r	ENSDARP00000078547	ENSDARG00000060012
DLPD02084_2	0,230041277	Chr5	8467411	8467763	DLAgn_00156320	tspan3b	ENSDARP00000050229	ENSDARG00000034753
DLPD02084_1	0,23461287	Chr5	8467411	8467763	DLAgn_00156320	tspan3b	ENSDARP00000050229	ENSDARG00000034753
DLPD02083_2	0,263280404	Chr12	18813782	18813913	DLAgn_00025360	smc6	ENSDARP00000110583	ENSDARG00000091821
DLPD02083_1	0,19063363	Chr12	18813782	18813913	DLAgn_00025360	smc6	ENSDARP00000110583	ENSDARG00000091821
DLPD02075_2	0,156109917	Chr1B	9034765	9035227	DLAgn_00102550	neurod2	ENSDARP00000023642	ENSDARG00000016854
DLPD02075_1	0,156887133	Chr1B	9034765	9035227	DLAgn_00102550	neurod2	ENSDARP00000023642	ENSDARG00000016854
DLPD02067_1	0,497471309	Chr8	5249027	5249286				
DLPD02038_2	0,188641579	Chr17	9014285	9014602	DLAgn_00068580	PRSS35	ENSDARP00000076534	ENSDARG00000059081
DLPD02038_1	0,171120682	Chr17	9014285	9014602	DLAgn_00068580	PRSS35	ENSDARP00000076534	ENSDARG00000059081
DLPD01990_2	0,239216176	Chr20	23339508	23339717	DLAgn_00123560	add2	ENSDARP00000099313	ENSDARG00000074581
DLPD01990_1	0,2473857	Chr20	23339508	23339717	DLAgn_00123560	add2	ENSDARP00000099313	ENSDARG00000074581
DLPD01860_2	0,293916187	Chr16	6607538	6607982	DLAgn_00058010			
DLPD01834_2	0,44276568	Chr13	839576	839865				
DLPD01834_1	0,439993145	Chr13	839576	839865				
DLPD01825_2	0,37621835	Chr10	20431086	20431446				
DLPD01825_1	0,381401506	Chr10	20431086	20431446				
DLPD01793_1	0,079077761	Chr22-25	7688764	7689119				
DLPD01760_2	0,340725602	Chr24	6004369	6004682				
DLPD01760_1	0,336179919	Chr24	6004369	6004682				
DLPD01750_2	0,291087996	Chr5	13412523	13412653				
DLPD01750_1	0,303880813	Chr5	13412523	13412653				
DLPD01709_2	0,130321763	Chr18-21	9457992	9458180				
DLPD01709_1	0,107071965	Chr18-21	9457992	9458180				
DLPD01672_2	0,423074148	Chr2	8810586	8810730	DLAgn_00109300	SPOCK1	ENSDARP00000101212	ENSDARG00000074644
DLPD01672_1	0,402608879	Chr2	8810586	8810730	DLAgn_00109300	SPOCK1	ENSDARP00000101212	ENSDARG00000074644
DLPD01661_2	0,182519257	Chr6	12623560	12623671	DLAgn_00168900	myrf	ENSDARP00000104600	ENSDARG00000078676
DLPD01661_1	0,220857559	Chr6	12623560	12623671	DLAgn_00168900	myrf	ENSDARP00000104600	ENSDARG00000078676
DLPD01591_1	0,435905209	UN	63527385	63527766				
DLPD01544_1	0,475832439	UN	43310826	43311592	DLAgn_00234790	btr25	ENSDARP00000137909	ENSDARG00000102018
DLPD01294_2	0,204427885	Chrx	4782460	4782543				
DLPD01269_1	0,456451445	Chr1A	4650466	4651209	DLAgn_00088530	hdac11	ENSDARP00000109681	ENSDARG00000087573
DLPD01216_1	0,491675919	UN	25321078	25321724	DLAgn_00223890		ENSDARP00000104912	ENSDARG00000090009
DLPD01183_1	0,431983165	Chr19	22460864	22461604	DLAgn_00087380	C7	ENSDARP00000114284	ENSDARG00000057121
DLPD01177_1	0,302626746	Chr20	13396323	13396561	DLAgn_00120050	ap3s2	ENSDARP00000058338	ENSDARG00000039882
DLPD01127_2	0,278775445	Chr12	10591052	10591345	DLAgn_00021700	tmed8	ENSDARP00000068724	ENSDARG00000052390
DLPD01001_2	0,394792079	Chr2	15067368	15067512				
DLPD01001_1	0,398768938	Chr2	15067368	15067512				
DLPD00995_2	0,327410268	Chr9	4918684	4918899	DLAgn_00198180	stmn1b	ENSDARP00000041006	ENSDARG00000033655
DLPD00995_1	0,348525288	Chr9	4918684	4918899	DLAgn_00198180	stmn1b	ENSDARP00000041006	ENSDARG00000033655
DLPD00803_2	0,445086899	Chr4	3790975	3791074	DLAgn_00145850	cfhl4	ENSDARP00000138093	ENSDARG00000102456
DLPD00803_1	0,468878118	Chr4	3790975	3791074	DLAgn_00145850	cfhl4	ENSDARP00000138093	ENSDARG00000102456
DLPD00702_2	0,238664584	Chr2	24707074	24708183				
DLPD00685_1	0,463531897	Chrx	4287522	4287846	DLAgn_00206280			
DLPD00357_2	0,293185193	Chr6	10731756	10732185	DLAgn_00167860	fkbp16	ENSDARP00000009132	ENSDARG00000001976

Supplementary Table S4: Differentially expressed genes between normal and jaw deformed juveniles (dissected mandible; 58 dph).

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Supplementary Table S5

Enriched BP terms at 38 alpha (whole head)

Category	Term	Genes	Count	%	EASE	Fold Enrichment	FDR
GOTERM_BP	GO:0035967~cellular response to topologically incorrect protein	ENSDARG00000045362		3,8462	0,0647	29,1372549	61,4
GOTERM_BP	GO:0006796~phosphate-containing compound metabolic process	ENSDARG00000016909		17,308	0,068	1,956979807	63,264
GOTERM_BP	GO:0044237~cellular metabolic process	ENSDARG00000169023		44,231	0,0684	1,29263004	63,514
GOTERM_BP	GO:0006793~phosphorus metabolic process	ENSDARG00000016909		17,308	0,0719	1,934522662	65,378
GOTERM_BP	GO:0008152~metabolic process	ENSDARG00000169026		50	0,078	1,223639204	68,515
GOTERM_BP	GO:0035966~response to topologically incorrect protein	ENSDARG00000045362		3,8462	0,0786	23,83957219	68,768

Enriched BP terms at 58 alpha (lower jaw)

Category	Term	Genes	Count	%	EASE	Fold Enrichment	FDR
GOTERM_BP	GO:0007186~G-protein coupled receptor signaling pathway	ENSDARG00000021912		7,5949	9E-06	5,373576799	0,0147
GOTERM_BP	GO:0070507~regulation of microtubule cytoskeleton organization	ENSDARG00000030105		3,1646	0,0001	18,28508772	0,1623
GOTERM_BP	GO:0007019~microtubule depolymerization	ENSDARG00000030104		2,5316	0,0001	35,10736842	0,1669
GOTERM_BP	GO:0032886~regulation of microtubule-based process	ENSDARG00000030105		3,1646	0,0002	15,67293233	0,3167
GOTERM_BP	GO:0006811~ion transport	ENSDARG00000327314		8,8608	0,0007	2,953744939	1,0494
GOTERM_BP	GO:0031110~regulation of microtubule polymerization or depolymerization	ENSDARG00000030104		2,5316	0,0012	17,55368421	1,8299
GOTERM_BP	GO:0007017~microtubule-based process	ENSDARG00000055219		5,6962	0,0017	3,949578947	2,5784
GOTERM_BP	GO:0031109~microtubule polymerization or depolymerization	ENSDARG00000030104		2,5316	0,0027	13,50283401	4,1068
GOTERM_BP	GO:0051261~protein depolymerization	ENSDARG00000030104		2,5316	0,005	10,97105263	7,5288
GOTERM_BP	GO:0007188~adenylate cyclase-modulating G-protein coupled receptor signaling pathway	ENSDARG00000016674		2,5316	0,005	10,97105263	7,5288
GOTERM_BP	GO:0007399~nervous system development	ENSDARG00000447118		11,392	0,0055	2,025425101	8,2844
GOTERM_BP	GO:0007187~G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	ENSDARG00000016674		2,5316	0,006	10,32569659	8,9312
GOTERM_BP	GO:0006814~sodium ion transport	ENSDARG00000058494		2,5316	0,0071	9,752046784	10,462
GOTERM_BP	GO:0030001~metal ion transport	ENSDARG00000001837		4,4304	0,0073	3,989473684	10,797
GOTERM_BP	GO:0006812~cation transport	ENSDARG00000009109		5,6962	0,0074	3,109904683	10,847
GOTERM_BP	GO:0000226~microtubule cytoskeleton organization	ENSDARG00000030106		3,7975	0,0087	4,619390582	12,692
GOTERM_BP	GO:0043624~cellular protein complex disassembly	ENSDARG00000030104		2,5316	0,0142	7,632036613	19,903
GOTERM_BP	GO:0007601~visual perception	ENSDARG00000002193		1,8987	0,0163	14,62807018	22,581
GOTERM_BP	GO:0050953~sensory perception of light stimulus	ENSDARG00000002193		1,8987	0,0163	14,62807018	22,581
GOTERM_BP	GO:0043241~protein complex disassembly	ENSDARG00000030104		2,5316	0,0178	7,021473684	24,388
GOTERM_BP	GO:0031175~neuron projection development	ENSDARG00000030109		5,6962	0,0204	2,598407202	27,352
GOTERM_BP	GO:0032984~macromolecular complex disassembly	ENSDARG00000030104		2,5316	0,022	6,501364522	29,173
GOTERM_BP	GO:0051493~regulation of cytoskeleton organization	ENSDARG00000030105		3,1646	0,0325	4,063352827	40,153
GOTERM_BP	GO:0048468~cell development	ENSDARG00000155515		9,4937	0,0333	1,798533218	40,931
GOTERM_BP	GO:0048699~generation of neurons	ENSDARG00000168511		6,962	0,0338	2,089724311	41,4
GOTERM_BP	GO:0048878~chemical homeostasis	ENSDARG00000042856		3,7975	0,0366	3,211039795	43,954
GOTERM_BP	GO:0015672~monovalent inorganic cation transport	ENSDARG00000058495		3,1646	0,0453	3,657017544	51,333
GOTERM_BP	GO:0030030~cell projection organization	ENSDARG00000301010		6,3291	0,0479	2,07000993	53,399
GOTERM_BP	GO:0006813~potassium ion transport	ENSDARG00000015793		1,8987	0,0492	8,228289474	54,32
GOTERM_BP	GO:0007267~cell-cell signaling	ENSDARG00000006864		2,5316	0,0497	4,744238976	54,691
GOTERM_BP	GO:0048666~neuron development	ENSDARG00000030109		5,6962	0,0514	2,170098323	55,927
GOTERM_BP	GO:0003008~system process	ENSDARG00000102507		4,4304	0,056	2,517946506	59,149
GOTERM_BP	GO:0022008~neurogenesis	ENSDARG00000168511		6,962	0,0595	1,893044376	61,426
GOTERM_BP	GO:0061024~membrane organization	ENSDARG00000023756		3,7975	0,0666	2,714487249	65,741
GOTERM_BP	GO:0006820~anion transport	ENSDARG00000009105		3,1646	0,072	3,134586466	68,697
GOTERM_BP	GO:0007600~sensory perception	ENSDARG00000002193		1,8987	0,0735	6,582631579	69,459
GOTERM_BP	GO:0022411~cellular component disassembly	ENSDARG00000030104		2,5316	0,0799	3,900818713	72,576
GOTERM_BP	GO:0007268~synaptic transmission	ENSDARG00000006863		1,8987	0,0801	6,269172932	72,66

Supplementary Table S5: Functional enrichment analysis carried out on DEGs

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Supplementary Table S6-1

L39743 region (ChrX, position 3443463 ± 500 kb)							
DLPD probe	**Chromosome**	**Start position**	**End position**	**DicLabv1.0c gene match**	**Gene Name**	**Fold Change**	
DLPD04709_1	ChrX	3007201	3008640	DLAgn_00205880	sbf1	1,002491138	
DLPD04709_2	ChrX	3007201	3008640	DLAgn_00205880	sbf1	1,118643161	
DLPD00515_1	ChrX	3008608	3009239	DLAgn_00205880	sbf1	1,115906839	*
DLPD00515_2	ChrX	3008608	3009239	DLAgn_00205880	sbf1	1,037996752	
DLPD09365_1	ChrX	3092657	3092903	DLAgn_00205940	psmc2	0,901066071	
DLPD09365_2	ChrX	3092657	3092903	DLAgn_00205940	psmc2	0,862388063	
DLPD03988_1	ChrX	3095050	3095252	DLAgn_00205940	psmc2	0,981112533	
DLPD03988_2	ChrX	3095050	3095252	DLAgn_00205940	psmc2	1,025411263	
DLPD17852_2	ChrX	3101503	3101725	DLAgn_00205950	dnajc2	0,8498834	
DLPD13338_1	ChrX	3109193	3109392	DLAgn_00205960	pmpcb	0,908816159	
DLPD13338_2	ChrX	3109193	3109392	DLAgn_00205960	pmpcb	0,933008075	
DLPD18433_1	ChrX	3211396	3212105			1,625370684	
DLPD18433_2	ChrX	3211396	3212105			1,666403443	
DLPD16772_1	ChrX	3287504	3288126	DLAgn_00206000	si:ch73-278m9.1	1,130469679	*
DLPD16772_2	ChrX	3287504	3288126	DLAgn_00206000	si:ch73-278m9.1	1,099697123	
DLPD00529_1	ChrX	3301058	3302020			0,982393284	
DLPD00529_2	ChrX	3301058	3302020			1,161337723	
DLPD17270_2	ChrX	3309253	3309464	DLAgn_00206010	cmah	6,02153872	*
DLPD17667_2	ChrX	3332913	3333114	DLAgn_00206020	glipr1a	1,393303428	*
DLPD12695_1	ChrX	3338596	3339043	DLAgn_00206020	glipr1a	1,431545301	*
DLPD10335_1	ChrX	3377893	3378293	DLAgn_00206050	si:dkey-30c15.17	1,013440407	
DLPD10335_2	ChrX	3377893	3378293	DLAgn_00206050	si:dkey-30c15.17	1,001720199	
DLPD08515_1	ChrX	3474022	3474466			1,130887676	*
DLPD08515_2	ChrX	3474022	3474466			1,089338208	
DLPD02509_1	ChrX	3479513	3480069	DLAgn_00206080	nap1l1	1,013903553	
DLPD02509_2	ChrX	3479513	3480069	DLAgn_00206080	nap1l1	1,031159834	
DLPD09704_1	ChrX	3496797	3496899	DLAgn_00206080	nap1l1	1,005874777	
DLPD09704_2	ChrX	3496797	3496899	DLAgn_00206080	nap1l1	0,955409482	
DLPD17765_1	ChrX	3708340	3708508	DLAgn_00206140	si:ch211-251f6.6	1,389068189	
DLPD17765_2	ChrX	3708340	3708508	DLAgn_00206140	si:ch211-251f6.6	1,50054294	
DLPD04853_1	ChrX	3745453	3745966			1,0085241	
DLPD04853_2	ChrX	3745453	3745966			1,028028459	

L12903 region (Chr17, position 1566309 ± 500 kb)							
DLPD probe	**Chromosome**	**Start position**	**End position**	**DicLabv1.0c gene match**	**Gene Name**	**Fold Change**	
DLPD06187_1	Chr17	15138678	15139478	DLAgn_00070930	ppp2r5a	0,978326704	
DLPD06187_2	Chr17	15138678	15139478	DLAgn_00070930	ppp2r5a	1,007510909	
DLPD13178_1	Chr17	15140263	15140999	DLAgn_00070930	ppp2r5a	0,975958877	
DLPD13178_2	Chr17	15140263	15140999	DLAgn_00070930	ppp2r5a	0,90851612	
DLPD16813_2	Chr17	15225183	15226095			0,784745923	*
DLPD03852_1	Chr17	15254165	15255602	DLAgn_00070970	lpgat1	1,027195329	
DLPD03852_2	Chr17	15254165	15255602	DLAgn_00070970	lpgat1	1,080616271	
DLPD02975_1	Chr17	15270778	15271673	DLAgn_00070990	slc30a1a	0,91055647	
DLPD02975_2	Chr17	15270778	15271673	DLAgn_00070990	slc30a1a	0,930345001	
DLPD18542_1	Chr17	15302825	15303263	DLAgn_00071010		1,046469229	
DLPD18542_2	Chr17	15302825	15303263	DLAgn_00071010		1,002557667	
DLPD18055_1	Chr17	15305451	15305621	DLAgn_00071010		1,246485208	*
DLPD18055_2	Chr17	15305451	15305621	DLAgn_00071010		1,10605577	
DLPD08166_1	Chr17	15309103	15309756	DLAgn_00071020	rhag	0,828229292	*
DLPD08166_2	Chr17	15309103	15309756	DLAgn_00071020	rhag	0,843897008	*
DLPD10525_1	Chr17	15354041	15354919	DLAgn_00071040	nrbp1	1,019343838	
DLPD10525_2	Chr17	15354041	15354919	DLAgn_00071040	nrbp1	0,945893317	
DLPD06255_1	Chr17	15355516	15355815	DLAgn_00071040	nrbp1	1,377206605	*
DLPD06255_2	Chr17	15355516	15355815	DLAgn_00071040	nrbp1	1,467200047	*
DLPD02799_1	Chr17	15393100	15393430	DLAgn_00071070	churc1	1,003566364	
DLPD02799_2	Chr17	15393100	15393430	DLAgn_00071070	churc1	0,994487129	
DLPD07211_1	Chr17	15457204	15457736	DLAgn_00071130	ttc32	0,875988767	
DLPD07211_2	Chr17	15457204	15457736	DLAgn_00071130	ttc32	0,895274187	

DLPD02483_1Chr17	15482067	15482653	DLAgn_00071160		1,026619584	
DLPD02483_2Chr17	15482067	15482653	DLAgn_00071160		0,951720393	
DLPD02482_1Chr17	15482239	15482671	DLAgn_00071160		1,067080244	
DLPD02482_2Chr17	15482239	15482671	DLAgn_00071160		1,063842863	
DLPD02101_1Chr17	15487095	15487919	DLAgn_00071170	pum2	1,17437057	*
DLPD02101_2Chr17	15487095	15487919	DLAgn_00071170	pum2	1,100686849	
DLPD00721_2Chr17	15507512	15508323	DLAgn_00071180	RHOB	0,971134286	
DLPD00720_2Chr17	15507548	15508289	DLAgn_00071180	RHOB	0,981226537	
DLPD02082_1Chr17	15510224	15511593	DLAgn_00071190	apoba	1,593101579	
DLPD02691_1Chr17	15550777	15551807	DLAgn_00071210	apobb.1	0,814774341	
DLPD02691_2Chr17	15550777	15551807	DLAgn_00071210	apobb.1	1,017092529	
DLPD10512_2Chr17	15551660	15551956	DLAgn_00071210	apobb.1	1,023340082	
DLPD12513_1Chr17	15594690	15595274	DLAgn_00071230	pdia6	0,917835658	
DLPD12513_2Chr17	15594690	15595274	DLAgn_00071230	pdia6	0,777664526	*
DLPD12640_1Chr17	15625236	15625918	DLAgn_00071270	rock2a	0,878662984	
DLPD12640_2Chr17	15625236	15625918	DLAgn_00071270	rock2a	0,877749185	
DLPD05641_1Chr17	15657785	15658534	DLAgn_00071280	sobpb	0,866333672	
DLPD05641_2Chr17	15657785	15658534	DLAgn_00071280	sobpb	0,930039107	
DLPD00305_1Chr17	15658423	15659348	DLAgn_00071280	sobpb	0,793672419	*
DLPD00305_2Chr17	15658423	15659348	DLAgn_00071280	sobpb	0,777715191	*
DLPD12951_1Chr17	15664095	15664384	DLAgn_00071280	sobpb	0,972153387	
DLPD02605_1Chr17	15911667	15911981	DLAgn_00071360	prep	0,585888941	*
DLPD02605_2Chr17	15911667	15911981	DLAgn_00071360	prep	0,608361468	*
DLPD06714_1Chr17	15937100	15937902	DLAgn_00071370	prdm1a	0,886757725	
DLPD06714_2Chr17	15937100	15937902	DLAgn_00071370	prdm1a	0,823606889	
DLPD06829_1Chr17	16039106	16039283	DLAgn_00071420	mrpl19	0,951171113	
DLPD06829_2Chr17	16039106	16039283	DLAgn_00071420	mrpl19	0,864572178	*

Supplementary Table S6-1: SAM significant test on all probes mapping in the regions spanning ± 500kb respectively around the two most significant GWAS analysis loci (38 dph).

*P<0.01

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Supplementary Table S6-2

L39743 region (ChrX, position 3443463 ± 500 kb)							
DLPD probe	**Chromosome**	**Start position**	**End position**	**DicLabv1.0c gene match**	**Gene Name**	**Fold Change**	
DLPD04709_1	ChrX	3007201	3008640	DLAgn_00205880	sbf1	0,928144985	
DLPD04709_2	ChrX	3007201	3008640	DLAgn_00205880	sbf1	0,991644429	*
DLPD00515_1	ChrX	3008608	3009239	DLAgn_00205880	sbf1	0,900728227	*
DLPD00515_2	ChrX	3008608	3009239	DLAgn_00205880	sbf1	0,81404543	
DLPD09365_1	ChrX	3092657	3092903	DLAgn_00205940	psmc2	0,994543047	
DLPD09365_2	ChrX	3092657	3092903	DLAgn_00205940	psmc2	0,967712616	
DLPD03988_1	ChrX	3095050	3095252	DLAgn_00205940	psmc2	0,948105552	
DLPD03988_2	ChrX	3095050	3095252	DLAgn_00205940	psmc2	0,961234082	
DLPD17852_2	ChrX	3101503	3101725	DLAgn_00205950	dnajc2	1,291533804	
DLPD13338_1	ChrX	3109193	3109392	DLAgn_00205960	pmpcb	1,264394714	
DLPD13338_2	ChrX	3109193	3109392	DLAgn_00205960	pmpcb	1,227936535	
DLPD18433_1	ChrX	3211396	3212105			1,14481188	
DLPD18433_2	ChrX	3211396	3212105			1,293093394	
DLPD16772_1	ChrX	3287504	3288126	DLAgn_00206000	si:ch73-278m9.1	1,124497304	
DLPD16772_2	ChrX	3287504	3288126	DLAgn_00206000	si:ch73-278m9.1	1,186575912	*
DLPD00529_1	ChrX	3301058	3302020			0,757340255	
DLPD00529_2	ChrX	3301058	3302020			0,962201828	
DLPD12695_1	ChrX	3338596	3339043	DLAgn_00206020	glipr1a	0,983791307	*
DLPD10335_1	ChrX	3377893	3378293	DLAgn_00206050	si:dkey-30c15.17	0,779419795	*
DLPD10335_2	ChrX	3377893	3378293	DLAgn_00206050	si:dkey-30c15.17	0,779336256	
DLPD08515_1	ChrX	3474022	3474466			1,03677053	
DLPD08515_2	ChrX	3474022	3474466			1,174129394	
DLPD02509_1	ChrX	3479513	3480069	DLAgn_00206080	nap1l1	0,996835917	
DLPD02509_2	ChrX	3479513	3480069	DLAgn_00206080	nap1l1	1,051503115	
DLPD09704_1	ChrX	3496797	3496899	DLAgn_00206080	nap1l1	1,053777375	
DLPD09704_2	ChrX	3496797	3496899	DLAgn_00206080	nap1l1	1,012788962	
DLPD17765_1	ChrX	3708340	3708508	DLAgn_00206140	si:ch211-251f6.6	1,088884398	
DLPD17765_2	ChrX	3708340	3708508	DLAgn_00206140	si:ch211-251f6.6	1,101920532	
DLPD04853_1	ChrX	3745453	3745966			0,948370831	
DLPD04853_2	ChrX	3745453	3745966			0,915574102	

L12903 region (Chr17, position 1566309 ± 500 kb)						
DLPD probe	**Chromosome**	**Start position**	**End position**	**DicLabv1.0c gene match**	**Gene Name**	**Fold Change**
DLPD17442_1	Chr17	15137326	15138036			1,096152589
DLPD06187_1	Chr17	15138678	15139478	DLAgn_00070930	ppp2r5a	0,890481646
DLPD06187_2	Chr17	15138678	15139478	DLAgn_00070930	ppp2r5a	0,89802538
DLPD13178_1	Chr17	15140263	15140999	DLAgn_00070930	ppp2r5a	1,000042303
DLPD13178_2	Chr17	15140263	15140999	DLAgn_00070930	ppp2r5a	1,077555231
DLPD16813_2	Chr17	15225183	15226095			0,63956928
DLPD03852_1	Chr17	15254165	15255602	DLAgn_00070970	lpgat1	0,848809433
DLPD03852_2	Chr17	15254165	15255602	DLAgn_00070970	lpgat1	0,895399999
DLPD02975_1	Chr17	15270778	15271673	DLAgn_00070990	slc30a1a	0,962901383
DLPD02975_2	Chr17	15270778	15271673	DLAgn_00070990	slc30a1a	0,961959505
DLPD18542_1	Chr17	15302825	15303263	DLAgn_00071010		1,058305288
DLPD18542_2	Chr17	15302825	15303263	DLAgn_00071010		1,032973237
DLPD18055_1	Chr17	15305451	15305621	DLAgn_00071010		1,285966806
DLPD18055_2	Chr17	15305451	15305621	DLAgn_00071010		1,251621767
DLPD08166_1	Chr17	15309103	15309756	DLAgn_00071020	rhag	0,820513634
DLPD08166_2	Chr17	15309103	15309756	DLAgn_00071020	rhag	0,840009749
DLPD10525_1	Chr17	15354041	15354919	DLAgn_00071040	nrbp1	0,934036837
DLPD10525_2	Chr17	15354041	15354919	DLAgn_00071040	nrbp1	0,948683775
DLPD06255_1	Chr17	15355516	15355815	DLAgn_00071040	nrbp1	1,015748871
DLPD06255_2	Chr17	15355516	15355815	DLAgn_00071040	nrbp1	1,028569664
DLPD02799_1	Chr17	15393100	15393430	DLAgn_00071070	churc1	0,963348605
DLPD02799_2	Chr17	15393100	15393430	DLAgn_00071070	churc1	0,998852931
DLPD07211_1	Chr17	15457204	15457736	DLAgn_00071130	ttc32	1,050552946
DLPD07211_2	Chr17	15457204	15457736	DLAgn_00071130	ttc32	1,084417251
DLPD02483_1	Chr17	15482067	15482653	DLAgn_00071160		1,001967903

DLPD02483_2	Chr17	15482067	15482653	DLAgn_00071160		0,902633539	
DLPD02482_1	Chr17	15482239	15482671	DLAgn_00071160		0,975781893	
DLPD02482_2	Chr17	15482239	15482671	DLAgn_00071160		0,930544292	
DLPD02101_1	Chr17	15487095	15487919	DLAgn_00071170	pum2	0,964083404	
DLPD02101_2	Chr17	15487095	15487919	DLAgn_00071170	pum2	1,001731894	
DLPD00721_2	Chr17	15507512	15508323	DLAgn_00071180	RHOB	1,123044135	
DLPD00720_2	Chr17	15507548	15508289	DLAgn_00071180	RHOB	1,085221497	
DLPD02082_1	Chr17	15510224	15511593	DLAgn_00071190	apoba	0,894204292	
DLPD02691_1	Chr17	15550777	15551807	DLAgn_00071210	apobb.1	0,477829088	
DLPD02691_2	Chr17	15550777	15551807	DLAgn_00071210	apobb.1	0,53454309	
DLPD10512_2	Chr17	15551660	15551956	DLAgn_00071210	apobb.1	0,901279141	
DLPD12513_1	Chr17	15594690	15595274	DLAgn_00071230	pdia6	0,645182623	*
DLPD12513_2	Chr17	15594690	15595274	DLAgn_00071230	pdia6	0,517139638	
DLPD12640_1	Chr17	15625236	15625918	DLAgn_00071270	rock2a	1,210330349	
DLPD12640_2	Chr17	15625236	15625918	DLAgn_00071270	rock2a	0,99533253	
DLPD05641_1	Chr17	15657785	15658534	DLAgn_00071280	sobpb	1,018539036	
DLPD05641_2	Chr17	15657785	15658534	DLAgn_00071280	sobpb	1,039592597	
DLPD00305_1	Chr17	15658423	15659348	DLAgn_00071280	sobpb	1,116805228	
DLPD00305_2	Chr17	15658423	15659348	DLAgn_00071280	sobpb	1,048219679	
DLPD12951_1	Chr17	15664095	15664384	DLAgn_00071280	sobpb	0,627421618	
DLPD02605_1	Chr17	15911667	15911981	DLAgn_00071360	prep	1,268510408	
DLPD02605_2	Chr17	15911667	15911981	DLAgn_00071360	prep	1,233403301	
DLPD06714_1	Chr17	15937100	15937902	DLAgn_00071370	prdm1a	1,407340016	
DLPD06714_2	Chr17	15937100	15937902	DLAgn_00071370	prdm1a	1,197404026	
DLPD06829_1	Chr17	16039106	16039283	DLAgn_00071420	mrpl19	0,949312805	
DLPD06829_2	Chr17	16039106	16039283	DLAgn_00071420	mrpl19	0,91501556	

Supplementary Table S6-2: SAM significant test on all probes mapping in the regions spanning ± 500kb respectively around the two most significant GWAS analysis loci (58 dph).

*P<0.01

An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*).

Supplementary Table S7.

Locus	Repeat	Primers	T_a	Size range	Alleles
DLA0008	$(AC)_{24}$	F:AAGCTATCTGATCTCGCTTG R:ACGTGATTAAGTGTTTGTGAG	56	236-298	11
DLA0119	$(TG)_{10}$	F:GCAGGTTCAAATTATTTTTGCTC R:TCCTCCTTTTGCTTGCTAGG	54	219-261	10
DLA0016	$(TG)_{24}$	F:GTGACCGCAGATGAAGAAC R:ACTGTGGGCTCATAAACATC	54	228-258	11
DLA0020	$(TG)_{20}$	F:GTCTAATGAGCAGTGGAGCAG R:GCATGTTAGATCCACCTCTTTC	56	153-175	8
DLA0105	$(AC)_{16}$	F:GAGGCTGTATGCTGTTGCAG R:ACCCATGCATAAGGTCAGTG	56	138-172	9
DLA0145	$(TC)_{20}$	F:CCCACAATAGATTCAAATAG R:CACACATGCAATTATACTG	54	152-188	10
DLA0248	$(TC)_5ACAT(TC)_5(T)_2(TC)_7$ $(AC)_3(ACGC)_4$	F:TGCATGATGATGTGTGAGCA R:TGGCAGGCTAAAACCTCAAG	54	111-127	5
DLA0228	$(AAAG)_3(AG)_4(AAAG)_3$	F:CCAATGTTTTCATCCCCTCA R:TTGCTGCTTGTGAAGTGACC	54	86-98	3
DLA0244	$(TG)_{12}(AG)_5(TG)_2$	F:ACTGAAAGCACAGCCTGGTT R:CCCCCATCCAATACACTCAC	54	100-104	3

Supplementary Table S7: Microsatellites loci for parental assignment. Locus name (Locus), repeats number (Repeat), primers sequence (Primers), annealing temperature (Ta), fragment length (Size range), alleles number (Alleles).

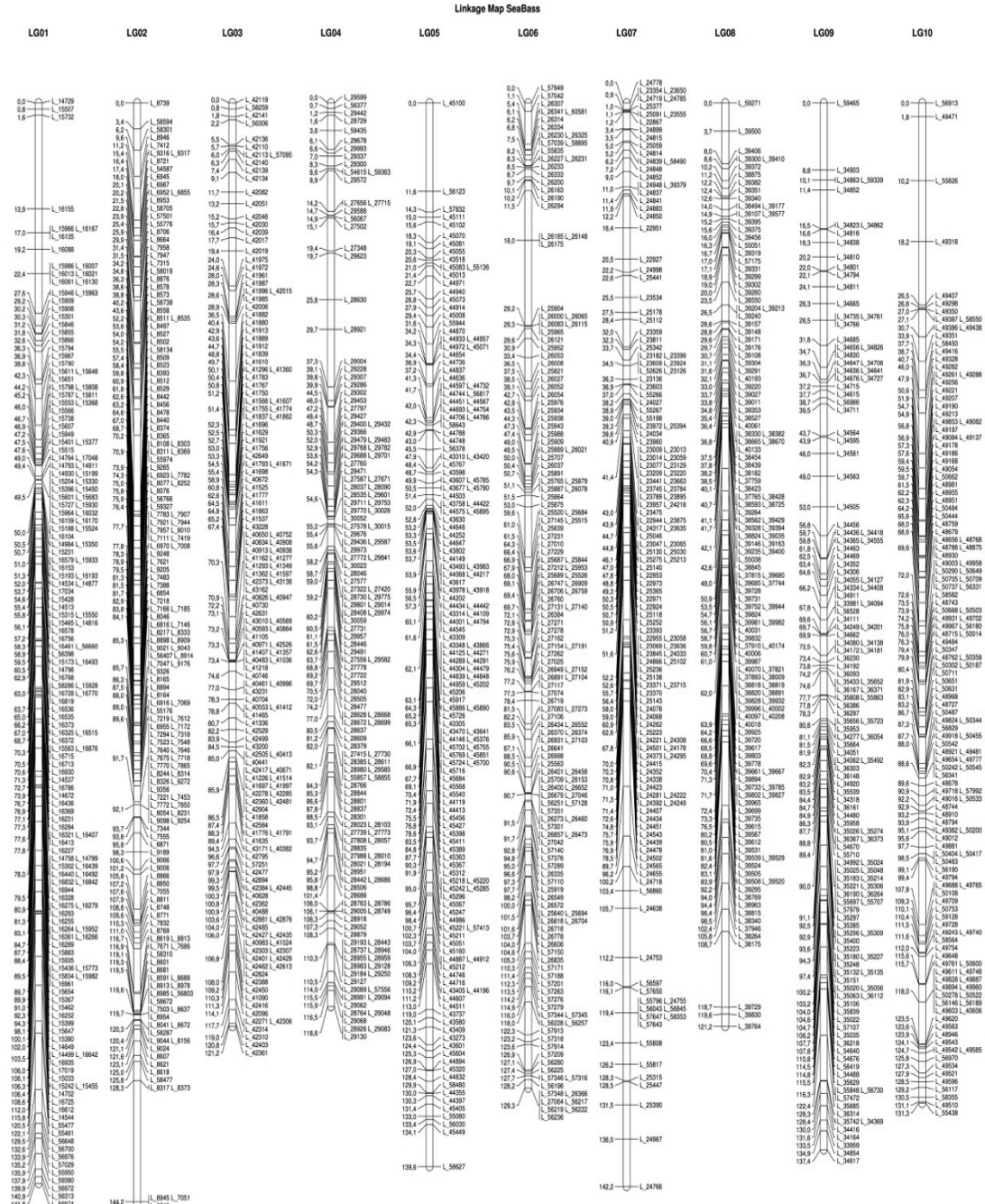
3. Conclusions

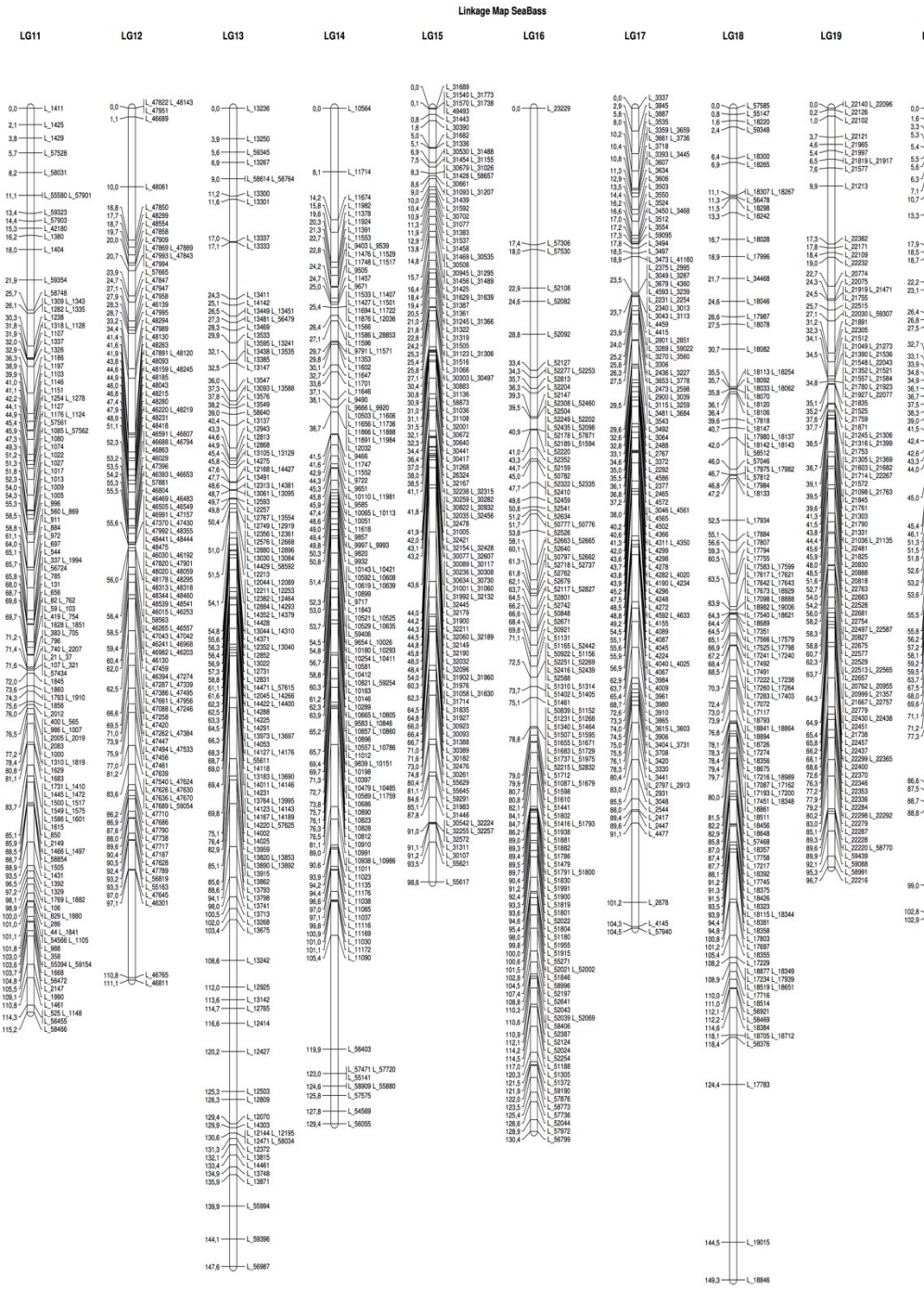
This study reports on the development and application of a 2bRAD genotyping technique on three species of veterinary interest and relevance. The main goal of this PhD research was to analyze the molecular basis of mandibular prognathism in farmed European seabass, using an integrated genomic approach. 2bRAD sequencing was initially carried out on parental and juvenile seabass DNAs, to construct a high-density linkage map and to search for QTLs associated with jaw deformity. A catalogue containing 7,390 loci was developed and used as reference for SNP discovery and genotyping to map families. The number of informative SNPs in the mapping panel was 3,266, distributed over 24 Linkage Groups (LGs) in a sex-averaged linkage map. A maternal *half-sib* regression analysis identified a total of 18 QTLs significant either at genome-wide or chromosome-wide level. The most significant one belongs to LG18, which corresponds to seabass Chr17 and explained 13.21% of total phenotypic variation. A second significant QTL was identified on LG22, matching ChrX, which explained 11.54% of total phenotypic variation. A GWAS analysis was then performed on a dataset of 7,390 loci and applied to 107 cases (prognathic samples) and 191 controls (non-prognatich samples). Two SNPs, on ChrX and Chr17, respectively, were found to be significantly associated with MP. Notably, the SNP marker on Chr17 was positioned within the Sobp gene coding region, know to play a pivotal role in craniofacial development. In human, homozygous missense mutation in *Sobp* has been associated with a syndrome causing mental retardation, anterior maxillary protrusion and strabismus. The most significant association with prognathism was found for a SNP marker located on ChrX, within a putative gene showing an *in silico* predicted transcript without any significant sequence similarity. However, the region around the gene appears to be well conserved across several teleost genomes, with the presence of conserved non-coding sequence elements (CNE) showing high similarity. Multiple lines of evidence suggest that CNEs have an important role in regulating gene expression, often encoding enhancers that act on nearby genes as well as distant ones and in several cases exerting their action on genetic loci involved in development. Finally, the analysis of differentially expressed genes in jaw-deformed animals highlighted the "nervous system development" as a crucial pathway in MP. Zic2 is a key gene for craniofacial morphogenesis in model species, and was significantly down-regulated in MP-affected seabass. Gene expression data revealed also a significant down-regulation of *Sobp* in deformed larvae, with concordant evidence from two probes. Such evidence further supports *Sobp* as a candidate gene contributing to mandibular prognathism. By integrating transcriptomic analysis with GWAS, the present study, has successfully provided evidence for the potential mechanisms underlying jaw deformity in the European seabass. QTL mapping and GWA analysis allowed the identification of

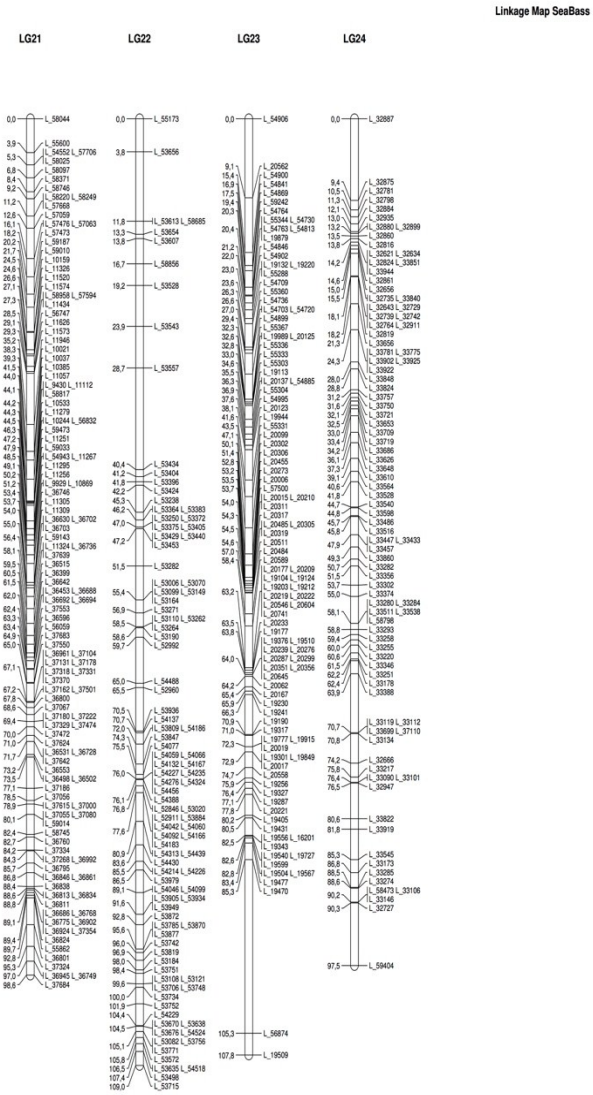
two regions implied in the determination of lower jaw deformity and pointed out a candidate gene, *Sobp*, as likely contributing to seabass MP. Moreover, the presence of a cluster of CNEs around to the most significant SNP on ChrX is suggestive that such elements might act as distant enhancers in craniofacial development. Finally, as previously reported in model species, differential regulation of several genes involved in neural development, such as putative *Zic2*a and stathmins, confirms the importance of this biological process to develop craniofacial deformities. The present work might be considered as a case-study proving the feasibility of an integrated genomic approach as a compelling strategy to unravel the molecular bases of skeletal anomalies in fish aquaculture. As a future perspective, these integrated methods should be used routinely to develop genetic tools intended to be applied in breeding selection.

4. Appendix

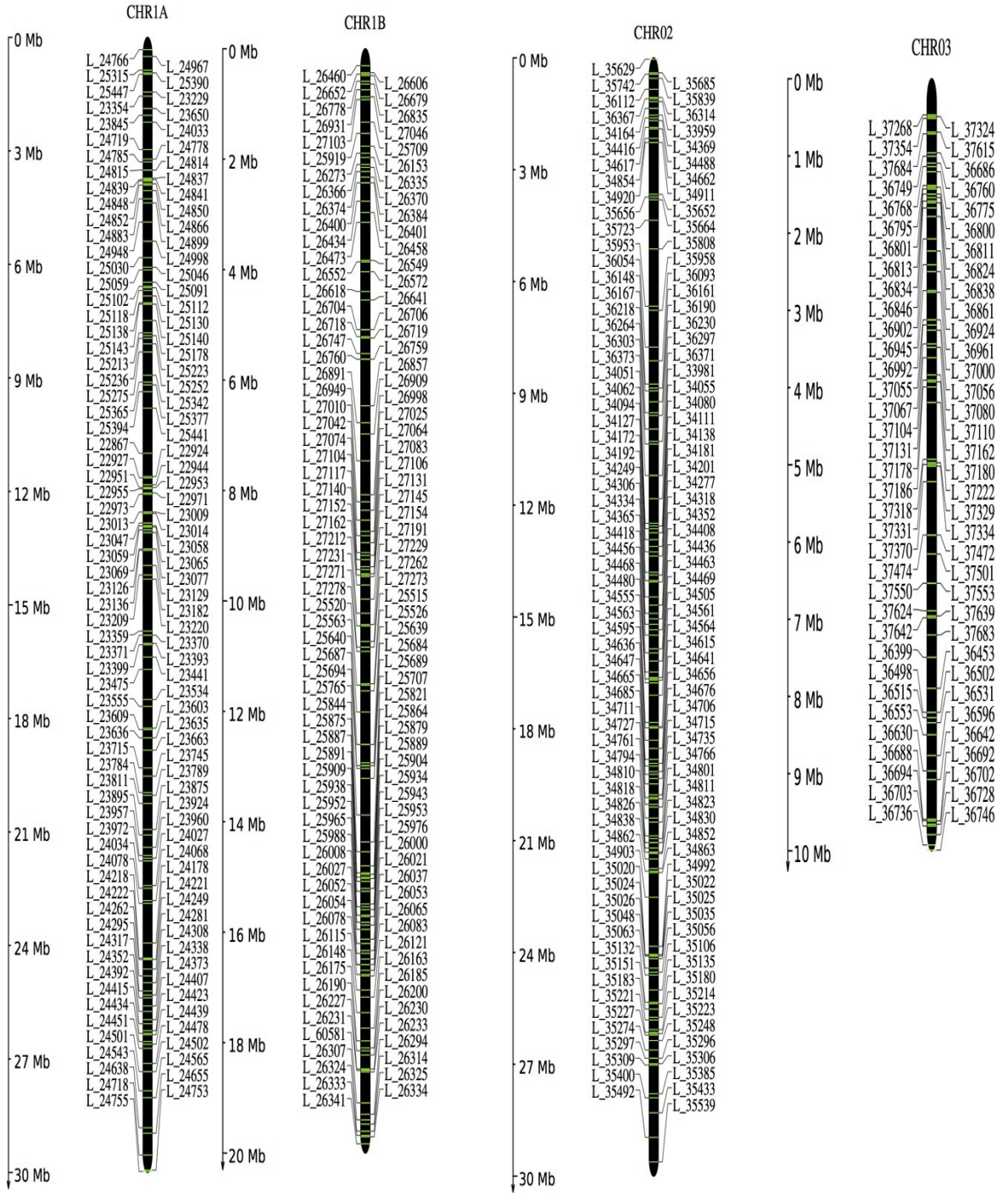
4.1 High-density single nucleotide polymorphism (SNP) based linkage map of European seabass (*Dicentrarchus labrax*).

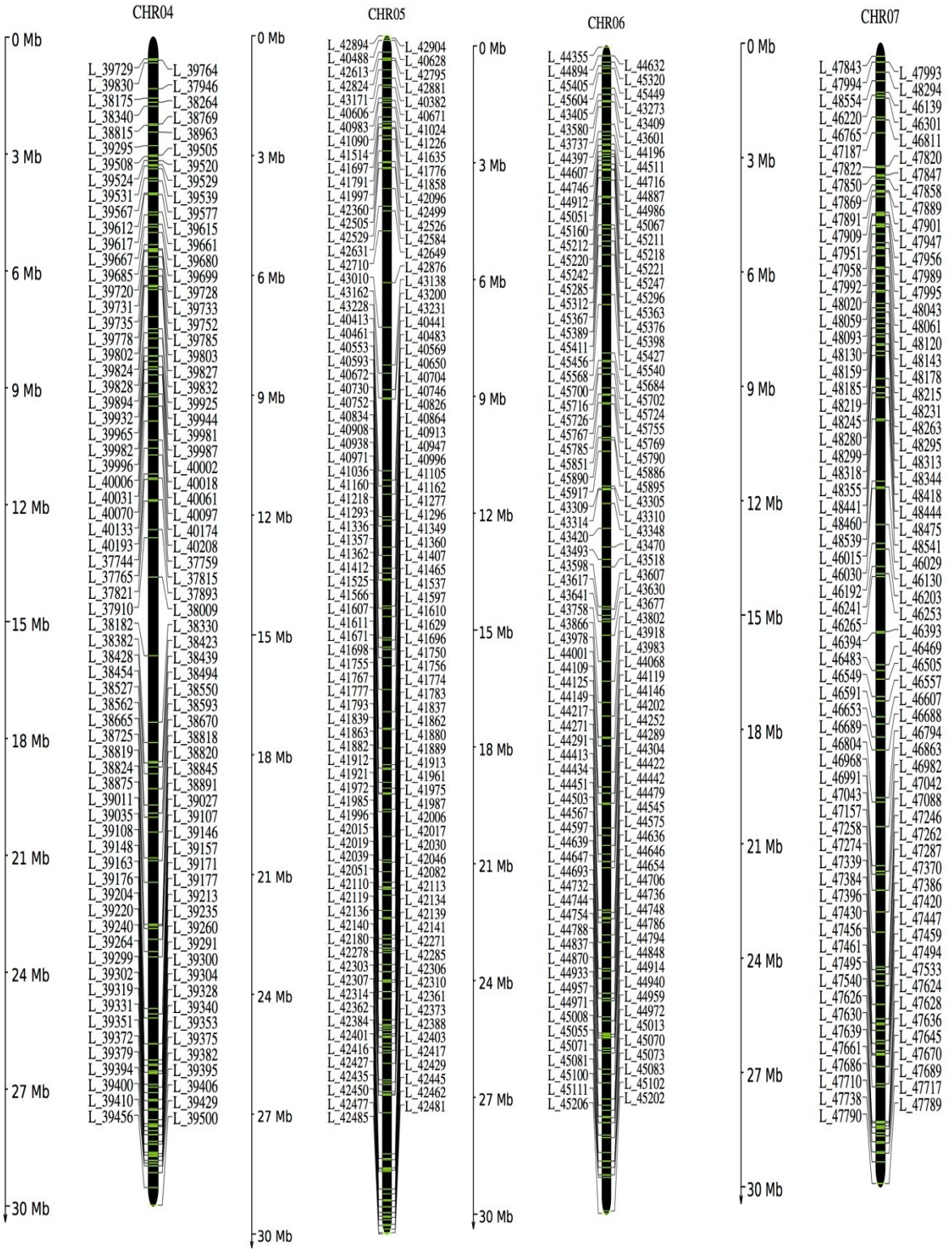


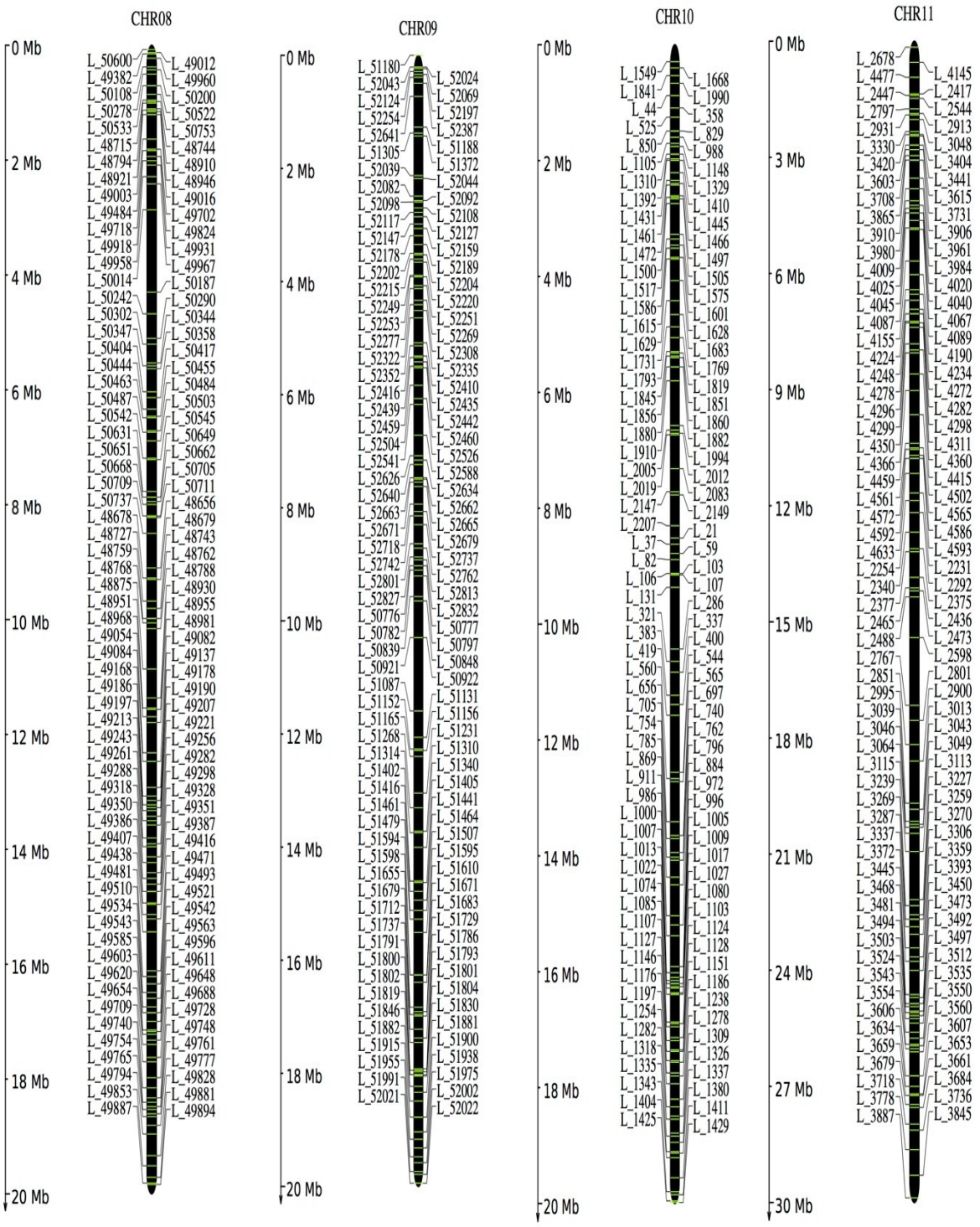

Linkage Map SeaBass
LG11
LG12
LG13
LG14
LG15
LG16
LG17
LG18
LG19
LG20

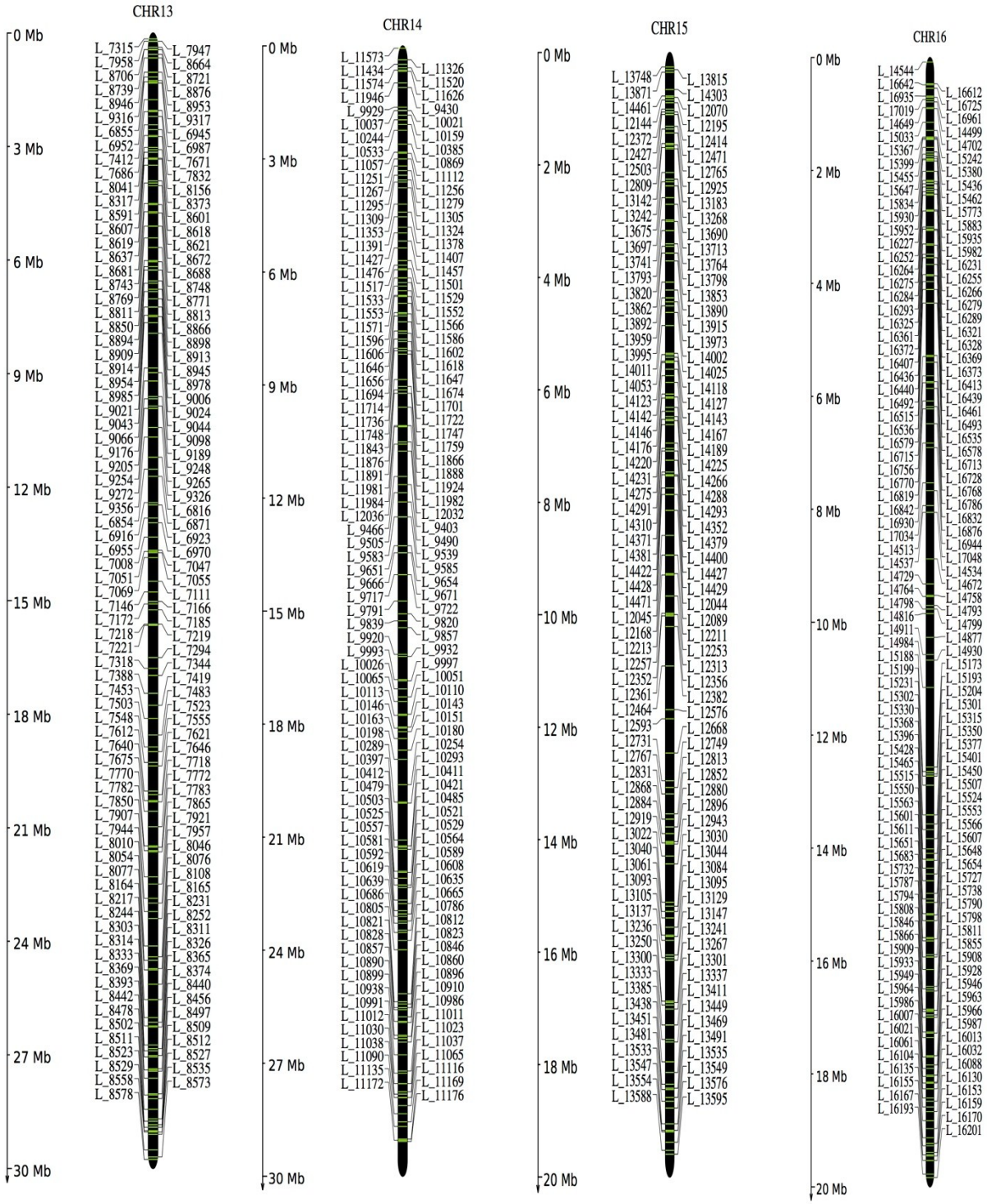

Linkage Map SeaBass
LG21
LG22
LG23
LG24

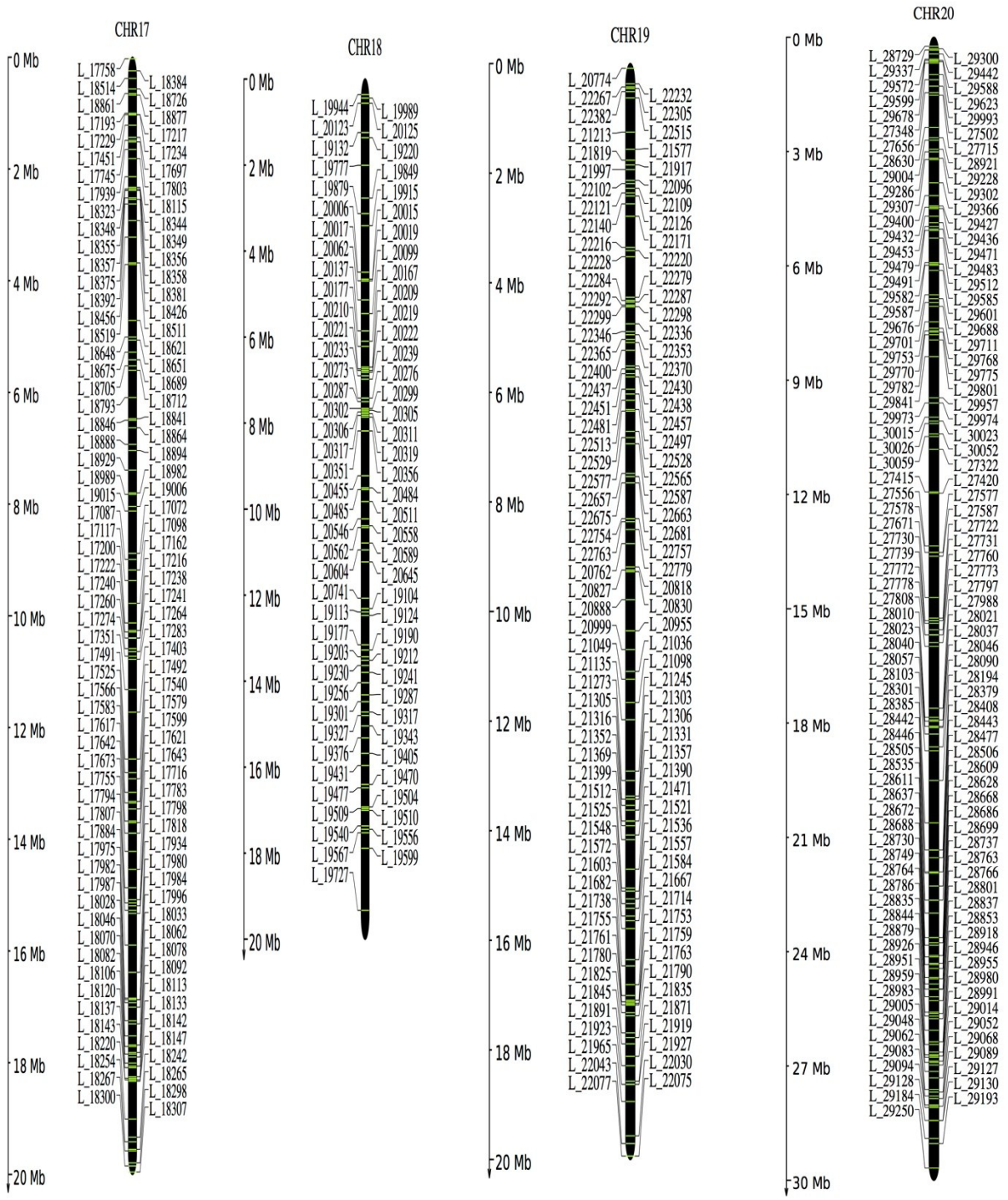
4.2 Physical map.

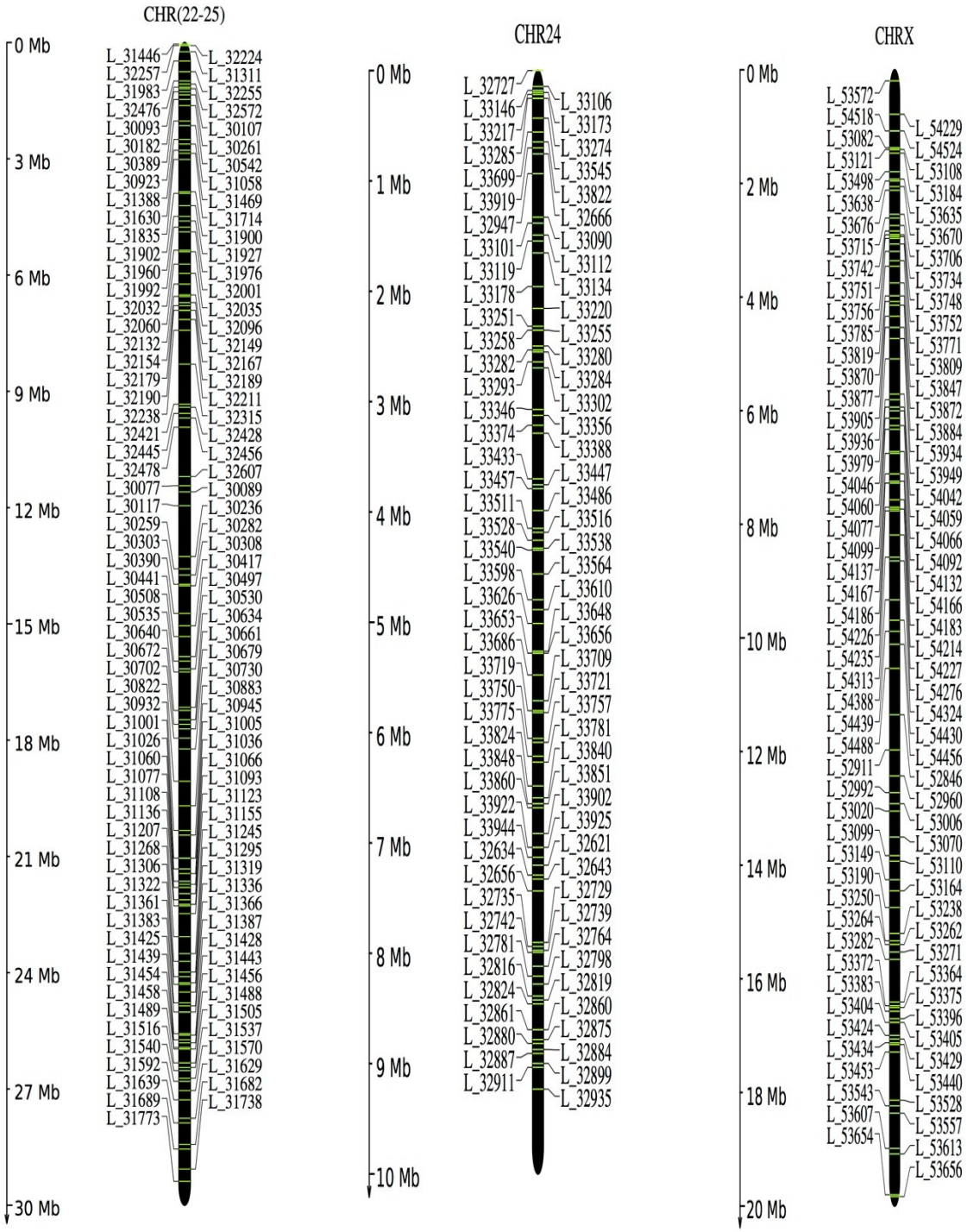


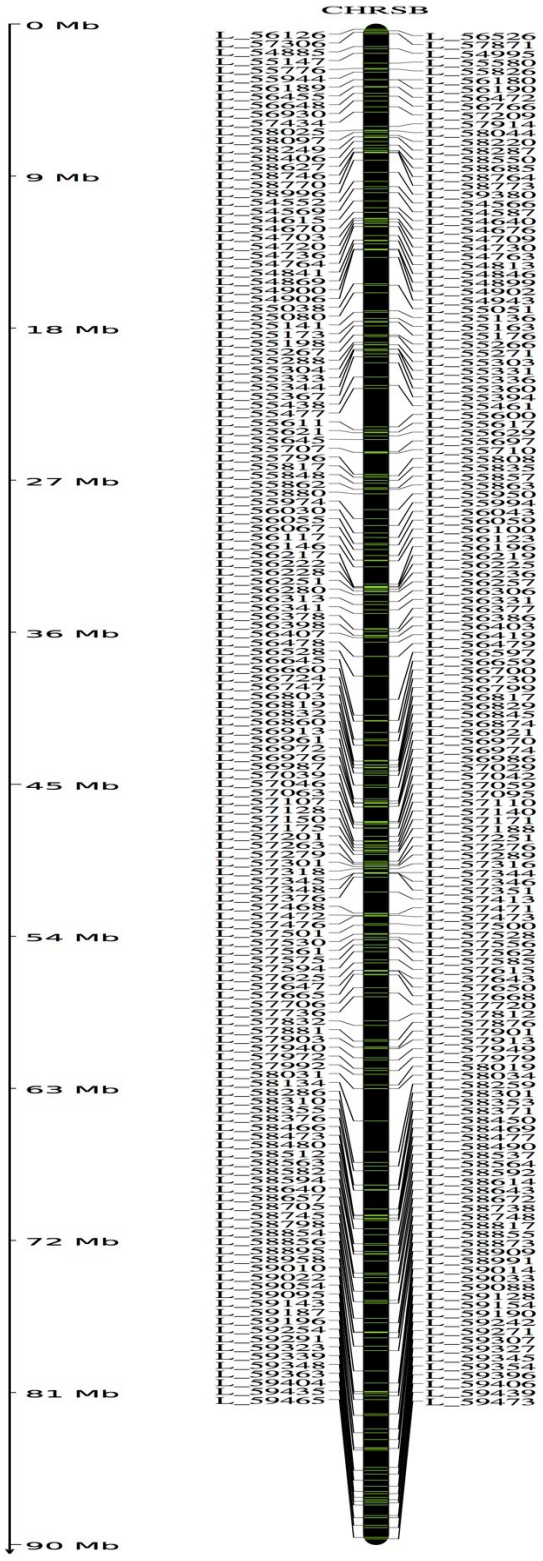

CHR04
CHR05
CHR06
CHR07
0 Mb
3 Mb
6 Mb
9 Mb
12 Mb
15 Mb
18 Mb
21 Mb
24 Mb
27 Mb
30 Mb


CHR08
CHR09
CHR10
CHR11


CHR13
CHR14
CHR15
CHR16


CHR17
CHR18
CHR19
CHR20


CHR(22-25)
0 Mb
3 Mb
6 Mb
9 Mb
12 Mb
15 Mb
18 Mb
21 Mb
24 Mb
27 Mb
30 Mb
L_31446
L_32257
L_31983
L_32476
L_30093
L_30182
L_30389
L_30923
L_31388
L_31630
L_31835
L_31902
L_31960
L_31992
L_32032
L_32060
L_32132
L_32154
L_32179
L_32190
L_32238
L_32421
L_32445
L_32478
L_30077
L_30117
L_30259
L_30303
L_30390
L_30441
L_30508
L_30535
L_30640
L_30672
L_30702
L_30822
L_30932
L_31001
L_31026
L_31060
L_31077
L_31108
L_31136
L_31207
L_31268
L_31306
L_31322
L_31361
L_31383
L_31425
L_31439
L_31454
L_31458
L_31489
L_31516
L_31540
L_31592
L_31639
L_31689
L_31773
L_32224
L_31311
L_32255
L_32572
L_30107
L_30261
L_30542
L_31058
L_31469
L_31714
L_31900
L_31927
L_31976
L_32001
L_32035
L_32096
L_32149
L_32167
L_32189
L_32211
L_32315
L_32428
L_32456
L_32607
L_30089
L_30236
L_30282
L_30308
L_30417
L_30497
L_30530
L_30634
L_30661
L_30679
L_30730
L_30883
L_30945
L_31005
L_31036
L_31066
L_31093
L_31123
L_31155
L_31245
L_31295
L_31319
L_31336
L_31366
L_31387
L_31428
L_31443
L_31456
L_31488
L_31505
L_31537
L_31570
L_31629
L_31682
L_31738
CHR24
0 Mb
1 Mb
2 Mb
3 Mb
4 Mb
5 Mb
6 Mb
7 Mb
8 Mb
9 Mb
10 Mb
L_32727
L_33146
L_33217
L_33285
L_33699
L_33919
L_32947
L_33101
L_33119
L_33178
L_33251
L_33258
L_33282
L_33293
L_33346
L_33374
L_33433
L_33457
L_33511
L_33528
L_33540
L_33598
L_33626
L_33653
L_33686
L_33719
L_33750
L_33775
L_33824
L_33848
L_33860
L_33922
L_33944
L_32634
L_32656
L_32735
L_32742
L_32781
L_32816
L_32824
L_32861
L_32880
L_32887
L_32911
L_33106
L_33173
L_33274
L_33545
L_33822
L_32666
L_33090
L_33112
L_33134
L_33220
L_33255
L_33280
L_33284
L_33302
L_33356
L_33388
L_33447
L_33486
L_33516
L_33538
L_33564
L_33610
L_33648
L_33656
L_33709
L_33721
L_33757
L_33781
L_33840
L_33851
L_33902
L_33925
L_32621
L_32643
L_32729
L_32739
L_32764
L_32798
L_32819
L_32860
L_32875
L_32884
L_32899
L_32935
CHRX
0 Mb
2 Mb
4 Mb
6 Mb
8 Mb
10 Mb
12 Mb
14 Mb
16 Mb
18 Mb
20 Mb
L_53572
L_54518
L_53082
L_53121
L_53498
L_53638
L_53676
L_53715
L_53742
L_53751
L_53756
L_53785
L_53819
L_53870
L_53877
L_53905
L_53936
L_53979
L_54046
L_54060
L_54077
L_54099
L_54137
L_54167
L_54186
L_54226
L_54235
L_54313
L_54388
L_54439
L_54488
L_52911
L_52992
L_53020
L_53099
L_53149
L_53190
L_53250
L_53264
L_53282
L_53372
L_53383
L_53404
L_53424
L_53434
L_53453
L_53543
L_53607
L_53654
L_54229
L_54524
L_53108
L_53184
L_53635
L_53670
L_53706
L_53734
L_53748
L_53752
L_53771
L_53809
L_53847
L_53872
L_53884
L_53934
L_53949
L_54042
L_54059
L_54066
L_54092
L_54132
L_54166
L_54183
L_54214
L_54227
L_54276
L_54324
L_54430
L_54456
L_52846
L_52960
L_53006
L_53070
L_53110
L_53164
L_53238
L_53262
L_53271
L_53364
L_53375
L_53396
L_53405
L_53429
L_53440
L_53528
L_53557
L_53613
L_53656


CHRSB
0 Mb
9 Mb
18 Mb
27 Mb
36 Mb
45 Mb
54 Mb
63 Mb
72 Mb
81 Mb
90 Mb