Genomics 112 (2020) 4297–4303

Contents lists available at ScienceDirect

Genomics

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

New genomic resources for three exploited Mediterranean fishes

Katharina Fietz[a,1], Elena Trofimenko[a,1], Pierre-Edouard Guerin[b,1], Véronique Arnal[b], Montserrat Torres-Oliva[c], Stéphane Lobréaux[d], Angel Pérez-Ruzafa[e], Stéphanie Manel[b,*], Oscar Puebla[f,a]

[a] GEOMAR Helmholtz Centre for Ocean Research Kiel, Evolutionary Ecology of Marine Fishes, Düsternbrooker Weg 20, 24105 Kiel, Germany
[b] CEFE, Univ Montpellier, CNRS, EPHE-PSL University, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France
[c] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, Kiel, Germany
[d] Laboratoire d'Ecologie Alpine, CNRS, Université Grenoble-Alpes, Grenoble, France
[e] Departmento de Ecología e Hidrología, Facultad de Biología, Campus de Espinardo, Regional Campus of International Excellence "Campus Mare Nostrum", University of Murcia, 30100 Murcia, Spain
[f] Leibniz Centre for Tropical Marine Research, Fahrenheitstrasse 6, 28359 Bremen, Germany

ARTICLE INFO

Keywords:
Mediterranean
Fishes
Genomes
Assembly
Annotation
RAD sequencing

ABSTRACT

Extensive fishing has led to fish stock declines throughout the last decades. While clear stock identification is required for designing management schemes, stock delineation is problematic due to generally low levels of genetic structure in marine species. The development of genomic resources can help to solve this issue. Here, we present the first mitochondrial and nuclear draft genome assemblies of three economically important Mediterranean fishes, the white seabream, the striped red mullet, and the comber. The assemblies are between 613 and 785 Mbp long and contain between 27,222 and 32,375 predicted genes. They were used as references to map Restriction-site Associated DNA markers, which were developed with a single-digest approach. This approach provided between 15,710 and 21,101 Single Nucleotide Polymorphism markers per species. These genomic resources will allow uncovering subtle genetic structure, identifying stocks, assigning catches to populations and assessing connectivity. Furthermore, the annotated genomes will help to characterize adaptive divergence.

1. Introduction

Extensive fishing has led to the decline of Mediterranean fish stocks over the last decades [17,63]. Yet the identification of stocks is often problematic due to generally low levels of population genetic structure [8,26,64]. In this situation, a large number of genetic markers is required to detect fine-scale population structure [11,20], assign catches to genetic populations [6] and assess levels of genetic and demographic connectivity [66]. A large number of genetic markers can also contribute to evaluate the effect of marine protected areas (MPAs) on fished areas and optimize the efficiency of MPA networks [67], since MPAs tend to be a reservoir of genetic richness [49]. When genetic markers are mapped to an annotated reference genome of the same or a closely related species, they also provide the opportunity to characterize adaptive divergence [23]. This aspect is particularly relevant in the Mediterranean Sea, which is exposed to extensive anthropogenic pressures [50,58] including global warming [25,47].

Restriction-site Associated DNA (RAD) sequencing [24] and related reduced-representation approaches [61,62,65] have become methods of choice to generate large numbers of Single Nucleotide Polymorphism (SNP) markers. Due to its applicability to non-model organisms, RAD sequencing has revolutionized the fields of ecological and conservation genomics [2]. Yet while the utility of RAD sequencing is well recognized, procedures for library preparation, sequencing and filtering sometimes lack details that are critical to assess the quality of the data and the robustness of the results. For example, PCR clones generated during library preparation can represent a significant proportion of the data and thereby bias allele frequencies if not identified and filtered [3]. This is particularly true when the number of PCR cycles is high, which is often the case when the starting DNA is degraded or present in low concentrations. The number of RAD markers needs to be known in order to adjust the sequencing effort, yet this number is difficult to predict in the absence of a previous study or reference genome. This often results in a sub-optimal sequencing effort, *i.e.* too low or too high

* Corresponding author.
E-mail address: stephanie.manel@ephe.psl.eu (S. Manel).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ygeno.2020.06.041
Received 23 February 2020; Received in revised form 22 June 2020; Accepted 24 June 2020
Available online 03 July 2020
0888-7543/

coverage. The availability of a reference genome allows to estimate the number of RAD markers generated by different restrictions enzymes [36] and can greatly improve genotyping quality by providing a template to call SNP markers [43]. A reference genome also allows to assess physical linkage among markers and to consider population genetic statistics along scaffolds as opposed to a SNP-by-SNP basis [13]. Yet the process of genome assembly and annotation is complex and computationally intensive. It requires high-molecular-weight DNA with high purity and structural integrity, especially when long-read technologies are used [22]. Finally, stringent filtering of SNP markers with respect to sequencing coverage, missing data, minimum allele frequency, and linkage is often required for downstream population genomic analyses.

Here, we present annotated genome assemblies of three exploited Mediterranean fish species from three families, the white seabream (*Diplodus sargus* (Linnaeus, 1758), Sparidae), the striped red mullet (*Mullus surmuletus* (Linnaeus, 1758), Mullidae), and the comber (*Serranus cabrilla* (Linnaeus, 1758), Serranidae). These three species are exploited economically in the Mediterranean Sea [27,40], and previous studies have found weak to no population genetic structure for all of them (*D. sargus* [28,29,35]; *M. surmuletus* [18,42]; *S. cabrilla* [52]). We used these nuclear assemblies as references to map RAD markers and characterize SNPs for the three species, which we filtered stringently with respect to PCR clones, coverage, missing data, minimum allele frequency and linkage.

2. Results

2.1. Genome assemblies

Whole-genome sequencing of *D. sargus*, *M. surmuletus* and *S. cabrilla* with the Illumina HiSeq 4000 platform produced 651, 649 and 755 million paired-end 150 bp reads, respectively. After quality filtering and trimming, 609, 588 and 730 million reads were kept, respectively, and used to assemble each genome with the Platanus assembler [34] (Table 1). First, all paired-end reads were assembled into contigs with N50s of 1101, 384 and 1135 kbp for *D. sargus*, *M. surmuletus* and *S. cabrilla*, respectively. Scaffolds were then built using the mate-pair reads to link contigs into 2344, 2190 and 2940 scaffolds, respectively. The assembly of *D. sargus* reached the highest contiguity, with a scaffold N50 of 3371 kbp. The assemblies of *M. surmuletus* and *S. cabrilla* were overall more fragmented (scaffold N50 of 488 kbp and 613 kbp, respectively), but they also contained very large scaffolds (Table 1) and almost all Benchmarking Universal Single Copy Ortholog genes (BUSCOs). The final size of these *de novo* genome assemblies was 785, 613 and 627 Mbp for *D. sargus*, *M. surmuletus* and *S. cabrilla*, which represents 72%, 103% and 79% of estimated genome size based on C-value [21], respectively. Summary statistics of several fish genome assemblies, including our study species and the best currently available fish assembly of *D. labrax*, are presented in Table S1.

The search for BUSCOs showed the high completeness of the three genome assemblies. From the set of 978 metazoan BUSCOs, the *D. sargus* assembly contains 97.5%, the *M. surmuletus* assembly 92.5% and the *S. cabrilla* assembly 96.7% (Fig. 1A, Table S2). From the set of 4584 Actinopterygii BUSCOs, the *D. sargus* assembly contains 96.6%, the *M. surmuletus* assembly 89.9% and the *S. cabrilla* assembly 95.3% (Fig. 1C, Table S3). These results show that the *D. sargus* assembly is not only the most contiguous, but also the most complete assembly.

The mitochondrial sequences assembled into circular sequences with length of 16,513 bp – 16,620 bp. The mtDNA comprised 37 genes, including 13 protein-coding genes (COX1, COX2, ATP8, ATP6, COX3, ND3, ND4L, ND4, ND5, ND6, CYTB, ND1), 22 transfer RNA genes (tRNA), two ribosomal RNA genes (rRNA) (12S rRNA and 16S rRNA) and the control region (Fig. S1).

2.2. Gene annotation and ortholog gene analysis

The number of predicted genes totaled 31,055 for *M. surmuletus*, 32,375 for *D. sargus* and 27,222 for *S. cabrilla.* To assess these annotations, we compared the percentage of BUSCOs present in the assemblies and investigated how many we could recover as annotated gene models. The *D. sargus* annotation contained 96.0% of complete BUSCOs, which is close to the 97.5% found in the genome assembly. Similar results were obtained for the other two species: the *S. cabrilla* and *M. surmuletus* annotations contained 95.6% and 90.8% metazoan BUSCOs, respectively (Fig. 1B, Table S4). Of the Actinopterygii BUSCOs, the *D. sargus* annotation contained 90.2%, the *S. cabrilla* annotation 87.6% and the *M. surmuletus* annotation 80.3% (Fig. 1D, Table S5).

Using OrthoMCL analysis, we identified genes that are conserved across our focal species and the *D. rerio* reference genome, as well as genes that are unique to our fish species (Fig. 2). Out of the total of 16,432 1:1 orthologs identified, 6446 genes (39%) were shared among all four species and 3577 genes (21%) were shared by our three target species. Just 195 (1.1%), 321 (1.9%) and 266 (1.6%) genes were only present in *D. sargus*, *M. surmuletus* and *S. cabrilla*, respectively. *D. sargus* and *M. surmuletus* share 814 genes, *M. surmuletus* and *S. cabrilla* share 761 genes, while *D. sargus* and *S. cabrilla* share 1312 genes.

2.3. RAD markers prediction

In silico digestion of the three genome assemblies with *SbfI* predicted 30,039, 23,078 and 29,931 restrictions sites for *D. sargus, M. surmuletus* and *S. cabrilla*, respectively, leading to an expected number of 60,078, 46,156 and 59,662 RAD markers for the three species since each restriction site generates two RAD markers (one on each side).

2.4. SNP description

RAD sequencing of 90 individuals generated a total of 49,009, 39,357 and 52,388 RAD markers for *D. sargus, M. surmuletus* and *S. cabrilla*, respectively, which provided a total of 39,678, 31,009, and 47,954 SNPs (Table 2). After applying stringent filtering, we retained 20,074, 15,710 and 21,101 SNPs for *D. sargus*, *M. surmuletus* and *S.*

Table 1
Summary statistics of the genome assembly for each species using Platanus [34]. All statistics are based on contig sizes larger than 4 kbp.

Species	Computing platform	Library	Total size (Mbp)	# of contigs	Contig N50 (kbp)	# of scaffolds	Scaffold N50 (Mbp)	Scaffold L50	Length of largest scaffold (Mbp)	Coverage
D. sargus	MESO@LR	Paired-end	785	2,408,078	1101	2344	3.37	58	2.27	57×
M. surmuletus	MESO@LR[1]	350 bp & 550 bp	613	3,146,055	384	2940	0.49	317	3.28	74×
S. cabrilla	MBB[2]	insert size, mate-pair 3kbp & 5kbp insert size	627	2,169,385	1135	2190	0.61	352	2.76	63×

[1] MESO@LR is 80 cores and 1 Tb RAM.
[2] MBB is 64 cores and 512Gb RAM.

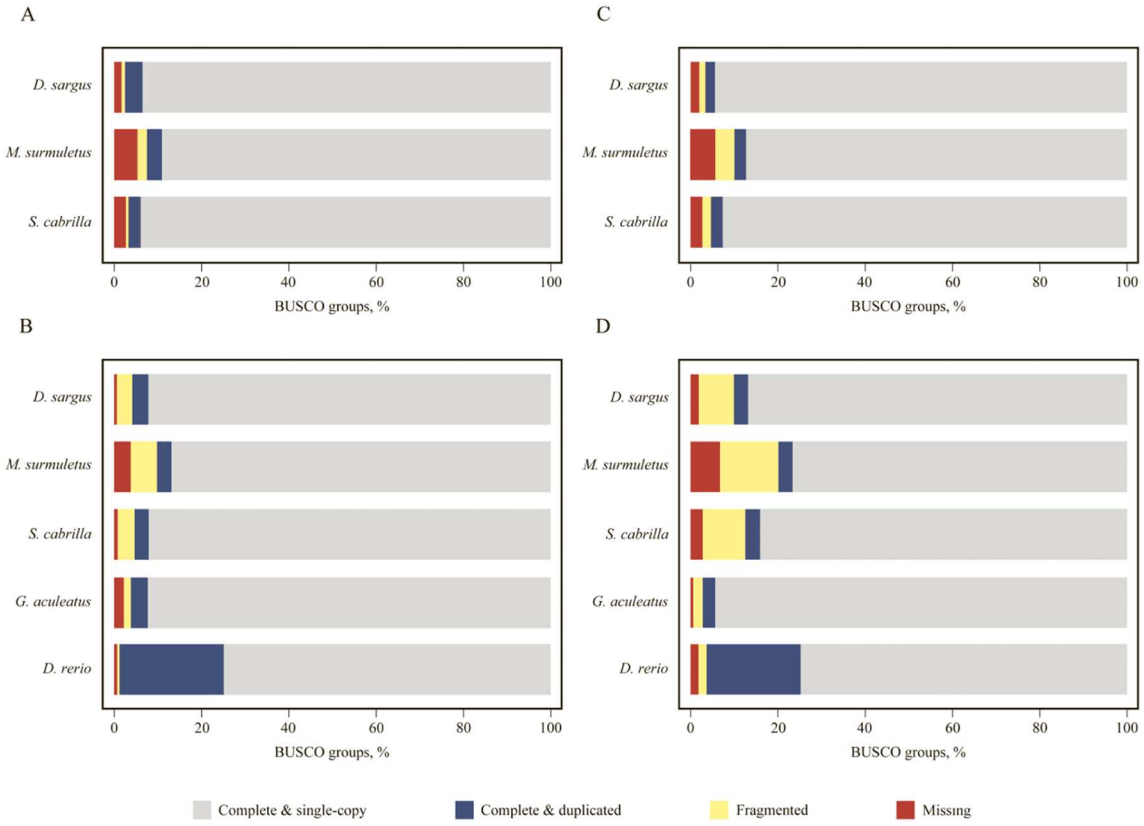

Fig. 1. A) Percentage of conserved Metazoan genes (BUSCOs) found in our Platanus genome assemblies (Table S2); B) percentage of conserved Metazoan genes found in our gene annotations, compared to the annotations of the *D. rerio* and *G. aculeatus* reference genomes used to train the Augustus gene prediction model (Table S4); C) percentage of conserved Actinopterygii genes found in our Platanus genome assemblies (Table S3); D) percentage of conserved Actinopterygii genes found in our gene annotations compared to the annotations the *D. rerio* and *G. aculeatus* reference genomes used to train the Augustus gene prediction model (Table S5). BUSCO stands for Benchmarking Universal Single Copy Ortholog genes.

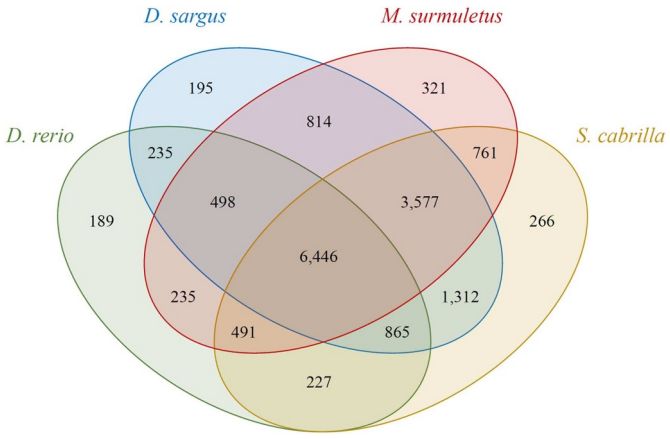

Fig. 2. Four-set Venn diagram of 1:1 orthologous genes shared by *M. surmuletus*, *D. sargus, S. cabrilla* and *D. rerio*. Each ellipse shows the total number of genes specific to each species. Intersections indicate orthologous genes.

cabrilla, respectively, corresponding to 45 – 65% of all SNPs (Table 2). Of these, 173, 178 and 226 were located in the mitochondrial genomes of *D. sargus*, *M. surmuletus* and *S. cabrilla*, respectively, representing less than 1% of the total number SNPs (Table 2). The distance between SNPs averaged 35,389, 30,717 and 28,240 bp per species, respectively. The SNPs were spread evenly across the genomes (Fig. 3 A,C,E), with a mean number of 9.81 SNPs per 400,000 bp window in scaffolds larger than this size. The mean sequencing coverage across individuals was comparable among the three species, with 38×, 45× and 48× for *D. sargus, M. surmuletus* and *S. cabrilla*, respectively (Fig. 3 B,D,F). Of all SNPs filtered, 15%, 18% and 17% were located in exons for *D. sargus, M. surmuletus* and *S. cabrilla,* respectively (Table 2).

3. Discussion

We presented annotated nuclear and mitochondrial genome assemblies of three exploited Mediterranean fishes from three different families, the striped red mullet (*M. surmuletus*, Mullidae), the white seabream (*D. sargus*, Sparidae) and the comber (*S. cabrilla*, Serranidae).

Table 2
Summary statistics for the SNP markers generated by RAD sequencing for each species.

Species	Number of SNPs[1]	Number of filtered SNPs[1]	Average distance (bp) and standard deviation (SD)	SNPs in coding regions[1]	SNPs in exons[2]	Number of mt SNPs
D. sargus	39,678	20,074	35,389 (SD 34,997)	11,978	3138	173
M. surmuletus	31,009	15,710	30,717 (SD 29,190)	10,304	2908	178
S. cabrilla	47,954	21,101	28,240 (SD 27,013)	13,107	3589	226

[1] Nuclear and mitochondrial genomes.
[2] Coding SNPs which are located in exon.

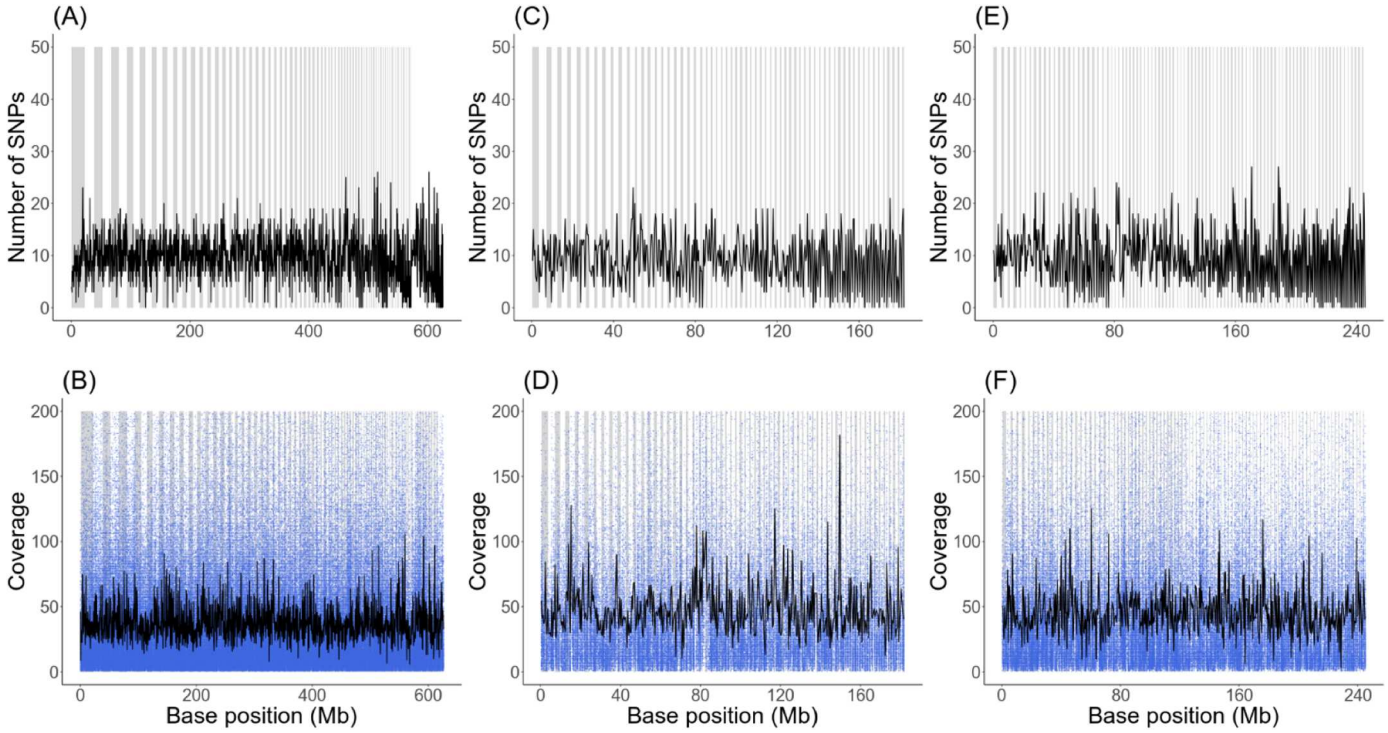

Fig. 3. RADseq coverage along the *D. sargus* (A, B), *M. surmuletus* (C, D) and *S. cabrilla* (E, F) genomes for a total of 90 individuals (30 per species). A), C), E) Number of SNPs per 400,000 bp sliding window along the genome; B), D), F) Coverage per SNP per individual. Each blue dot represents the coverage of one SNP in one individual and the black line represents mean coverage in 400,000 bp sliding windows. Grey and white rectangles represent the assembly scaffolds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To our knowledge, these genome assemblies represent the first genomes for these species.

The quality of a genome assembly in terms of both completeness and continuity greatly influences its usefulness for both genome-wide marker development and gene model prediction [68]. The quality of our three nuclear genome assemblies is attested by the almost complete gene content (89.9–96.6% of Actinopterygii BUSCOs) and by the fact that the sizes of our assemblies are in the expected range based on C-value. The difficulty in genome assembly generally increases with repeat content [53]. Therefore, discrepancies between expected genome size and assembly size from short-read sequencing technologies are still common. In a meta-study of avian genomes, Peona, Weissensteiner, and Suh [48] note that regions highly enriched in repetitive DNA or with strong deviations in nucleotide composition are often underrepresented in assemblies. The presence of such GC-rich or repeat-rich genome regions is a possible explanation for the *ca.* 20–30% gap between assembly sizes and estimated genome sizes for *D. sargus* and *S. cabrilla* in the present study. Comparing contiguity, we found the *D. sargus* genome to be more contiguous (higher scaffold N50, Table 1) than the *M. surmuletus* and *S. cabrilla* genomes. Possible explanations could be higher molecular weight of DNA or a higher homozygosity of *D. sargus* in comparison to *M. surmuletus* and *S. cabrilla* [34].

The mitochondrial genome is essential to eukaryote life and highly conserved across vertebrate species. Our mitochondrial genome assemblies match those of other fishes and vertebrates in terms of size (mean length = 16 kb), the presence of 37 genes (13 protein coding, 22 tRNA, and 2 rRNA genes) and the non-coding control region [51].

In all three genomes, the annotation has identified genes that are highly conserved across metazoans with great accuracy (Fig. 1). For the benefit of providing a resource as flexible as possible, we did not filter annotated gene lists with respect to the presence of a starting codon. For comparison, filtered *G. aculeatus* and *D. rerio* genome annotations contain 20,787 and 26,152 protein-coding genes, respectively, which are fewer compared to our unfiltered output [69,70]. To note is that a significant percentage of Actinopterygii specific orthologs are fragmented (8%, 13% and 9.8% in *D. sargus*, *M. surmuletus* and *S. cabrilla*, respectively). This is probably due to the lack of RNA-seq data for our focal species, which could have allowed the training of specific gene prediction models. However, we can confirm that the three genome annotations are exhaustive and almost complete, as the percentage of missing BUSCOs is low and almost the same in the assembly and annotation (Fig. 1A and B, Tables S2 and S4). The OrthoMCL output revealed that *D. sargus* and *S. cabrilla* share more ortholog genes than the two other species pairs. This is consistent with the phylogeny of the Perciformes, which shows that Mullidae have diverged during the early Lower Cretaceous (LC), while Sparidae and Serranidae have forked during the late LC [46]. As such, *D. sargus* and *S. cabrilla* are more closely related to each other than to *M. surmuletus* [46], which is also supported by phylogenetic findings [71] .

We used the reference genomes to generate rigorously filtered SNP datasets for the three species. Our approach with a single restriction enzyme (*SbfI*) recovered between 82% and 88% of the total number of RAD markers predicted by *in silico* digestion. These RAD markers provided between 31,000 and 47,000 SNPs pre-filtering and between 15,000 and 21,000 SNPs post-filtering that are evenly distributed across the genome. Besides providing a reference to align markers, this exemplifies that we also provide the expected number of markers. This allows knowing exactly what sequencing effort is needed to attain a given coverage. The number of high-quality markers generated here provides strong statistical power for future population genetic analyses. They can for instance be used for stock identification, investigations of population connectivity, and assignment studies. In addition, between 2908 and 3589 of our filtered SNPs lie in exonic regions. These markers may be used to start investigating functional variation (see *e.g.* [20,30]).

This study provides the first genomic resources for three economically important fish species in the Mediterranean Sea and as such lays a

solid foundation for future population and conservation genomic and adaptive studies.

4. Materials and methods

4.1. Genome sequencing

An individual of each species was sampled in the Western Mediterranean Sea (Table S6). Fin tissues of *M. surmuletus* and *D. sargus* were preserved in 96% ethanol at 4 °C prior to DNA extraction, which was done within less than 24 h. Tissues were cut into ~2 mm^2 pieces, dried at 70 °C for 20 min, lysed in proteinase K at 56 °C for 18 h and incubated in RNAse A solution for 10 min at ambient temperature. DNA was extracted with a Macherey-Nagel Nucleospin® kit. For *S. cabrilla*, DNA extraction was conducted directly upon sampling. Tissues were dried out with filter paper and either flash-frozen in liquid nitrogen and crushed or cut into ~2 mm^2 pieces. The fragmented tissues were lysed in proteinase K at 56 °C for 60 min and incubated in RNAse A for 10 min at ambient temperature. DNA was extracted using a Qiagen MagAttract HMW DNA kit.

For each genome, two paired-end libraries with insert sizes of 350 bp and 550 bp were generated from 1 to 2 μg of double-stranded DNA, as well as two mate-pair libraries with insert sizes of 3 kbp and 5 kbp from 4 μg of DNA. Libraries were sequenced on an Illumina HiSeq 4000 platform (150 bp paired-end reads). Library preparation and sequencing were conducted by FASTERIS (https://www.fasteris.com/dna).

4.2. Genome assemblies

Nuclear and mitochondrial genomes were assembled using three computing clusters, the Montpellier Bioinformatics Biodiversity platform (MBB: 64 cores, 512 Gb RAM), the High Performance Computing Platform of Occitanie/Pyrénées-Méditerranée Region of the Montpellier Mediterranean Metropole (MESO@LR: 80 cores, 1 Tb RAM), and CIMENT infrastructure in Grenoble (https://ciment.ujf-grenoble.fr, Froggy: 32 cores, 512 Gb RAM). The entire bioinformatics workflow for genome assembly is described in Fig. S1. Reads with < 50% bp with a phred quality > 20 were filtered out. Adapter sequences were also filtered out and the 3′ extremities of the retained reads were trimmed with ngsShoRT [14]. Finally, reads shorter than 90 bp were removed.

Nuclear genomes were assembled using the Platanus assembler [34] (Fig. S2). Platanus was selected due to its excellent performance with highly heterozygous genomes, as well as with simulated datasets that we produced (data not shown). The paired-end libraries were used to assemble reads into contigs, and both the paired-end and mate-pair libraries were used for scaffolding and gap closing. Mitochondrial genomes were assembled and annotated using MitoZ [45]. Five million sequences were randomly selected as a subset of the full paired-end sequence set. Mitochondrial sequences were then identified from this subset using a ranking method based on a Hidden Markov Model profile of known mitochondrial sequences from 2413 chordate species. Mitochondrial sequences were then used to assemble the mitochondrial genome.

4.3. Gene annotation

Each fish genome was annotated using the *ab initio* gene predictor Augustus v3.2.3 [57] and homology-based extrinsic hints. Each genome was first repeat-masked using RepeatMasker v4.0.8 [56]. Zebrafish (*Danio rerio*) and stickleback (*Gasterosteus aculeatus*) annotated protein sequences were downloaded from the Ensembl website (versions GRCz11 and BROADS1, respectively) and aligned to each repeat-masked fish genome using Exonerate [55]. Untranslated regions (UTRs) and alternative isoforms were not predicted due to the lack of species-specific RNA-seq data. Therefore, in each focal fish species, Augustus was run with the options "–species = zebrafish –UTR = off –alternatives-from-evidence = false" and with the respective Exonerate alignments as extrinsic hints.

All reviewed metazoan proteins were downloaded from UniProt [5] and used as database to run a search in Blast+ v2.2.30 [9]. The highest scoring hit was selected as the putative gene name for each gene model. To functionally annotate the predicted genes, InterProScan v5.19 [33] was run with options "-appl Pfam -b interpro -iprlookup -goterms" and functional information was added to the final annotation dataset using Annie v1.0 [59]. To identify ortholog gene families and species-specific genes in each Mediterranean fish genome, the OrthoMCL pipeline [39] was used on the three annotated protein datasets along with the *D. rerio* protein dataset. Results were visualized with the *venndiagram* R package [15]. Finally, mitochondrial assemblies were annotated using BLAST family alignments on known protein coding genes, transfer RNA genes and rRNA genes.

Quality of the nuclear genome assemblies and annotations were validated against the Metazoan and Actinopterygii Benchmarking Universal Single-Copy Orthologs (BUSCOs) with BUSCO v3.0.2 [54].

4.4. RAD markers prediction

SimRAD [36] was used to perform *in silico* digestion of the three genome assemblies with *SbfI* to predict the number of restriction sites and RAD markers in the three species.

4.5. RAD sequencing

A total of 90 samples (30 per species) from the Western Mediterranean was provided by local artisanal fishermen (Table S7, Fig. S3) and preserved in 96% ethanol. RADseq libraries were prepared using 1 μg of genomic DNA per sample in 50 μl reaction volume. Libraries were prepared following the protocol described in [24] with a few modifications. At step 3.1 (restriction enzyme digestion), DNA was digested with 3 μl of the restriction enzyme *SbfI-HF* (New England Biolabs Inc., USA) in a 50 μl reaction volume. At step 3.2 (P1 adapter ligation), we used 2 μl of barcoded P1 adapters (100 nM) in a 60 μl reaction volume and incubated the samples at room temperature for 1.5 h. Forty-eight samples were pooled per library. At steps 3.4 and 3.5, NEB Next® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs Inc., USA) was used following the manufacturer's instructions to combine DNA end repair, 3′-dA overhang addition and P2 adapter ligation, followed by purification with a Qiagen QIAQuick PCR Purification Kit (Qiagen N.V., Netherlands). Finally, step 3.6 (PCR amplification) was run with the following settings: 30 s 98 °C, 18× (10 s 98 °C, 30 s 68 °C, 30 s 72 °C), 5 min 72 °C, hold 4 °C. P1 and P2 adapter sequences as well as PCR primer sequences are provided in Table S8. Each library was sequenced on one lane of a HiSeq 4000 Illumina Sequencer (paired-end, 2 × 150 bp) at the Institute of Clinical Molecular Biology, Kiel University, Germany.

4.6. SNP calling, genotyping and filtering

PhiX174 sequences that were used for quality control and calibration of the sequencing run were filtered out using BBMap v38.06 [7]. Raw sequences were demultiplexed and filtered using the *process_radtags* pipeline in STACKS v2.2 [12,13]. This included keeping only individuals with >1,000,000 reads at this step, the removal of reads with more than one mismatch in the barcode sequence, and the removal of low-quality reads (with an average raw phred-score < 20 within a 0.2 sliding window). In addition, reads were trimmed to a final length of 139 bp due to a drop in read quality towards the end of the read. Taking advantage of paired-end information, *clone_filter* was used to remove pairs of paired-end reads that matched exactly, as the vast majority of these are expected to be PCR clones. Paired-end read sequences were subsequently aligned with BWA [37] to the reference

genomes of *M. surmuletus*, *D. sargus,* and *S. cabrilla,* thereby improving the reliability of stacks building. Aligned reads were sorted using SAMTOOLS 1.9 [38] and loci were built with *gstacks* providing genotype calls.

In order to retain only high-quality biallelic SNPs for population genetic analysis, called genotypes were further filtered with the *populations* pipeline and vcftools v0.1.16 [19] . Only the first SNP was retained per RAD marker, and a SNP was retained only if present in at least 85% of individuals with a minimum minor allele frequency (MAF) of 1%. In order to reduce linkage among markers, only one locus was retained for all pairs of loci that were closer than 5000 bp or that had an r^2 value >0.8. Finally, individuals with $>30\%$ missing data were filtered out.

Data accessibility statement

The three reference genomes have been uploaded in the European Nucleotide Archive (ENA, project accession number PRJEB38135 at https://www.ebi.ac.uk/ena/browser/view/PRJEB38135), RAD genotypes in: https://gitlab.mbb.univ-montp2.fr/reservebenefit/subset_30_samples_vcf and the scripts in a git repository: https://gitlab.mbb.univ-montp2.fr/reservebenefit/genomic_resources_for_med_fishes.

Animal experiments

All samples were obtained from fishermen. No experiments were conducted.

Author statement

Véronique Arnal processed samples for genome sequencing. Katharina Fietz processed samples for RAD sequencing, and contributed to write the manuscript. Elena Trofimenko processed samples for RAD sequencing. Stéphane Lobreau performed genome assemblies. Pierre-Edouard Guerin performed genome assemblies, processed samples for RAD sequencing, and contributed to write the manuscript. Montserrat Torres-Oliva performed gene annotation. Angel Pérez-Ruzafa contributed to the sampling design and sampling. Stéphanie Manel and Oscar Puebla: Funding acquisition, designed the research, supervised the project and contributed to write the manuscript.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This research was funded through the 2015-2016 BiodivERsA COFUND call for research proposals, with the national funders ANR (France), Formas (Sweden), DLR (Germany), AEI (Spain) and the CNRS for the PICS SEACONNECT.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.06.041.

References

[2] K.R. Andrews, J.M. Good, M.R. Miller, G. Luikart, P.A. Hohenlohe, Harnessing the power of RADseq for ecological and evolutionary genomics, Nat. Rev. Genet. 17 (2) (2016) 81–92, https://doi.org/10.1038/nrg.2015.28.

[3] K.R. Andrews, P.A. Hohenlohe, M.R. Miller, B.K. Hand, J.E. Seeb, G. Luikart, Trade-offs and utility of alternative RADseq methods: reply to Puritz et al. 2014, Mol. Ecol. 23 (24) (2014) 5943–5946, https://doi.org/10.1111/mec.12964.

[5] A. Bateman, M.J. Martin, S. Orchard, M. Magrane, E. Alpi, B. Bely, ... C. UniProt, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Research 47 (D1) (2019) D506–D515, https://doi.org/10.1093/nar/gky1049.

[6] L. Benestan, T. Gosselin, C. Perrier, B. Sainte-Marie, R. Rochette, L. Bernatchez, RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus), Mol. Ecol. 24 (13) (2015) 3299–3315, https://doi.org/10.1111/mec.13245.

[7] B. Bushnell, J. Rood, E. Singer, BBMerge - accurate paired shotgun read merging via overlap, PLoS One 12 (10) (2017), https://doi.org/10.1371/journal.pone.0185056.

[8] A. Calo, I. Muñoz, A. Pérez-Ruzafa, C. Vergara-Chen, J.A. García-Charton, Spatial genetic structure in the saddled sea bream (*Oblada melanura* Linnaeus, 1758) suggests multi-scaled patterns of connectivity between protected and unprotected areas in the Western Mediterranean Sea, Fish. Res. 176 (2016) 30–38, https://doi.org/10.1016/j.fishres.2015.12.001.

[9] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST plus: architecture and applications, Bmc Bioinforma. 10 (2009), https://doi.org/10.1186/1471-2105-10-421.

[11] C. Carreras, V. Ordonez, L. Zane, C. Kruschel, I. Nasto, E. Macpherson, M. Pascual, Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation, Sci. Rep. 7 (2017), https://doi.org/10.1038/srep43417.

[12] J. Catchen, A. Amores, P. Hohenlohe, W. Cresko, J. Postlethwait, Stacks: building and genotyping loci de novo from short-read sequences, G3: Genes Genomes Genet. 1 (2011) 171–182.

[13] J. Catchen, P.A. Hohenlohe, S. Bassham, A. Amores, W.A. Cresko, Stacks: an analysis tool set for population genomics, Mol. Ecol. 22 (11) (2013) 3124–3140, https://doi.org/10.1111/mec.12354.

[14] C. Chen, S.S. Khaleel, H. Huang, C.H. Wu, NgsShoRT: A software for pre-processing Illumina short read sequences for de novo genome assembly, Proceedings of the International Conference on Bioinformatics, 2013, p. 706.

[15] H. Chen, P.C. Boutros, VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R, Bmc Bioinforma. 12 (2011), https://doi.org/10.1186/1471-2105-12-35.

[17] F. Colloca, G. Scarcella, S. Libralato, Recent trends and impacts of fisheries exploitation on Mediterranean stocks and ecosystems, Front. Mar. Sci. 4 (2017), https://doi.org/10.3389/fmars.2017.00244.

[18] A. Dalongeville, M. Andrello, D. Mouillot, S. Lobreaux, M.J. Fortin, F. Lasram, ... S. Manel, Geographic isolation and larval dispersal shape seascape genetic patterns differently according to spatial scale, Evol Appl 11 (8) (2018) 1437–1447, https://doi.org/10.1111/eva.12638.

[19] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, ... Genomes Project Anal, G, The variant call format and VCFtools, Bioinformatics 27 (15) (2011) 2156–2158, https://doi.org/10.1093/bioinformatics/btr330.

[20] J.D. DiBattista, M.J. Travers, G.I. Moore, R.D. Evans, S.J. Newman, M. Feng, ... O. Berry, Seascape genomics reveals fine-scale patterns of dispersal for a reef fish along the ecologically divergent coast of Northwestern Australia, Molecular Ecology 26 (22) (2017) 6206–6223, https://doi.org/10.1111/mec.14352.

[21] J. Dolezel, J. Bartos, H. Voglmayr, J. Greilhuber, Nuclear DNA content and genome size of trout and human, Cytom. Part A 51A (2) (2003) 127–128, https://doi.org/10.1002/cyto.a.10013.

[22] V. Dominguez Del Angel, E. Hjerde, L. Sterck, S. Capella-Gutierrez, C. Notredame, O.V Pettersson, ... H. Lantz, Ten steps to get started in Genome Assembly and Annotation [version 1; referees: 2 approved], F1000Research 148 (2018), https://doi.org/10.12688/f1000research.13598.1 7(ELIXIR).

[23] H. Ellegren, Genome sequencing and population genomics in non-model organisms, Trends Ecol. Evol. 29 (1) (2014) 51–63, https://doi.org/10.1016/j.tree.2013.09.008.

[24] Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. (772), 157-178.

[25] F. Giorgi, Climate change hotspots, Geophys. Res. Lett. 33 (8) (2006).

[26] K. Gkagkavouzis, N. Karaiskou, T. Katopodi, I. Leonardos, T.J. Abatzopoulos, Trianta fyllidis, A., The genetic population structure and temporal genetic stability of gilthead sea bream Sparus aurata populations in the Aegean and Ionian seas, using microsatellite DNA markers, J. Fish Biol. 94 (4) (2019) 606–613, https://doi.org/10.1111/jfb.13932.

[27] R. Goni, S. Adlerstein, D. Alvarez-Berastegui, A. Forcada, O. Renones, G. Criquet, ... S. Planes, Spillover from six western Mediterranean marine protected areas: evidence from artisanal fisheries, Marine Ecology Progress Series 366 (2008) 159–174, https://doi.org/10.3354/meps07532.

[28] M. González-Wanguemert, A. Pérez-Ruzafa, F. Cánovas, J.A. García-Charton, C. Marcos, Temporal genetic variation in populations of Diplodus sargus from the SW Mediterranean Sea, Mar. Ecol. Prog. Ser. 334 (2007) 237–244.

[29] M. González-Wanguemert, A. Pérez-Ruzafa, J.A. García-Charton, C. Marcos, Genetic differentiation and gene flow of two sparidae subspecies, Diplodus sargus sargus and Diplodus sargus cadenati in Atlantic and south-West Mediterranean populations, Biol. J. Linn. Soc. 89 (4) (2006) 705–717, https://doi.org/10.1111/j.1095-8312.2006.00706.x.

[30] B.C. Guo, J. DeFaveri, G. Sotelo, A. Nair, J. Merila, Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks, BMC Biol. 13 (2015) 19, https://doi.org/10.1186/s12915-015-0130-8.

[33] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W.Z. Li, C. McAnulla, ... S. Hunter, InterProScan 5: genome-scale protein function classification, Bioinformatics 30 (9) (2014) 1236–1240, https://doi.org/10.1093/bioinformatics/btu031.

[34] R. Kajitani, K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, ... T. Itoh, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads, Genome Res 24 (8) (2014) 1384–1395, https://doi.org/10.

1101/gr.170720.113.
[35] P. Lenfant, S. Planes, Temporal genetic changes between cohorts in a natural population of a marine fish, Diplodus sargus, Biol. J. Linn. Soc. 76 (1) (2002) 9–20, https://doi.org/10.1046/j.1095-8312.2002.00041.x.
[36] O. Lepais, J.T. Weir, SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches, Mol. Ecol. Resour. 14 (6) (2014) 1314–1321.
[37] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.
[38] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, ... Genome Project Data, P, The Sequence Alignment/Map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.
[39] L. Li, C.J. Stoeckert, D.S. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes, Genome Res. 13 (9) (2003) 2178–2189, https://doi.org/10.1101/gr.1224503.
[40] J. Lloret, T. Font, A comparative analysis between recreational and artisanal fisheries in a Mediterranean coastal area, Fish. Manag. Ecol. 20 (2–3) (2013) 148–160, https://doi.org/10.1111/j.1365-2400.2012.00868.x.
[42] Z. Mamuris, C. Stamatis, C. Triantaphyllidis, Intraspecific genetic variation of striped red mullet (Mullus surmuletus L.) in the Mediterranean Sea assessed by allozyme and random amplified polymorphic DNA (RAPD) analysis, Heredity (Edinb) 83 (1999) 30–38, https://doi.org/10.1038/sj.hdy.6885400.
[43] S. Manel, C. Perrier, M. Pratlong, L. Abi-Rached, J. Paganini, P. Pontarotti, D. Aurelle, Genomic resources and their influence on the detection of the signal of positive selection in genome scans, Mol. Ecol. 25 (1) (2016) 170–184, https://doi.org/10.1111/mec.13468.
[45] G.L. Meng, Y.Y. Li, C.T. Yang, S.L. Liu, MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization, Nucleic Acids Res. 47 (11) (2019), https://doi.org/10.1093/nar/gkz173.
[46] C.N. Meynard, D. Mouillot, N. Mouquet, E.J.P. Douzery, A phylogenetic perspective on the evolution of Mediterranean teleost fishes, PLoS One 7 (5) (2012), https://doi.org/10.1371/journal.pone.0036443.
[47] M.L. Parry, Assessment of Potential Effects and Adaptations for Climate Change in Europe: The Europe ACACIA Project, (2000) Retrieved from.
[48] V. Peona, M.H. Weissensteiner, A. Suh, How complete are “complete” genome assemblies?-an avian perspective, Mol. Ecol. Resour. 18 (6) (2018) 1188–1195, https://doi.org/10.1111/1755-0998.12933.
[49] A. Pérez-Ruzafa, M. González-Wangüemert, P. Lenfant, C. Marcos, J.A. García-Charton, Effects of fishing protection on the genetic structure of fish populations, Biol. Conserv. 129 (2006) 244–255.
[50] F. Ramirez, M. Coll, J. Navarro, J. Bustamante, A.J. Green, Spatial congruence between multiple stressors in the Mediterranean Sea may reduce its resilience to climate impacts, Sci. Rep. 8 (2018), https://doi.org/10.1038/s41598-018-33237-w.
[51] T.P. Satoh, M. Miya, K. Mabuchi, M. Nishida, Structure and variation of the mitochondrial genome of fishes, BMC Genomics 17 (2016), https://doi.org/10.1186/s12864-016-3054-y.
[52] C. Schunter, J. Carreras-Carbonell, E. MacPherson, J. Tintore, E. Vidal-Vijande, A. Pascual, ... M. Pascual, Matching genetics with oceanography: directional gene flow in a Mediterranean fish species, Molecular Ecology 20 (24) (2011) 5167–5181, https://doi.org/10.1111/j.1365-294X.2011.05355.x.
[53] F.J. Sedlazeck, H. Lee, C.A. Darby, M.C. Schatz, Piercing the dark matter: bioinformatics of long-range sequencing and mapping, Nat. Rev. Genet. 19 (6) (2018) 329–346, https://doi.org/10.1038/s41576-018-0003-4.
[54] F.A. Simao, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (19) (2015) 3210–3212, https://doi.org/10.1093/bioinformatics/btv351.
[55] G.S. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison, Bmc Bioinforma. (2005) 6, https://doi.org/10.1186/1471-2105-6-31.
[56] A.F.A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0, (2013).
[57] M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes, Nucleic Acids Res. 32 (2004) W309–W312, https://doi.org/10.1093/nar/gkh379.
[58] A. Stock, L.B. Crowder, B.S. Halpern, F. Micheli, Uncertainty analysis and robust areas of high and low modeled human impact on the global oceans, Conserv. Biol. 32 (6) (2018) 1368–1379, https://doi.org/10.1111/cobi.13141.
[59] R. Tate, B. Hall, T. Derego, Annie the Functional Annotator-Initial Release, (2014) (ZENODO).
[61] R.J. Toonen, J.B. Puritz, Z.H. Forsman, J.L. Whitney, I. Fernandez-Silva, K.R. Andrews, C.E. Bird, ezRAD: a simplified method for genomic genotyping in non-model organisms, Peerj 1 (2013), https://doi.org/10.7717/peerj.203.
[62] N.J. van Orsouw, R.C.J. Hogers, A. Janssen, F. Yalcin, S. Snoeijers, E. Verstege, ... M.J.T. van Eijk, Complexity Reduction of Polymorphic Sequences (CRoPS (TM)): A Novel Approach for Large-Scale Polymorphism Discovery in Complex Genomes, Plos One 2 (11) (2007), https://doi.org/10.1371/journal.pone.0001172.
[63] P. Vasilakopoulos, C.D. Maravelias, G. Tserpes, The alarming decline of Mediterranean fish stocks, Curr. Biol. 24 (14) (2014) 1643–1648, https://doi.org/10.1016/j.cub.2014.05.070.
[64] A. Viret, D. Tsaparis, C.S. Tsigenopoulos, P. Berrebi, A. Sabatini, M. Arculeo, ... E.D.H. Durieux, Absence of spatial genetic structure in common dentex (Dentex dentex Linnaeus, 1758) in the Mediterranean Sea as evidenced by nuclear and mitochondrial molecular markers, Plos One 13 (9) (2018), https://doi.org/10.1371/journal.pone.0203866.
[65] S. Wang, E. Meyer, J.K. McKay, M.V. Matz, 2b-RAD: a simple and flexible method for genome-wide genotyping, Nat. Methods 9 (8) (2012), https://doi.org/10.1038/nmeth.2023 808 − +.
[66] R.S. Waples, Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species, J. Hered. 89 (5) (1998) 438–450, https://doi.org/10.1093/jhered/89.5.438.
[67] A. Xuereb, C.C. D’Aloia, R.M. Daigle, M. Andrello, A. Dalongeville, S. Manel, ... M.J. Fortin, Marine Conservation and Marine Protected Areas, in: S.N.S. AG (Ed.), Population Genomics, Springer Nature, Switzerland, 2019.
[68] X.F. Yang, H.P. Liu, Z.H. Ma, Y. Zou, M. Zou, Y.Z. Mao, ... R.B. Yang, Chromosome-level genome assembly of Triplophysa tibetana, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau, Molecular Ecology Resources 19 (4) (2019) 1027–1036, https://doi.org/10.1111/1755-0998.13021.
[69] Howe K, The zebrafish reference genome sequence and its relationship to the human genome, Nature 496 (2013), https://doi.org/10.1038/nature12111.
[70] Jones FC, et al., The genomic basis of adaptive evolution in threespine sticklebacks, Nature 484 (2012) 55–61, https://doi.org/10.1038/nature10944.
[71] Albouy C, FishMed: traits phylogeny, current and projected species distribution of Mediterranean fishes, and environmental data. Ecology 96 (2015) 2312–2313, https://doi.org/10.1890/14-2279.1.