



HAL
open science

New genomic resources for three exploited Mediterranean fishes

Katharina Fietz, Elena Trofimenko, Pierre-Edouard Guerin, Véronique Arnal, Montserrat Torres-Oliva, Stéphane Lobréaux, Angel Pérez-Ruzafa, Stéphanie Manel, Oscar Puebla

► To cite this version:

Katharina Fietz, Elena Trofimenko, Pierre-Edouard Guerin, Véronique Arnal, Montserrat Torres-Oliva, et al.. New genomic resources for three exploited Mediterranean fishes. *Genomics*, 2020, 112 (6), pp.4297-4303. 10.1016/j.ygeno.2020.06.041 . hal-02997412

HAL Id: hal-02997412

<https://hal.science/hal-02997412v1>

Submitted on 20 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **NEW GENOMIC RESOURCES FOR THREE EXPLOITED MEDITERRANEAN FISHES**

2 **RUNNING TITLE: GENOMIC RESOURCES FOR MEDITERRANEAN FISHES**

3 **Authors**

4 Katharina Fietz^{1*}, Elena Trofimenko^{1*}, Pierre-Edouard Guerin^{2*}, Véronique Arnal², Montserrat
5 Torres-Oliva³, Stéphane Lobréaux⁴, Angel Pérez-Ruzafa⁵, Stephanie Manel^{2**}, Oscar Puebla^{6,1}

6 ¹ GEOMAR Helmholtz Centre for Ocean Research Kiel, Evolutionary Ecology of Marine
7 Fishes, Düsternbrooker Weg 20, 24105 Kiel, Germany

8 ² CEFE, PSL Research University, EPHE, CNRS, Université de Montpellier, Université Paul-
9 Valéry Montpellier 3, IRD, Montpellier, France

10 ³ Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, University
11 Hospital Schleswig-Holstein, Kiel, Germany

12 ⁴ Laboratoire d'Ecologie Alpine, CNRS; Université Grenoble-Alpes, Grenoble, France

13 ⁵ Departamento de Ecología e Hidrología, Facultad de Biología, Campus de Espinardo,
14 Regional Campus of International Excellence "Campus Mare Nostrum", University of
15 Murcia, 30100 Murcia, Spain

16 ⁶ Leibniz Centre for Tropical Marine Research, Fahrenheitstrasse 6, 28359, Bremen, Germany

17

18 * these authors contributed equally to this work

19 ** **corresponding author**: Stephanie Manel, stephanie.manel@ephe.psl.eu

20 **Abstract**

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

34

35 **Keywords**

36 Mediterranean, fishes, genomes, assembly, annotation, RAD sequencing

37 **Introduction**

38 Extensive fishing has led to the decline of Mediterranean fish stocks over the last decades
39 (Colloca, Scarcella, & Libralato, 2017; Vasilakopoulos, Maravelias, & Tserpes, 2014). Yet the
40 identification of stocks is often problematic due to generally low levels of population genetic
41 structure (Calo, Muñoz, Pérez-Ruzafa, Vergara-Chen, & García-Charton, 2016; Gkagkavouzis et
42 al., 2019; Viret et al., 2018). In this situation, a large number of genetic markers is required to
43 detect fine-scale population structure (Carreras et al., 2017; DiBattista et al., 2017), assign
44 catches to genetic populations (Benestan et al., 2015) and assess levels of genetic and
45 demographic connectivity (Waples, 1998). A large number of genetic markers can also
46 contribute to evaluate the effect of marine protected areas (MPAs) on fished areas and optimize
47 the efficiency of MPA networks (Xuereb et al., 2019), since Marine Protected Areas (MPAs)
48 tend to be a reservoir of genetic richness (Pérez-Ruzafa, González-Wangüemert, Lenfant,
49 Marcos, & García-Charton, 2006). When genetic markers are mapped to an annotated reference
50 genome of the same or a closely related species, they also provide the opportunity to characterize
51 adaptive divergence (Ellegren, 2014). This aspect is particularly relevant in the Mediterranean
52 Sea, which is ~~a hotspot in terms of both~~submitted to strong anthropogenic pressures (Ramirez,
53 Coll, Navarro, Bustamante, & Green, 2018; Stock, Crowder, Halpern, & Micheli, 2018) ~~and~~
54 including global warming (Giorgi, 2006; Parry, 2000).

55 Restriction-site Associated DNA (RAD) sequencing (Etter, Bassham, Hohenlohe, Johnson,
56 & Cresko, 2011) and related reduced-representation approaches (Toonen et al., 2013; van
57 Orsouw et al., 2007; Wang, Meyer, McKay, & Matz, 2012) have become methods of choice to
58 generate large numbers of Single Nucleotide Polymorphism (SNP) markers. Due to its
59 applicability to non-model organisms, RAD sequencing has revolutionized the fields of
60 ecological and conservation genomics (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016).

61 Yet while the utility of RAD sequencing is well recognized, procedures for library preparation,
62 sequencing and filtering sometimes lack details that are critical to assess the quality of the data
63 and the robustness of the results. For example, PCR clones generated during library preparation
64 can represent a significant proportion of the data and thereby bias allele frequencies if not
65 identified and filtered (Andrews et al., 2014). This is particularly true when the number of PCR
66 cycles is increased, which is often needed when the starting DNA is degraded or in low
67 concentrations. The number of RAD markers needs to be known in order to adjust the
68 sequencing effort, yet this number is difficult to predict in the absence of a previous study or
69 reference genome. This often results in a sub-optimal sequencing effort, i.e. too low or too high
70 coverage. The availability of a reference genome allows to estimate the number of RAD markers
71 generated by different restrictions enzymes (Lepais & Weir 2014) and can greatly improve
72 genotyping quality by providing a template to call SNP markers (Manel et al., 2016). A reference
73 genome also allows to assess physical linkage among markers and consider population genetic
74 statistics along scaffolds as opposed to a SNP-by-SNP basis (Catchen, Hohenlohe, Bassham,
75 Amores, & Cresko, 2013). Yet the process of genome assembly and annotation is complex and
76 computationally intensive. It requires high-molecular-weight DNA with high purity and
77 structural integrity, especially when long-read technologies are used (Dominguez Del Angel et
78 al., 2018). Finally, stringent filtering of SNP markers with respect to sequencing coverage,
79 missing data, minimum allele frequency, and linkage is often required for downstream
80 population genomic analyses.

81 Here, we present annotated genome assemblies of three exploited Mediterranean fish species
82 from three families, the white seabream (*Diplodus sargus* (Linnaeus, 1758), Sparidae), the
83 striped red mullet (*Mullus surmuletus* (Linnaeus, 1758), Mullidae), and the comber (*Serranus*
84 *cabrilla* (Linnaeus, 1758), Serranidae). These three species are exploited in the Mediterranean

85 Sea (Goni et al., 2008; Lloret & Font, 2013), and previous studies have found weak to no
86 population genetic structure for all of them (*D. sargus*: González-Wanguemert, Pérez-Ruzafa,
87 Canovas, García-Charton, and Marcos (2007); González-Wanguemert, Pérez-Ruzafa, García-
88 Charton, and Marcos (2006); Lenfant and Planes (2002); *M. surmuletus*: Dalongeville et al.
89 (2018); Mamuris, Stamatis, and Triantaphyllidis (1999); *S. cabrilla*: Schunter et al. (2011)). We
90 use these nuclear assemblies as references to map RAD markers and characterize SNPs for the
91 three species, which we filter stringently with respect to PCR clones, coverage, missing data,
92 minimum allele frequency and linkage.

93

94 **Results**

95 *Genome assemblies*

96 Whole-genome sequencing of *D. sargus*, *M. surmuletus* and *S. cabrilla* with the Illumina HiSeq
97 4000 platform produced 651, 649 and 755 million paired-end 150 bp reads, respectively. After
98 quality filtering and trimming, 609, 588 and 730 million reads were kept, respectively, and used
99 to assemble each genome with the Platanus assembler (Kajitani et al., 2014) (Table 1). First, all
100 paired-end reads were assembled into contigs with N50s of 1,101, 384 and 1,135 kbp for *D.*
101 *sargus*, *M. surmuletus* and *S. cabrilla*, respectively. Scaffolds were then built using the mate-pair
102 reads to link contigs into 2,344, 2,190 and 2,940 scaffolds, respectively. The assembly of *D.*
103 *sargus* reached the highest contiguity, with a scaffold N50 of 3,371 kbp. The assemblies of *M.*
104 *surmuletus* and *S. cabrilla* were overall more fragmented (scaffold N50 of 488 kbp and 613 kbp,
105 respectively), but they also contained very large scaffolds (Table 1) and almost all BUSCOs
106 (Benchmarking Universal Single Copy Ortholog genes, see below). The final size of these *de*
107 *novo* genome assemblies was 785, 613 and 627 Mbp for *D. sargus*, *M. surmuletus* and *S.*
108 *cabrilla*, which represents 72%, 103% and 79% of estimated genome size based on C-value

109 (Dolezel, Bartos, Voglmayr, & Greilhuber, 2003), respectively. Summary statistics of several
110 fish genome assemblies, including our study species and the best currently available fish
111 assembly of *D. labrax*, are presented in Table S1.

112 The search for BUSCOs showed the high completeness of the three genome assemblies.
113 From the set of 978 metazoan BUSCOs, the *D. sargus* assembly contains 97.5%, the *M.*
114 *surmuletus* assembly 92.5% and the *S. cabrilla* assembly 96.7% (Fig. 1A, Table S2). From the
115 set of 4,584 Actinopterygii BUSCOs, the *D. sargus* assembly contains 96.6%, the *M. surmuletus*
116 assembly 89.9% and the *S. cabrilla* assembly 95.3% (Fig. 1C, Table S3). These results show that
117 the *D. sargus* assembly is not only the most contiguous, but also the most complete assembly.

118 The mitochondrial sequences assembled into circular sequences with a length of 16,513bp –
119 16,620bp. The mtDNA comprised 37 genes, including 13 protein-coding genes (COX1, COX2,
120 ATP8, ATP6, COX3, ND3, ND4L, ND4, ND5, ND6, CYTB, ND1), 22 transfer RNA genes
121 (tRNA), two ribosomal RNA genes (rRNA) (12S rrn and 16S rrn) and the control region (Fig.
122 S1).

123

124 *Gene annotation and Ortholog gene analysis*

125 The number of predicted genes totaled 31,055 for *M. surmuletus*, 32,375 for *D. sargus* and
126 27,222 for *S. cabrilla*. To assess these annotations, we compared the percentage of BUSCOs
127 present in the assemblies and investigated how many we could recover as annotated gene
128 models. The *D. sargus* annotation contained 96.0% of complete BUSCOs, which is close to the
129 97.5% found in the genome assembly. Similar results were obtained for the other two species:
130 the *S. cabrilla* and *M. surmuletus* annotations contained 95.6% and 90.8% metazoan BUSCOs,
131 respectively (Fig. 1B, Table S4). Of the Actinopterygii BUSCOs, the *D. sargus* annotation

132 contained 90.2%, the *S. cabrilla* annotation 87.6% and the *M. surmuletus* annotation 80.3% (Fig.
133 1D, Table S5).

134 Using OrthoMCL analysis, we identified genes that are conserved across our focal species
135 and the *D. rerio* reference genome, as well as genes that are unique to our fish species (Fig. 2).
136 Out of the total of 16,432 1:1 orthologs identified, 6,446 genes (39%) were shared among all 4
137 species and 3,577 genes (21%) were shared by our 3 target species. Just 195 (1.1%), 321 (1.9%)
138 and 266 (1.6%) genes were only present in *D. sargus*, *M. surmuletus* and *S. cabrilla*,
139 respectively. *D. sargus* and *M. surmuletus* share 814 genes, *M. surmuletus* and *S. cabrilla* share
140 761 genes, while *D. sargus* and *S. cabrilla* share 1,312 genes.

141

142 *RAD markers prediction*

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

147

148 *SNP description*

149 RAD sequencing generated a total of 49,009, 39,357 and 52,388 RAD markers for *D. sargus*, *M.*
150 *surmuletus* and *S. cabrilla*, respectively, which provided a total of 39,678, 31,009, and 47,954
151 SNPs (Table 2). After applying stringent filtering, we retained 20,074, 15,710 and 21,101 SNPs
152 for *D. sargus*, *M. surmuletus* and *S. cabrilla*, respectively, corresponding to 45 – 65 % of all
153 SNPs (Table 2). Of these, 173, 178 and 226 were located in the mitochondrial genomes of *D.*
154 *sargus*, *M. surmuletus* and *S. cabrilla*, respectively, representing less than 1% of the total number
155 SNPs (Table 2). The distance between SNPs averaged 35,389, 30,717 and 28,240 bp per species,

156 respectively. The SNPs were spread evenly across the genomes (Fig. 3 A,C,E), with a mean
157 number of 9.81 SNPs per 400,000 bp window in scaffolds larger than this size. The mean
158 sequencing coverage across individuals was comparable among the three species, with 38X, 45X
159 and 48X for *D. sargus*, *M. surmuletus* and *S. cabrilla*, respectively (Fig. 3 B,D,F). Of all SNPs
160 filtered, 15%, 18% and 17% were located in exons for *D. sargus*, *M. surmuletus* and *S. cabrilla*,
161 respectively (Table 2).

162

163 **Discussion**

164 We presented annotated nuclear and mitochondrial genome assemblies of three exploited
165 Mediterranean fishes from three different families, the striped red mullet (*M. surmuletus*,
166 Mullidae), the white seabream (*D. sargus*, Sparidae) and the comber (*S. cabrilla*, Serranidae).
167 To our knowledge, these genome assemblies represent the first genomes for these species.

168 The quality of a genome assembly in terms of both completeness and continuity greatly
169 influences its usefulness for both genome-wide marker development and gene model
170 prediction (Yang et al., 2019). The quality of our three nuclear genome assemblies is attested
171 by the almost complete gene content (89.9 – 96.6% of Actinopterygii BUSCOs) and by the
172 fact that the sizes of our assemblies are in the expected range based on C-value. The difficulty
173 in genome assembly generally increases with repeat content (Sedlazeck, Lee, Darby, &
174 Schatz, 2018). Therefore, discrepancies between expected genome size and assembly size
175 from short-read sequencing technologies are still common. In a meta-study of avian genomes,
176 Peona, Weissensteiner, and Suh (2018) note that regions highly enriched in repetitive DNA or
177 with strong deviations in nucleotide composition are often underrepresented in assemblies.
178 The presence of such GC-rich or repeat-rich genome regions is a possible explanation for the
179 ca. 20-30% gap between assembly sizes and estimated genome sizes for *D. sargus* and *S.*

180 *cabrilla* in the present study. Comparing contiguity, we found the *D. sargus* genome to be
181 more contiguous (higher scaffold N50, Table 1) than the *M. surmuletus* and *S. cabrilla*
182 genomes. Possible explanations could be higher molecular weight of DNA or a higher
183 homozygosity of *D. sargus* in comparison to *M. surmuletus* and *S. cabrilla* (Kajitani et al.,
184 2014).

185 The mitochondrial genome is essential to eukaryote life and highly conserved across
186 vertebrate species. Our mitochondrial genome assemblies match those of other fishes and
187 vertebrates in terms of size (mean length = 16kb), the presence of 37 genes (13 protein
188 coding, 22 tRNA, and 2 rRNA genes) and the non-coding control region (Satoh, Miya,
189 Mabuchi, & Nishida, 2016).

190 In all three genomes, the annotation has identified genes that are highly conserved across
191 metazoans with great accuracy (Fig. 1). For the benefit of providing a resource as flexible as
192 possible, we did not filter annotated gene lists with respect to the presence of a starting codon.
193 For comparison, filtered *G. aculeatus* and *D. rerio* genome annotations contain 20,787 and
194 26,152 protein-coding genes, respectively, which are fewer compared to our unfiltered output
195 (Howe et al., 2013; Jones et al., 2012). To note is that a significant percentage of
196 Actinopterygii specific orthologs are fragmented (8%, 13% and 9.8% in *D. sargus*, *M.*
197 *surmuletus* and *S. cabrilla*, respectively). This is probably due to the lack of RNA-seq data for
198 our focal species, which could have allowed the training of specific gene prediction models.
199 However, we can confirm that the three genome annotations are exhaustive and **almost**
200 complete, as the percentage of missing BUSCOs is low and almost the same in the assembly
201 and annotation (Fig. 1A and B, Tables S2 and S4). The OrthoMCL output revealed that *D.*
202 *sargus* and *S. cabrilla* share more ortholog genes than the two other species pairs. This is
203 consistent with the phylogeny of the Perciformes, which shows that Mullidae have diverged

204 during the early Lower Cretaceous (LC), while Sparidae and Serranidae have forked during
205 the late LC (Meynard, Mouillot, Mouquet, & Douzery, 2012). As such, *D. sargus* and *S.*
206 *cabrilla* are more closely related to each other than to *M. surmuletus* (Meynard et al., 2012),
207 which is also supported by phylogenetic findings of Albouy *et al.* (2015).

208 We used the reference genomes to generate rigorously filtered SNP datasets for the three
209 species. Our approach with a single restriction enzyme (*SbfI*) recovered between 82% and
210 88% of the total number of RAD markers predicted by *in silico* digestion. These RAD
211 markers provided between 31,000 and 47,000 SNPs pre-filtering and between 15,000 and
212 21,000 SNPs post-filtering that are evenly distributed across the genome. Besides providing a
213 reference to align markers, this exemplifies that we also provide the expected number of
214 markers. This allows knowing exactly what sequencing effort is needed to attain a given
215 coverage. The number of high-quality markers generated here provides strong statistical
216 power for future population genetic analyses. They can for instance be used for stock
217 identification, investigations of population connectivity, and assignment studies. In addition,
218 between 2,908 and 3,589 of our filtered SNPs lie in exonic regions. These markers may be
219 used to start investigating functional variation (see e.g. Guo, DeFaveri, Sotelo, Nair, and
220 Merila (2015), DiBattista et al. (2017)).

221 This study provides the first genomic resources for three economically important fish
222 species in the Mediterranean Sea and as such lays a solid foundation for future population and
223 conservation genomic and adaptive studies.

224

225 **Materials and Methods**

226 *Genome sequencing*

227 An individual of each species was sampled in the Western Mediterranean Sea (Table S6). Fin
228 tissues of *M. surmuletus* and *D. sargus* were preserved in 96% ethanol at 4°C prior to DNA
229 extraction, which was done within less than 24 hours. Tissues were cut into ~2 mm² pieces,
230 dried at 70°C for 20 minutes, lysed in proteinase K at 56°C for 18 hours and incubated in
231 RNase A solution for 10 minutes at ambient temperature. DNA was extracted with a
232 Macherey-Nagel Nucleospin® kit. For *S. cabrilla*, DNA extraction was conducted directly
233 upon sampling. Tissues were dried out with filter paper and either flash-frozen in liquid
234 nitrogen and crushed or cut into ~2 mm² pieces. The fragmented tissues were lysed in
235 proteinase K at 56°C for 60 minutes and incubated in RNase A for 10 minutes at ambient
236 temperature. DNA was extracted using a Qiagen MagAttract HMW DNA kit.

237 For each genome, two paired-end libraries with insert sizes of 350 bp and 550 bp were
238 generated from 1-2 µg of double-stranded DNA, as well as two mate-pair libraries with insert
239 sizes of 3 kbp and 5 kbp from 4 µg of DNA. Libraries were sequenced on an Illumina HiSeq
240 4000 platform (150 bp paired-end reads). Library preparation and sequencing was conducted by
241 FASTERIS (<https://www.fasteris.com/dna>).

242

243 *Genome assemblies*

244 Nuclear and mitochondrial genomes were assembled using three computing clusters, the
245 Montpellier Bioinformatics Biodiversity platform (MBB: 64 cores, 512 Gb RAM), the High
246 Performance Computing Platform of Occitanie / Pyrénées-Méditerranée Region of the
247 Montpellier Mediterranean Metropole (MESO@LR: 80 cores, 1 Tb RAM), and CIMENT
248 infrastructure in Grenoble (<https://ciment.ujf-grenoble.fr>, Froggy: 32 cores, 512 Gb RAM). The
249 entire bioinformatics workflow for genome assembly is described in Appendix S1. Reads with <
250 50% bp with a phred quality > 20 were filtered out. Adapter sequences were also filtered out and

251 the 3' extremities of the retained reads were trimmed with ngsShoRT (C. Chen, Khaleel, Huang,
252 & Wu, 2013). Finally, reads shorter than 90 bp were removed.

253 Nuclear genomes were assembled using the Platanus assembler (Kajitani et al., 2014) (Fig.
254 S2). Platanus was selected due to its excellent performance with highly heterozygous genomes
255 (Kajitani et al., 2014), as well as with simulated datasets that we produced (data not shown). The
256 paired-end libraries were used to assemble reads into contigs, and both the paired-end and mate-
257 pair libraries were used for scaffolding and gap closing. Mitochondrial genomes were assembled
258 and annotated using MitoZ (Meng, Li, Yang, & Liu, 2019). Five million sequences were
259 randomly selected as a subset of the full paired-end sequence set. Mitochondrial sequences were
260 then identified from this subset using a ranking method based on a Hidden Markov Model profile
261 of known mitochondrial sequences from 2,413 chordate species. Mitochondrial sequences were
262 then used to assemble the mitochondrial genome.

263

264 *Gene annotation*

265 Each fish genome was annotated using the *ab initio* gene predictor Augustus v3.2.3 (Stanke,
266 Steinkamp, Waack, & Morgenstern, 2004) and homology-based extrinsic hints. Each genome
267 was first repeat-masked using RepeatMasker v4.0.8 (Smit, Hubley, & Green, 2013). Zebrafish
268 (*Danio rerio*) and stickleback (*Gasterosteus aculeatus*) annotated protein sequences were
269 downloaded from the Ensembl website (versions GRCz11 and BROADS1, respectively) and
270 aligned to each repeat-masked fish genome using Exonerate (Slater & Birney, 2005).
271 Untranslated regions (UTRs) and alternative isoforms were not predicted due to the lack of
272 species-specific RNA-seq data. Therefore, in each focal fish species, Augustus was run with the
273 options “--species=zebrafish --UTR=off --alternatives-from-evidence=false” and with the
274 respective Exonerate alignments as extrinsic hints.

275 All reviewed metazoan proteins were downloaded from UniProt (Bateman et al., 2019) and
276 used as database to run a search in Blast+ v2.2.30 (Camacho et al., 2009). The highest scoring hit
277 was selected as the putative gene name for each gene model. To functionally annotate the
278 predicted genes, InterProScan v5.19 (Jones et al., 2014) was run with options “-appl Pfam -b
279 interpro -iprlookup -goterms” and functional information was added to the final annotation
280 dataset using Annie v1.0 (Tate, Hall, & Derego, 2014). To identify ortholog gene families and
281 species-specific genes in each Mediterranean fish genome, the OrthoMCL pipeline (L. Li,
282 Stoeckert, & Roos, 2003) was used on the three annotated protein datasets along with the *D.*
283 *rerio* protein dataset. Results were visualized with the *venndiagram* R package (H. Chen &
284 Boutros, 2011). Finally, mitochondrial assemblies were annotated using BLAST family
285 alignments on known protein coding genes, transfer RNA genes and rRNA genes.

286 Quality of the nuclear genome assemblies and annotations were validated against the
287 Metazoan and Actinopterygii Benchmarking Universal Single-Copy Orthologs (BUSCOs) with
288 BUSCO v3.0.2 (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015).

289

290 *RAD markers prediction*

291 SimRAD (Lepais & Weir 2014) was used to perform *in silico* digestion of the three genome
292 assemblies with *SbfI* to predict the number of restriction sites and RAD markers in the three
293 species.

294

295 *RAD sequencing*

296 A total of 90 samples (30 per species) from the Western Mediterranean were provided by local
297 artisanal fishermen (Table S7, Fig. S3) and preserved in 96% ethanol. RADseq libraries were
298 prepared using 1 µg of genomic DNA per sample in 50 µl reaction volume. Libraries were

299 prepared following the protocol described in (Etter et al., 2011) with a few modifications. At step
300 3.1 (restriction enzyme digestion), DNA was digested with 3 μ L of the restriction enzyme *SbfI*-
301 *HF* (New England Biolabs Inc., USA) in a 50 μ L reaction volume. At step 3.2 (P1 adapter
302 ligation), we used 2 μ L of barcoded P1 adapters (100nM) in a 60 μ L reaction volume and
303 incubated the samples at room temperature for 1.5 h. Forty-eight samples were pooled per
304 library. At steps 3.4 and 3.5, NEB Next® Ultra™ II DNA Library Prep Kit for Illumina (New
305 England Biolabs Inc., USA) was used following the manufacturer's instructions to combine
306 DNA end repair, 3'-dA overhang addition and P2 adapter ligation, followed by purification with
307 a Qiagen QIAQuick PCR Purification Kit (Qiagen N.V., Netherlands). Finally, step 3.6 (PCR
308 amplification) was run with the following settings: 30 s 98°C, 18x (10 s 98°C, 30 s 68°C, 30 s
309 72°C), 5 min 72°C, hold 4°C. P1 and P2 adapter sequences as well as PCR primer sequences are
310 provided in Table S8. Each library was sequenced on one lane of a HiSeq 4000 Illumina
311 Sequencer (paired-end, 2x 150bp) at the Institute of Clinical Molecular Biology, Kiel University,
312 Germany.

313

314 *SNP calling, genotyping and filtering*

315 PhiX174 sequences that were used for quality control and calibration of the sequencing run were
316 filtered out using BBMap v38.06 (Bushnell, Rood, & Singer, 2017). Raw sequences were
317 demultiplexed and filtered using the *process_radtags* pipeline in STACKS v2.2 (Catchen,
318 Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen et al., 2013). This included keeping
319 only individuals with > 1,000,000 reads at this step, the removal of reads with more than one
320 mismatch in the barcode sequence, and the removal of low-quality reads (with an average raw
321 phred-score <20 within a 0.2 sliding window). In addition, reads were trimmed to a final length
322 of 139bp due to a drop in read quality towards the end of the read. Taking advantage of paired-

323 end information, *clone_filter* was used to remove pairs of paired-end reads that matched exactly,
324 as the vast majority of these are expected to be PCR clones. Paired-end read sequences were
325 subsequently aligned with BWA (H. Li & Durbin, 2009) to the reference genomes of *M.*
326 *surmuletus*, *D. sargus*, and *S. cabrilla*, thereby improving the reliability of stacks building.
327 Aligned reads were sorted using SAMTOOLS 1.9 (H. Li et al., 2009) and loci were built with
328 *gstacks* providing genotype calls.

329 In order to retain only high-quality biallelic SNPs for population genetic analysis, called
330 genotypes were further filtered with the *populations* pipeline and vcftools v0.1.16 (Danecek et
331 al., 2011). Only the first SNP was retained per RAD marker, and a SNP was retained only if
332 present in at least 85% of individuals with a minimum minor allele frequency (MAF) of 1%. In
333 order to reduce linkage among markers, only one locus was retained for all pairs of loci that were
334 closer than 5000 bp or that had an r^2 value >0.8 . Finally, individuals with $>30\%$ missing data
335 were filtered out.

336

337 **Acknowledgements**

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

341

342 **Conflict of interest statement**

343 The authors declare that they have no conflicts of interest.

344

345

346

347 **Data Accessibility Statement**

348 The three reference genomes have been uploaded in the European Nucleotide Archive (ENA,
349 project accession number PRJEB38135 at <https://www.ebi.ac.uk/ena/data/view/PRJEB38135>),
350 the RAD data a git repository: [https://gitlab.mbb.univ-](https://gitlab.mbb.univ-montp2.fr/reservebenefit/subset_30_samples_vcf)
351 [montp2.fr/reservebenefit/subset_30_samples_vcf](https://gitlab.mbb.univ-montp2.fr/reservebenefit/subset_30_samples_vcf) and the scripts in a git repository:
352 https://gitlab.mbb.univ-montp2.fr/reservebenefit/genomic_resources_for_med_fishes.

353

354 **Author contributions**

355 VA processed samples for genome sequencing. KF processed samples for RAD sequencing, and
356 contributed to write the manuscript. ET processed samples for RAD sequencing and wrote a first
357 draft of the manuscript. SL performed genome assemblies. PEG performed genome assemblies,
358 processed samples for RAD sequencing, and contributed to write the manuscript. MTO
359 performed gene annotation. APR contributed to the sampling design and sampling. SM and OP
360 designed the research, supervised the project and contributed to write the manuscript.

361

362 **Animal experiments**

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

364 **Tables and Figures**

365
366 **Table 1:** Summary statistics of the genome assembly for each species using Platannus (Kajitani et al., 2014). All statistics are based on contig
367 sizes larger than 4 kbp.
368

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
<i>D. sargus</i>	MESO@LR	Paired-end 350bp & 550bp	785	2,408,078	1,101	2,344	3.37	58	2.27	57X
<i>M. surmuletus</i>	MESO@LR ¹	insert size, mate-pair 3kbp & 5kbp	613	3,146,055	384	2,940	0.49	317	3.28	74X
<i>S. cabrilla</i>	MBB ²	insert size	627	2,169,385	1,135	2,190	0.61	352	2.76	63X

369

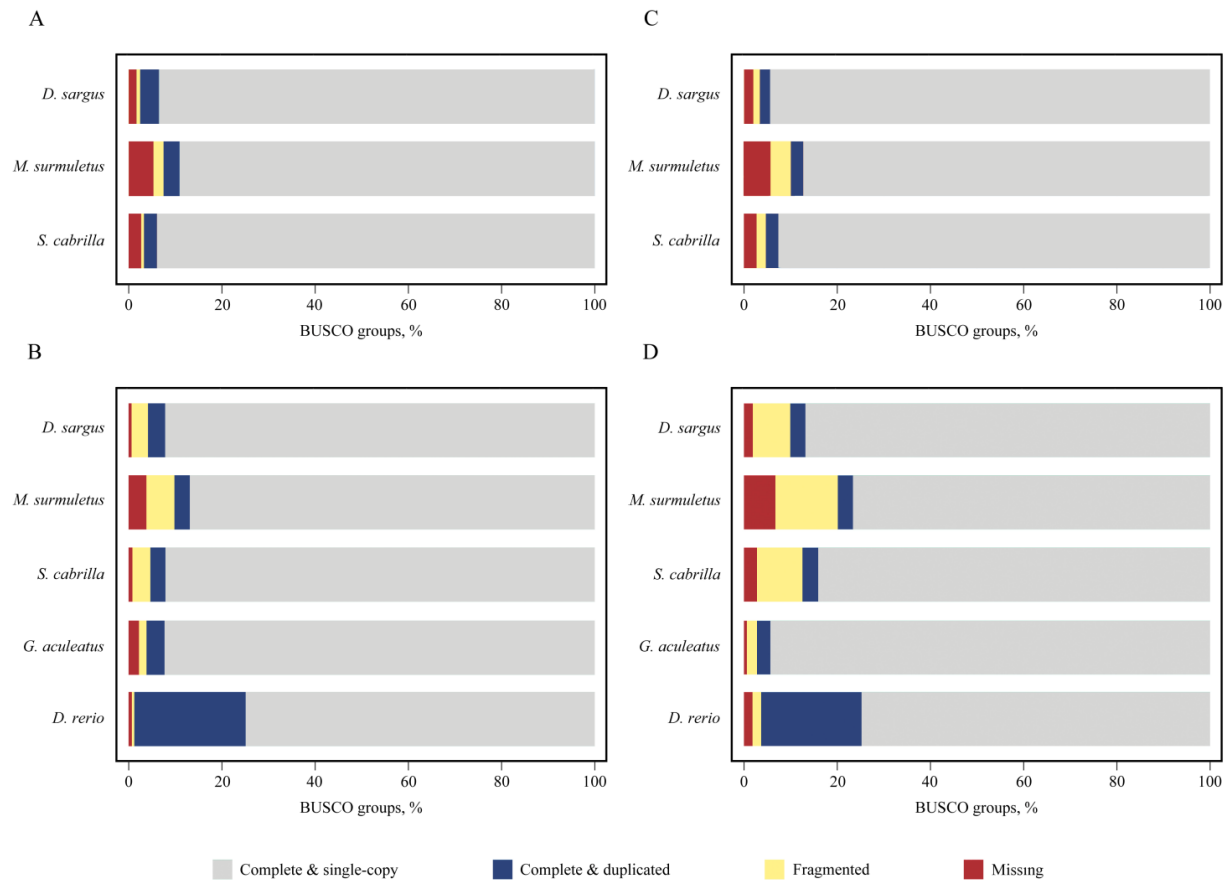
370 ¹ MESO@LR is 80 cores and 1Tb RAM

371 ² MBB is 64 cores and 512Gb RAM

372 **Table 2:** Summary statistics for the SNP markers generated by RAD sequencing for each
373 species.
374

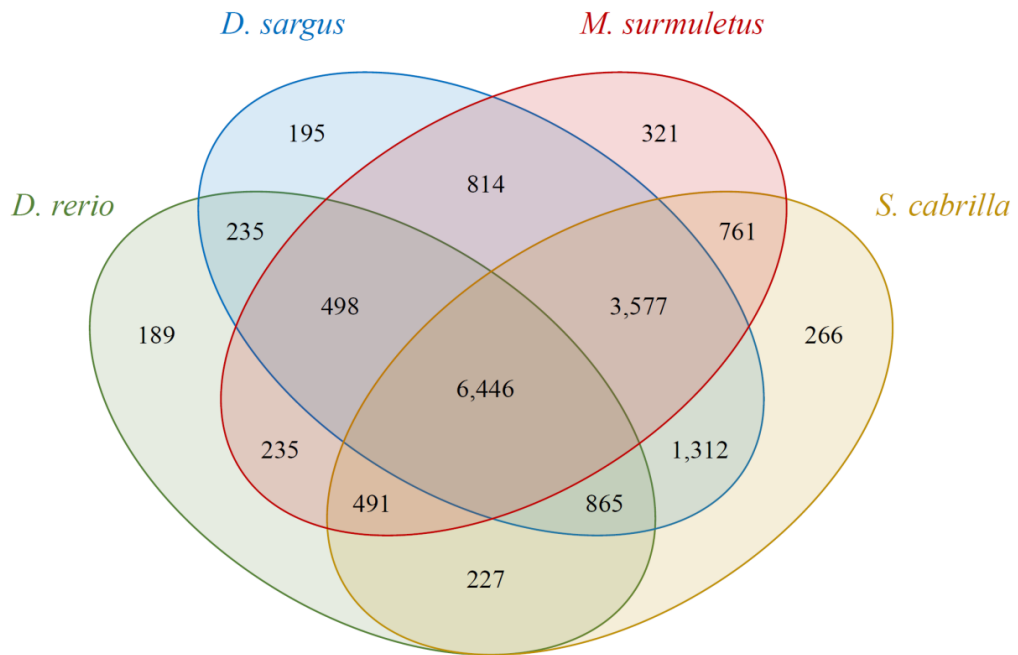
Species	Number of SNPs¹	Number of filtered SNPs¹	Average distance (bp) and standard deviation (SD)	SNPs in coding regions¹	SNPs in exons²	Number of mt SNPs
<i>D. sargus</i>	39,678	20,074	35,389 (SD 34,997)	11,978	3138	173
<i>M. surmuletus</i>	31,009	15,710	30,717 (SD 29,190)	10,304	2908	178
<i>S. cabrilla</i>	47,954	21,101	28,240 (SD 27,013)	13,107	3589	226

375
376 ¹Nuclear and mitochondrial genomes
377 ² Coding SNPs which are located in a exon
378



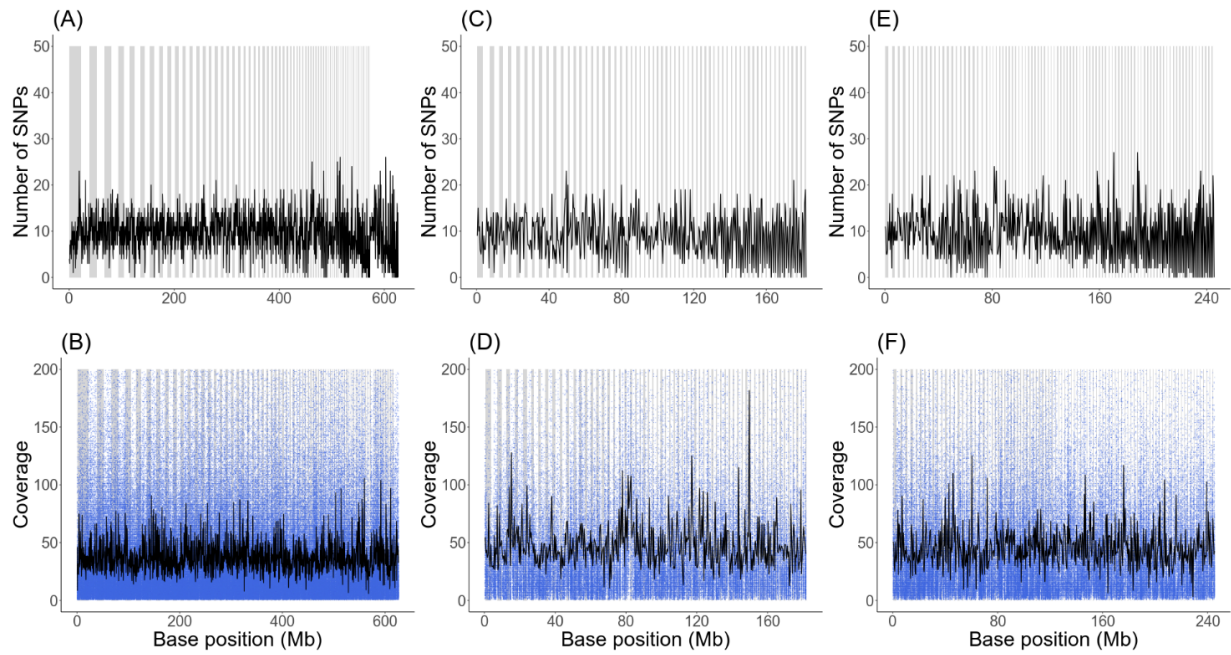
379

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



388
389
390
391

Figure 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.



393
 394
 395
 396
 397
 398
 399

Figure 3: RADseq coverage along the *D. sargus* (A, B), *M. surmuletus* (C, D) and *S. cabrilla* (E, F) genomes. 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.

400 **References**

- 401
- 402 Al-Nakeeb, K., Petersen, T. N., & Sicheritz-Ponten, T. (2017). Norgal: extraction and de novo
403 assembly of mitochondrial DNA from whole-genome sequencing data. *Bmc*
404 *Bioinformatics*, *18*. doi:10.1186/s12859-017-1927-y
- 405 Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016).
406 Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature*
407 *Reviews Genetics*, *17*(2), 81-92. doi:10.1038/nrg.2015.28
- 408 Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart, G.
409 (2014). Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al.
410 2014. *Molecular Ecology*, *23*(24), 5943-5946. doi:10.1111/mec.12964
- 411 Baichoo, S., & Ouzounis, C. A. (2017). Computational complexity of algorithms for sequence
412 comparison, short-read assembly and genome alignment. *Biosystems*, *156*, 72-85.
413 doi:10.1016/j.biosystems.2017.03.003
- 414 Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Alpi, E., Bely, B., . . . UniProt, C.
415 (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*,
416 *47*(D1), D506-D515. doi:10.1093/nar/gky1049
- 417 Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L.
418 (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful
419 population assignment in a widely distributed marine species, the American lobster
420 (*Homarus americanus*). *Molecular Ecology*, *24*(13), 3299-3315.
421 doi:10.1111/mec.13245
- 422 Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge - Accurate paired shotgun read
423 merging via overlap. *Plos One*, *12*(10). doi:10.1371/journal.pone.0185056
- 424 Calo, A., Muñoz, I., Pérez-Ruzafa, A., Vergara-Chen, C., & García-Charton, J. A. (2016).
425 Spatial genetic structure in the saddled sea bream (*Oblada melanura* Linnaeus, 1758)
426 suggests multi-scaled patterns of connectivity between protected and unprotected
427 areas in the Western Mediterranean Sea. *Fisheries Research*, *176*, 30-38.
428 doi:10.1016/j.fishres.2015.12.001
- 429 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden,
430 T. L. (2009). BLAST plus : architecture and applications. *Bmc Bioinformatics*, *10*.
431 doi:10.1186/1471-2105-10-421
- 432 Carnielli, C. M., Winck, F. V., & Leme, A. F. P. (2015). Functional annotation and biological
433 interpretation of proteomics data. *Biochimica Et Biophysica Acta-Proteins and*
434 *Proteomics*, *1854*(1), 46-54. doi:10.1016/j.bbapap.2014.10.019
- 435 Carreras, C., Ordonez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., & Pascual, M.
436 (2017). Population genomics of an endemic Mediterranean fish: differentiation by fine
437 scale dispersal and adaptation. *Scientific Reports*, *7*. doi:10.1038/srep43417
- 438 Catchen, J., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. (2011). Stacks:
439 building and genotyping loci de novo from short-read sequences. *G3: Genes,*
440 *Genomes, Genetics*, *1*, 171-182.
- 441 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an
442 analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140.
443 doi:10.1111/mec.12354
- 444 Chen, C., Khaleel, S. S., Huang, H., & Wu, C. H. (2013). NgsShoRT: a software for pre-
445 processing Illumina short read sequences for de novo genome assembly. *Proceedings*
446 *of the International Conference on Bioinformatics*, 706.

- 447 Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-
448 customizable Venn and Euler diagrams in R. *Bmc Bioinformatics*, *12*.
449 doi:10.1186/1471-2105-12-35
- 450 Chen, Q. F., Lan, C. W., Zhao, L., Wang, J. X., Chen, B. S., & Chen, Y. P. P. (2017). Recent
451 advances in sequence assembly: principles and applications. *Brief Funct Genomics*,
452 *16*(6), 361-378. doi:10.1093/bfgp/elx006
- 453 Colloca, F., Scarcella, G., & Libralato, S. (2017). Recent Trends and Impacts of Fisheries
454 Exploitation on Mediterranean Stocks and Ecosystems. *Frontiers in Marine Science*,
455 *4*. doi:10.3389/fmars.2017.00244
- 456 Dalongeville, A., Andrello, M., Mouillot, D., Lobreaux, S., Fortin, M. J., Lasram, F., . . .
457 Manel, S. (2018). Geographic isolation and larval dispersal shape seascape genetic
458 patterns differently according to spatial scale. *Evol Appl*, *11*(8), 1437-1447.
459 doi:10.1111/eva.12638
- 460 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes
461 Project Anal, G. (2011). The variant call format and VCFtools. *Bioinformatics*,
462 *27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330
- 463 DiBattista, J. D., Travers, M. J., Moore, G. I., Evans, R. D., Newman, S. J., Feng, M., . . .
464 Berry, O. (2017). Seascape genomics reveals fine-scale patterns of dispersal for a reef
465 fish along the ecologically divergent coast of Northwestern Australia. *Molecular*
466 *Ecology*, *26*(22), 6206-6223. doi:10.1111/mec.14352
- 467 Dolezel, J., Bartos, J., Voglmayr, H., & Greilhuber, J. (2003). Nuclear DNA content and
468 genome size of trout and human. *Cytometry Part A*, *51A*(2), 127-128.
469 doi:10.1002/cyto.a.10013
- 470 Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C.,
471 Pettersson, O. V., . . . Lantz, H. (2018). Ten steps to get started in Genome Assembly
472 and Annotation [version 1; referees: 2 approved]. *F1000Research*, *7*(ELIXIR):148.
473 doi:10.12688/f1000research.13598.1
- 474 Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms.
475 *Trends Ecol Evol*, *29*(1), 51-63. doi:10.1016/j.tree.2013.09.008
- 476 Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP
477 discovery and genotyping for evolutionary genetics using RAD sequencing *Methods*
478 *in Molecular Biology*(772), 157-178.
- 479 Giorgi, F. (2006). Climate change hotspots. *Geophysical Research Letters*, *33*(8).
- 480 Gkagkavouzis, K., Karaiskou, N., Katopodi, T., Leonardos, I., Abatzopoulos, T. J., & Trianta
481 fyllidis, A. (2019). The genetic population structure and temporal genetic stability of
482 gilthead sea bream *Sparus aurata* populations in the Aegean and Ionian Seas, using
483 microsatellite DNA markers. *Journal of Fish Biology*, *94*(4), 606-613.
484 doi:10.1111/jfb.13932
- 485 Goni, R., Adlerstein, S., Alvarez-Berastegui, D., Forcada, A., Renones, O., Criquet, G., . . .
486 Planes, S. (2008). Spillover from six western Mediterranean marine protected areas:
487 evidence from artisanal fisheries. *Marine Ecology Progress Series*, *366*, 159-174.
488 doi:10.3354/meps07532
- 489 González-Wanguemert, M., Pérez-Ruzafa, A., Cánovas, F., García-Charton, J. A., & Marcos,
490 C. (2007). Temporal genetic variation in populations of *Diplodus sargus* from the SW
491 Mediterranean Sea. *Marine Ecology Progress Series*, *334*, 237-244.
- 492 González-Wanguemert, M., Pérez-Ruzafa, A., García-Charton, J. A., & Marcos, C. (2006).
493 Genetic differentiation and gene flow of two sparidae subspecies, *Diplodus sargus*
494 *sargus* and *Diplodus sargus cadenati* in Atlantic and south-west Mediterranean

495 populations. *Biological Journal of the Linnean Society*, 89(4), 705-717.
496 doi:10.1111/j.1095-8312.2006.00706.x

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

500 Hahn, C., Bachmann, L., & Chevreur, B. (2013). Reconstructing mitochondrial genomes
501 directly from genomic next-generation sequencing reads—a baiting and iterative
502 mapping approach. *Nucleic Acids Research*, 41(13). doi:10.1093/nar/gkt371

503 Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools:
504 paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids*
505 *Research*, 37(1), 1-13. doi:10.1093/nar/gkn923

506 Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W. Z., McAnulla, C., . . . Hunter, S. (2014).
507 InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9),
508 1236-1240. doi:10.1093/bioinformatics/btu031

509 Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., . . . Itoh, T.
510 (2014). Efficient de novo assembly of highly heterozygous genomes from whole-
511 genome shotgun short reads. *Genome Res*, 24(8), 1384-1395.
512 doi:10.1101/gr.170720.113

513 Lenfant, P., & Planes, S. (2002). Temporal genetic changes between cohorts in a natural
514 population of a marine fish, *Diplodus sargus*. *Biological Journal of the Linnean*
515 *Society*, 76(1), 9-20. doi:10.1046/j.1095-8312.2002.00041.x

516 Lepais, O., & Weir, J. T. (2014). SimRAD: an R package for simulation-based prediction of
517 the number of loci expected in RADseq and similar genotyping by sequencing
518 approaches. *Molecular Ecology Resources*, 14(6), 1314-1321.

519 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
520 transform. *Bioinformatics*, 25(14), 1754-1760. doi:10.1093/bioinformatics/btp324

521 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project
522 Data, P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*,
523 25(16), 2078-2079. doi:10.1093/bioinformatics/btp352

524 Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups
525 for eukaryotic genomes. *Genome Res*, 13(9), 2178-2189. doi:10.1101/gr.1224503

526 Lloret, J., & Font, T. (2013). A comparative analysis between recreational and artisanal
527 fisheries in a Mediterranean coastal area. *Fisheries Management and Ecology*, 20(2-
528 3), 148-160. doi:10.1111/j.1365-2400.2012.00868.x

529 Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome
530 scans for selection based on principal component analysis. *Molecular Ecology*
531 *Resources*, 17(1), 67-77. doi:10.1111/1755-0998.12592

532 Mamuris, Z., Stamatis, C., & Triantaphyllidis, C. (1999). Intraspecific genetic variation of
533 striped red mullet (*Mullus surmuletus* L.) in the Mediterranean Sea assessed by
534 allozyme and random amplified polymorphic DNA (RAPD) analysis. *Heredity*
535 (*Edinb*), 83, 30-38. doi:10.1038/sj.hdy.6885400

536 Manel, S., Perrier, C., Pralong, M., Abi-Rached, L., Paganini, J., Pontarotti, P., & Aurelle, D.
537 (2016). Genomic resources and their influence on the detection of the signal of
538 positive selection in genome scans. *Molecular Ecology*, 25(1), 170-184.
539 doi:10.1111/mec.13468

540 Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Pinero, D., & Emerson, B. C.
541 (2015). Restriction site-associated DNA sequencing, genotyping error estimation and

542 de novo assembly optimization for population genetic inference. *Molecular Ecology*
543 *Resources*, 15(1), 28-41. doi:10.1111/1755-0998.12291

544 Meng, G. L., Li, Y. Y., Yang, C. T., & Liu, S. L. (2019). MitoZ: a toolkit for animal
545 mitochondrial genome assembly, annotation and visualization. *Nucleic Acids*
546 *Research*, 47(11). doi:10.1093/nar/gkz173

547 Meynard, C. N., Mouillot, D., Mouquet, N., & Douzery, E. J. P. (2012). A Phylogenetic
548 Perspective on the Evolution of Mediterranean Teleost Fishes. *Plos One*, 7(5).
549 doi:10.1371/journal.pone.0036443

550 Parry, M. L. (2000). *Assessment of potential effects and adaptations for climate change in*
551 *Europe: The Europe ACACIA project*. Retrieved from

552 Peona, V., Weissensteiner, M. H., & Suh, A. (2018). How complete are "complete" genome
553 assemblies?-An avian perspective. *Molecular Ecology Resources*, 18(6), 1188-1195.
554 doi:10.1111/1755-0998.12933

555 Pérez-Ruzafa, A., González-Wangüemert, M., Lenfant, P., Marcos, C. & García-Charton,
556 J.A., (2006). Effects of fishing protection on the genetic structure of fish populations.
557 *Biological Conservation*, 129: 244-255.

558 Ramirez, F., Coll, M., Navarro, J., Bustamante, J., & Green, A. J. (2018). Spatial congruence
559 between multiple stressors in the Mediterranean Sea may reduce its resilience to
560 climate impacts. *Scientific Reports*, 8. doi:10.1038/s41598-018-33237-w

561 Satoh, T. P., Miya, M., Mabuchi, K., & Nishida, M. (2016). Structure and variation of the
562 mitochondrial genome of fishes. *Bmc Genomics*, 17. doi:10.1186/s12864-016-3054-y

563 Schunter, C., Carreras-Carbonell, J., MacPherson, E., Tintore, J., Vidal-Vijande, E., Pascual,
564 A., . . . Pascual, M. (2011). Matching genetics with oceanography: directional gene
565 flow in a Mediterranean fish species. *Molecular Ecology*, 20(24), 5167-5181.
566 doi:10.1111/j.1365-294X.2011.05355.x

567 Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018). Piercing the dark matter:
568 bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics*,
569 19(6), 329-346. doi:10.1038/s41576-018-0003-4

570 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015).
571 BUSCO: assessing genome assembly and annotation completeness with single-copy
572 orthologs. *Bioinformatics*, 31(19), 3210-3212. doi:10.1093/bioinformatics/btv351

573 Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence
574 comparison. *Bmc Bioinformatics*, 6. doi:10.1186/1471-2105-6-31

575 Smit, A. F. A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0.

576 Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a web server
577 for gene finding in eukaryotes. *Nucleic Acids Research*, 32, W309-W312.
578 doi:10.1093/nar/gkh379

579 Stock, A., Crowder, L. B., Halpern, B. S., & Micheli, F. (2018). Uncertainty analysis and
580 robust areas of high and low modeled human impact on the global oceans.
581 *Conservation Biology*, 32(6), 1368-1379. doi:10.1111/cobi.13141

582 Tate, R., Hall, B., & Derego, T. (2014). Annie the functional annotator-initial release.
583 ZENODO.

584 Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S. T., . . .
585 Reinhardt, R. (2014). European sea bass genome and its variation provide insights into
586 adaptation to euryhalinity and speciation. *Nature Communications*, 5.
587 doi:10.1038/ncomms6770

588 Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K.
589 R., & Bird, C. E. (2013). ezRAD: a simplified method for genomic genotyping in non-
590 model organisms. *PeerJ*, *1*. doi:10.7717/peerj.203
591 van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., . . .
592 van Eijk, M. J. T. (2007). Complexity Reduction of Polymorphic Sequences (CRoPS
593 (TM)): A Novel Approach for Large-Scale Polymorphism Discovery in Complex
594 Genomes. *Plos One*, *2*(11). doi:10.1371/journal.pone.0001172
595 Vasilakopoulos, P., Maravelias, C. D., & Tserpes, G. (2014). The Alarming Decline of
596 Mediterranean Fish Stocks. *Current Biology*, *24*(14), 1643-1648.
597 doi:10.1016/j.cub.2014.05.070
598 Viret, A., Tsaparis, D., Tsigenopoulos, C. S., Berrebi, P., Sabatini, A., Arculeo, M., . . .
599 Durieux, E. D. H. (2018). Absence of spatial genetic structure in common dentex
600 (*Dentex dentex* Linnaeus, 1758) in the Mediterranean Sea as evidenced by nuclear and
601 mitochondrial molecular markers. *Plos One*, *13*(9). doi:10.1371/journal.pone.0203866
602 Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible
603 method for genome-wide genotyping. *Nature Methods*, *9*(8), 808-+.
604 doi:10.1038/nmeth.2023
605 Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation
606 in high gene flow species. *Journal of Heredity*, *89*(5), 438-450.
607 doi:10.1093/jhered/89.5.438
608 Xuereb, A., D'Aloia, C. C., Daigle, R. M., Andrello, M., Dalongeville, A., Manel, S., . . .
609 Fortin, M. J. (2019). Marine Conservation and Marine Protected Areas. In S. N. S. AG
610 (Ed.), *Population Genomics*. Switzerland: Springer Nature.
611 Yang, X. F., Liu, H. P., Ma, Z. H., Zou, Y., Zou, M., Mao, Y. Z., . . . Yang, R. B. (2019).
612 Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the
613 harsh high-altitude environment of the Tibetan Plateau. *Molecular Ecology Resources*,
614 *19*(4), 1027-1036. doi:10.1111/1755-0998.13021

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.

NEW GENOMIC RESOURCES FOR THREE EXPLOITED MEDITERRANEAN FISHES

Supplementary Information

Katharina Fietz, Elena Trofimenko, Pierre-Edouard Guerin, Véronique Arnal, Montserrat Torres-Oliva, Stéphane Lobréaux, Angel Pérez-Ruzafa, Stephanie Manel, Oscar Puebla ([see main text](#) for affiliations)

Table S1: Summary statistics of several [fish](#) genomes, including [the ones assembled in this study](#) and [one of the](#) the best currently available [fish](#) genome (*D. labrax*). Sources: AlMomin et al. (2016)¹, Tine et al. (2014)², Domingos, Zenger, and Jerry (2015)³, Xu et al. (2016)⁴, Shin et al. (2014)⁵, Nakamura et al. (2013)⁶

Species	Year	Assembler	Library	Coverage	N50 contig (kbp)	N50 scaffold (kbp)
<i>Pampus argenteus</i> ¹	2016	SOAP <i>de novo</i>	Paired-end 500bp insert size	58X	0.5	1.5
<i>Dicentrarchus labrax</i> ²	2015	Celera assembler	Paired-end 300bp & 500bp & 1kbp insert size, Mate-pair 8kbp & 20 kbp insert size	30X	53	5100
<i>Lates calcarifer</i> ³	2015	Velvet	Paired-end 330bp insert size	46X	15	311
<i>Miichthys miiuy</i> ⁴	2016	AllPaths-LG	Paired-end 180bp to 800bp insert size, Mate-pair 3kbp & 8kbp & 20kbp insert size	158X	81	1150
<i>Notothenia coriiceps</i> ⁵	2014	Celera / Pbjelly	Paired-end 150bp & 300bp & 500bp & 600bp insert size, Mate-pair 3kbp & 5kbp & 8kbp & 20kbp insert size	38X	17	217
<i>Thunnus orientalis</i> ⁶	2013	Newbler	Paired-end 300bp & 600bp insert size, Mate-pair 3kbp & 20kbp insert size	43X	82	136
<i>Diplodus sargus</i>	This study	Platanus	Paired-end 350bp & 550bp insert size,	57X	1.1	3371
<i>Mullus surmuletus</i>	This study	Platanus	Mate-pair 3kbp & 5kbp insert size	74X	0.38	488
<i>Serranus cabrilla</i>	This study	Platanus		63X	1.1	614

Table S2: Percentage of conserved Metazoan genes (BUSCOs) found in the genome assemblies of *D. sargus*, *M. surmuletus* and *S. cabrilla*.

	<i>D. sargus</i>	<i>M. surmuletus</i>	<i>S. cabrilla</i>
Complete BUSCOs	97.5%	92.5%	96.7%
Complete and single-copy BUSCOs	93.5%	89.1%	93.9%
Complete and duplicated BUSCOs	4.0%	3.4%	2.8%
Fragmented BUSCOs	0.8%	2.2%	0.6%
Missing BUSCOs	1.7%	5.3%	2.7%

Table S3: Percentage of conserved Actinopterygii genes (BUSCOs), found in the genome assemblies of *D. sargus*, *M. surmuletus* and *S. cabrilla*.

	<i>D. sargus</i>	<i>M. surmuletus</i>	<i>S. cabrilla</i>
Complete BUSCOs	96.6%	89.9%	95.3%
Complete and single-copy BUSCOs	94.4%	87.2%	92.6%
Complete and duplicated BUSCOs	2.2%	2.7%	2.7%
Fragmented BUSCOs	1.4%	4.3%	2.0%
Missing BUSCOs	2.0%	5.8%	2.7%

Table S4: Percentage of conserved Metazoan genes (BUSCOs) found in the gene annotations of *D. sargus*, *M. surmuletus* and *S. cabrilla*.

	<i>D. sargus</i>	<i>M. surmuletus</i>	<i>S. cabrilla</i>
Complete BUSCOs	96.0%	90.8%	95.6%
Complete and single-copy BUSCOs	92.3%	87.3%	92.3%
Complete and duplicated BUSCOs	3.7%	3.5%	3.3%
Fragmented BUSCOs	3.4%	5.9%	3.6%
Missing BUSCOs	0.6%	3.3%	0.8%

Table S5: Percentage of conserved Actinopterygii genes (BUSCOs), found in the gene annotations of *D. sargus*, *M. surmuletus* and *S. cabrilla*.

	<i>D. sargus</i>	<i>M. surmuletus</i>	<i>S. cabrilla</i>
Complete BUSCOs	90.2%	80.3%	87.6%
Complete and single-copy BUSCOs	86.9%	77.1%	84.1%
Complete and duplicated BUSCOs	3.3%	3.2%	3.5%
Fragmented BUSCOs	8.0%	13.3%	9.8%
Missing BUSCOs	1.8%	6.4%	2.6%

Table S6: Tissue collection site, date and depth for the individuals used to assemble the reference genomes.

Species	Date	Location	Latitude (decimal degree)	Longitude (decimal degree)	Depth (m)
<i>M. surmuletus</i>	22.02.17	Mallorca, Spain	39.911	3.132	24
<i>D. sargus</i>	15.07.16	Corsica, France	42.323	8.567	14
<i>S. cabrilla</i>	08.06.18	Banyuls, France	42.472	3.157	7

Table S7: Tissue collection site, date and depth of the samples used for RAD sequencing.

<i>sp.</i>	Label	Date	Latitude (decimal degree)	Longitude (decimal degree)	Location	Depth (m)
<i>Mullus surmuletus</i>	1	10.07.2017	41.964	3.228	Aiguafreda	23.3
	2	12.07.2017	42.469	3.163	Banyuls	12.4
	3	12.07.2017	42.469	3.163	Banyuls	11.7
	4	12.07.2017	42.472	3.157	Banyuls	11.5
	5	12.07.2017	42.472	3.157	Banyuls	10.2
	6	12.07.2017	42.472	3.157	Banyuls	10.2
	7	10.10.2017	37.554	-0.917	Cabo del Agua	32.4
	8	27.09.2017	42.285	3.297	Cadaques	16.8
	9	27.09.2017	42.285	3.297	Cadaques	17
	10	25.09.2017	39.605	3.413	Cala Bona	21
	11	05.10.2017	39.638	3.450	Cala Bona	28
	12	13.11.2017	36.892	-1.935	Cala de San Pedro	63
	13	19.03.2018	39.707	3.468	Cala Ratjada	18.7
	14	06.09.2017	42.323	3.310	Cap de Creus	13.2
	15	18.08.2017	37.079	-1.842	Castillo Macenas	32.4
	16	05.10.2017	39.325	2.970	Colònia Sant Jordi	22
	17	04.12.2017	39.280	2.945	Colònia Sant Jordi	56
	18	26.10.2017	40.034	4.187	Fornells	27.7
	19	20.04.2018	40.073	4.080	Illa Bledes	9.8
	20	28.05.2017	39.829	0.708	Islas Columbretes	94
	21	04.06.2017	41.202	2.142	L'Hospitalet	693
	22	09.08.2017	36.864	-1.980	Las Negras	55.8
	23	06.09.2017	38.220	-0.459	Monte Faro	21.5
	24	17.10.2017	39.880	3.160	Pollença	26
	25	26.06.2017	42.521	3.148	Port Vendres	10
	26	31.08.2017	37.031	-1.867	Río Alias	36
	27	05.09.2017	38.181	-0.444	Santa Pola	21
	28	01.12.2017	38.144	-0.613	Santa Pola	13.5
	29	23.05.2017	39.276	-0.137	Sueca	64.5
	30	31.05.2017	41.027	1.377	Tarragona	129.8
<i>Diplodus sargus</i>	1	12.07.2017	42.514	3.138	Banyuls	19.5
	2	12.07.2017	42.515	3.139	Banyuls	24.5
	3	12.07.2017	42.515	3.139	Banyuls	24.5
	4	12.07.2017	42.515	3.139	Banyuls	24.5
	5	06.09.2017	42.490	3.135	Banyuls	21.7
	6	28.04.2018	39.756	3.250	Betlem	23.8
	7	20.10.2017	39.652	3.474	Cala Ratjada	39.7
	8	27.10.2017	39.776	3.414	Cala Ratjada	32.5
	9	20.11.2017	39.764	3.405	Cala Ratjada	32.2
	10	05.12.2017	39.786	3.440	Cala Ratjada	36.5

11	26.09.2017	39.797	3.201	Can Picafort	27
12	30.03.2018	39.756	3.222	Can Picafort	14.2
13	25.10.2017	43.277	3.517	Cap d'Agde	28.4
14	25.10.2017	43.277	3.517	Cap d'Agde	28.6
15	06.07.2017	42.323	3.310	Cap de Creus	15.4
16	06.07.2017	42.323	3.310	Cap de Creus	15.4
17	06.07.2017	42.323	3.310	Cap de Creus	15.6
18	08.08.2017	39.894	0.687	Columbretes	7
19	09.08.2017	39.889	0.667	Columbretes	3
20	10.08.2017	39.894	0.687	Columbretes	7
21	13.09.2017	38.281	-0.505	El Altet	8
22	10.10.2017	37.550	-0.981	Escombreras	59.4
23	20.10.2017	37.753	-0.744	La Manga	1.8
24	02.11.2017	37.702	-0.752	La Manga	3.6
25	26.07.2017	41.801	3.071	Palamos	17.1
26	26.07.2017	41.801	3.071	Palamos	17.1
27	30.08.2017	41.857	3.149	Palamos	23.4
28	27.07.2017	42.239	3.181	Roses	21.4
29	17.10.2017	41.113	1.302	Tarragone	14.2
30	18.09.2017	37.931	-0.672	Torre Vieja	36
1	17.10.2017	39.874	3.259	Alcudia	56
2	07.09.2017	38.303	-0.464	Alicante	21.5
3	04.09.2017	38.161	-0.528	Alicante	18
4	04.09.2017	38.161	-0.528	Alicante	18
5	04.09.2017	38.161	-0.528	Alicante	18
6	21.09.2017	38.317	-0.447	Alicante	25
7	12.07.2017	42.479	3.148	Banyuls	17.8
8	12.07.2017	42.493	3.134	Banyuls	17
9	16.10.2017	37.583	-0.794	Cabo Negro	21.6
10	23.10.2017	37.585	-0.782	Cabo Negro	21.6
11	15.11.2017	36.892	-1.935	Cala de San Pedro	63
12	27.10.2017	39.824	3.471	Cala Ratjada	53.1
13	02.11.2017	39.650	3.607	Cala Ratjada	76.6
14	20.11.2017	39.804	3.380	Cala Ratjada	44.6
15	11.10.2017	37.591	-0.725	Calblanque	25.2
16	20.07.2017	40.076	4.085	Cap de Cavalleria	18
17	31.08.2017	37.184	-1.807	Garrucha	27
18	31.08.2017	37.197	-1.766	Garrucha	72
19	05.10.2017	39.792	3.734	Holanda	116
20	05.09.2017	38.172	-0.448	Isla de Tabarca	19
21	05.09.2017	38.172	-0.448	Isla de Tabarca	19
22	05.08.2017	39.852	0.682	Islas Columbretes	39.1
23	08.08.2017	39.905	0.632	La Calç	47.3
24	13.09.2017	39.831	4.345	Maó	65.8

Serranus cabrilla

25	01.06.2017	41.816	3.152	Palamos	73
26	17.08.2017	36.795	-2.285	Rambla Morales	34.2
27	08.09.2017	39.825	3.251	S'Aucanada	39
28	16.03.2018	39.881	3.406	Sa Muntanyota	64
29	04.12.2017	39.280	2.945	Santanyí	56
30	04.06.2017	41.148	1.795	Vilanova y la Geltru	55.75

Table S8: RADseq adapter sequences, F=Forward, R=Reverse. Concentrations of P1 adapters are 25nmole, P1 adapter treatment = standard desalting, Phosphorothioate linkage between bases marked with * in forward P1 adapters, P2 reverse adapter and PCR primers, 5' Phosphorylation as indicated in reverse P1 and forward P2 adapters. Concentrations of P2 adapters and PCR primers are 100 nmole, P2 adapter and PCR primer treatment = HPLC purification,

ID	Name	Sequence
1F	P1-FOR-AAACGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAACGGTGC*A
2F	P1-FOR-AACGTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACGTTTGC*A
3F	P1-FOR-AACTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACTGATGC*A
4F	P1-FOR-AAGACG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGACGTGC*A
5F	P1-FOR-AAGCTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCTATGC*A
6F	P1-FOR-AATATC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATATCTGC*A
7F	P1-FOR-AATGAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATGAGTGC*A
8F	P1-FOR-ACAAGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACAAGATGC*A
9F	P1-FOR-ACAGCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACAGCGTGC*A
10F	P1-FOR-ACATAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATACTGC*A
11F	P1-FOR-ACCATG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCATGTGC*A
12F	P1-FOR-ACCCCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCCCCTGC*A
13F	P1-FOR-ACTCTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTCTTTGC*A
14F	P1-FOR-ACTGGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGGCTGC*A
15F	P1-FOR-AGCCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCCATTGC*A
16F	P1-FOR-AGCGCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCGCATGC*A
17F	P1-FOR-AGGGTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGGTCTGC*A
18F	P1-FOR-AGGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTGTTGC*A
19F	P1-FOR-AGTAGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTAGGTGC*A
20F	P1-FOR-AGTTAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTTAATGC*A
21F	P1-FOR-ATAGTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGTATGC*A
22F	P1-FOR-ATCAAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCAAATGC*A
23F	P1-FOR-ATGCAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCACTGC*A
24F	P1-FOR-ATGTTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGTTGTGC*A
25F	P1-FOR-ATTCCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTCCGTGC*A
26F	P1-FOR-CAAAAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAAAATGC*A
27F	P1-FOR-CAATCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAATCGTGC*A
28F	P1-FOR-CACCTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACCTCTGC*A
29F	P1-FOR-CAGGCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGCATGC*A
30F	P1-FOR-CATACT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATACTTGC*A
31F	P1-FOR-CCATTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCATTTTGC*A
32F	P1-FOR-CCCGGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCCGGTTGC*A
33F	P1-FOR-CCCTAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCCTAATGC*A
34F	P1-FOR-CCGAGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGAGGTGC*A
35F	P1-FOR-CCGCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGCATTGC*A
36F	P1-FOR-CCTAAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTAACTGC*A
37F	P1-FOR-CGAGGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGAGGCTGC*A
38F	P1-FOR-CGCAGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCAGATGC*A
39F	P1-FOR-CGCGTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCGTGTGC*A
40F	P1-FOR-CGGTCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGTCTGC*A
41F	P1-FOR-CGTCTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTCTATGC*A
42F	P1-FOR-CGTGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATTGC*A

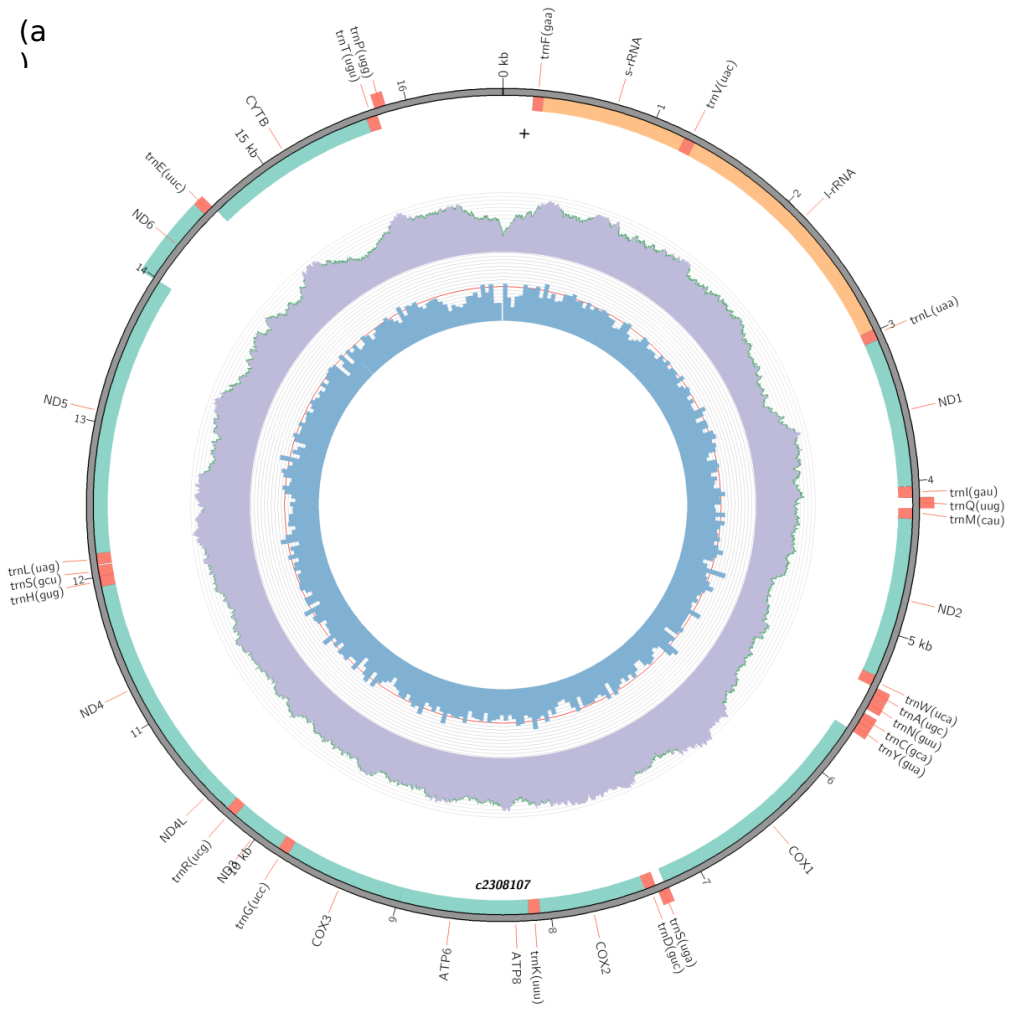
43F	P1-FOR-CTACAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTACAGTGC*A
44F	P1-FOR-CTCGCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCGCCTGC*A
45F	P1-FOR-CTGCGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGCGATGC*A
46F	P1-FOR-CTGGTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGTTTGC*A
47F	P1-FOR-CTTATG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTATGTGC*A
48F	P1-FOR-CTTTGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTTGCTGC*A
1R	P1-REV-AAACGG	/5Phos/CCGTTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
2R	P1-REV-AACGTT	/5Phos/AACGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
3R	P1-REV-AACTGA	/5Phos/TCAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
4R	P1-REV-AAGACG	/5Phos/CGTCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
5R	P1-REV-AAGCTA	/5Phos/TAGCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
6R	P1-REV-AATATC	/5Phos/GATATTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
7R	P1-REV-AATGAG	/5Phos/CTCATTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
8R	P1-REV-ACAAGA	/5Phos/TCTTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
9R	P1-REV-ACAGCG	/5Phos/CGCTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
10R	P1-REV-ACATAC	/5Phos/GTATGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
11R	P1-REV-ACCATG	/5Phos/CATGGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
12R	P1-REV-ACCCCC	/5Phos/GGGGGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
13R	P1-REV-ACTCTT	/5Phos/AAGAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
14R	P1-REV-ACTGGC	/5Phos/GCCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
15R	P1-REV-AGCCAT	/5Phos/ATGGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
16R	P1-REV-AGCGCA	/5Phos/TGCGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
17R	P1-REV-AGGGTC	/5Phos/GACCCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
18R	P1-REV-AGGTGT	/5Phos/ACACCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
19R	P1-REV-AGTAGG	/5Phos/CCTACTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
20R	P1-REV-AGTTAA	/5Phos/TTAACTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
21R	P1-REV-ATAGTA	/5Phos/TACTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
22R	P1-REV-ATCAAA	/5Phos/TTTGATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
23R	P1-REV-ATGCAC	/5Phos/GTGCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
24R	P1-REV-ATGTTG	/5Phos/CAACATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
25R	P1-REV-ATTCCG	/5Phos/CGGAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
26R	P1-REV-CAAAAA	/5Phos/TTTTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
27R	P1-REV-CAATCG	/5Phos/CGATTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
28R	P1-REV-CACCTC	/5Phos/GAGGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
29R	P1-REV-CAGGCA	/5Phos/TGCCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
30R	P1-REV-CATACT	/5Phos/AGTATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
31R	P1-REV-CCATTT	/5Phos/AAATGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
32R	P1-REV-CCCGGT	/5Phos/ACCGGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
33R	P1-REV-CCCTAA	/5Phos/TTAGGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
34R	P1-REV-CCGAGG	/5Phos/CCTCGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
35R	P1-REV-CCGCAT	/5Phos/ATGCGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
36R	P1-REV-CCTAAC	/5Phos/GTTAGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
37R	P1-REV-CGAGGC	/5Phos/GCCTCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
38R	P1-REV-CGCAGA	/5Phos/TCTGCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
39R	P1-REV-CGCGTG	/5Phos/CACGCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
40R	P1-REV-CGGTCC	/5Phos/GGACCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
41R	P1-REV-CGTCTA	/5Phos/TAGACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

42R P1-REV-CGTGAT /5Phos/ATCACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
43R P1-REV-CTACAG /5Phos/CTGTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
44R P1-REV-CTCGCC /5Phos/GGCGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
45R P1-REV-CTGCGA /5Phos/TCGCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
46R P1-REV-CTGGTT /5Phos/AACCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
47R P1-REV-CTTATG /5Phos/CATAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
48R P1-REV-CTTTGC /5Phos/GCAAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

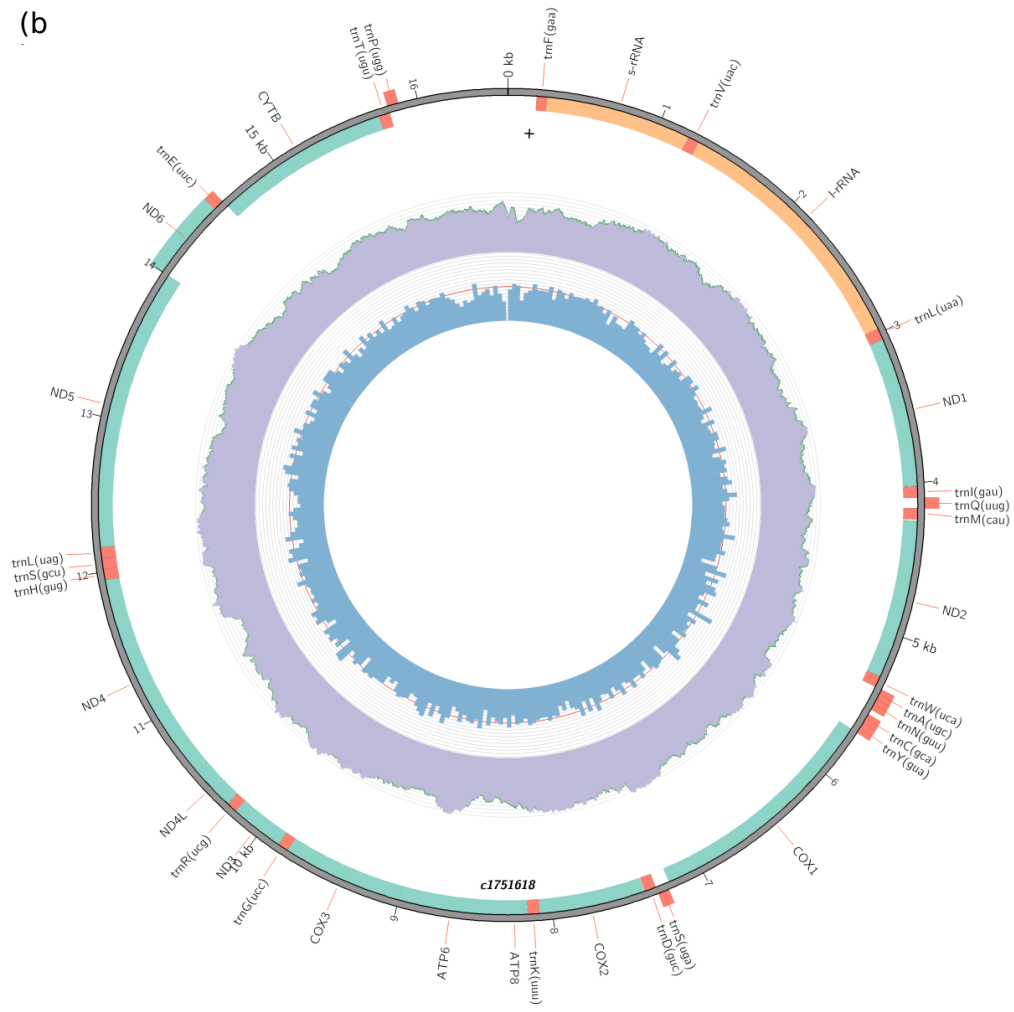
P2-FOR /5Phos/GATCGGAAGAGCGGTTTCAGCAGGAATGCCGAGACCGATCAGAACA*A
CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC
P2-REV *T

FOR-PCR AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
REV-PCR CAAGCAGAAGACGGCATAACG*A

(a)



(b)



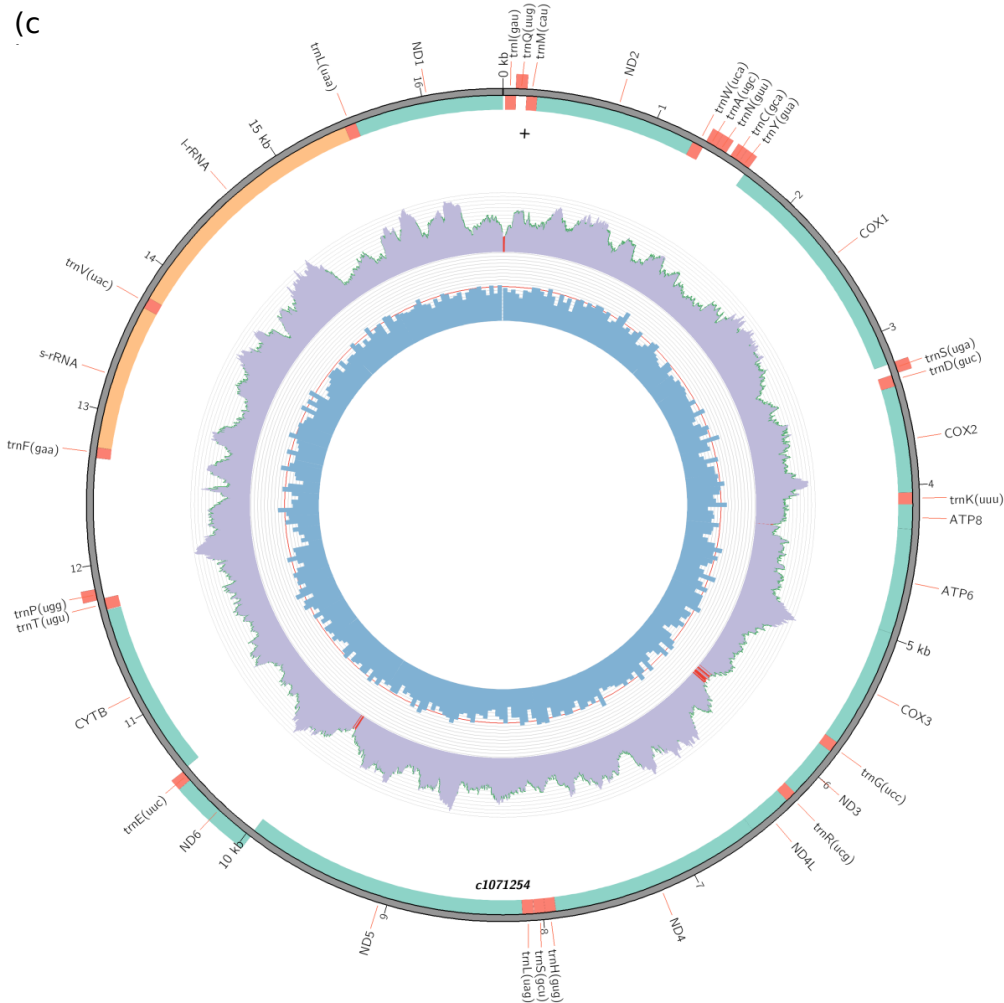


Figure S1: Visualization of circular mtDNA for (a) *D. sargus* (16,620 bp), (b) *M. surmuletus* (16,577 bp) and (c) *S. cabrilla* (16,513 bp). The innermost circle shows the GC content, which is calculated with a sliding-window method. The middle circle shows the coverage depth distribution with the dark green outline. Parts of the genome that have a coverage lower than 20X are shaded in red, whereas parts of genome that have coverage larger than the upper quartile are shaded in dark green. The outer circle shows the gene annotations and is shaded in orange for rRNA, in red for tRNA, and in blue for CDS.

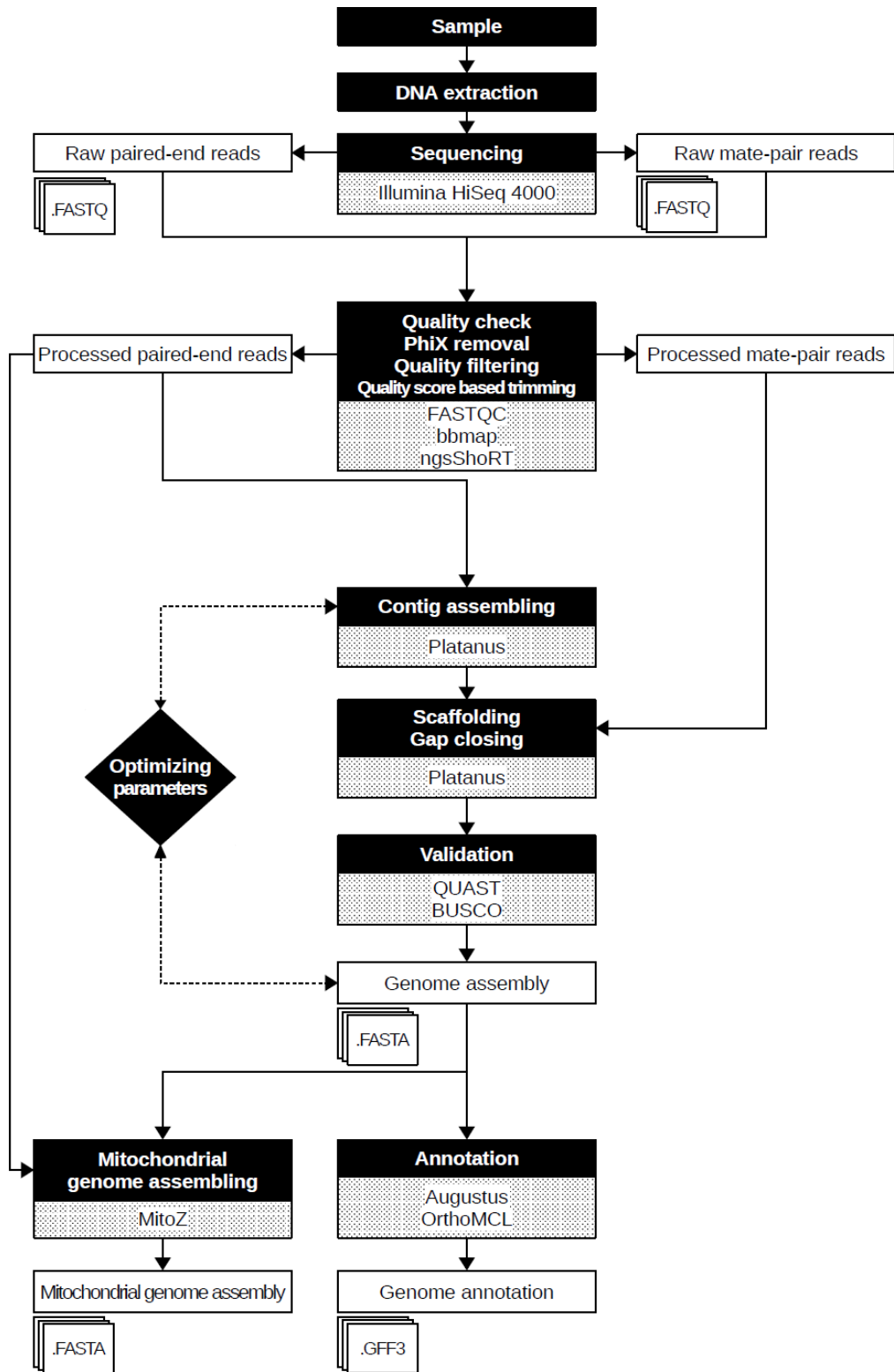


Figure S2: Schematic workflow of DNA extraction, genome assembly and annotation.

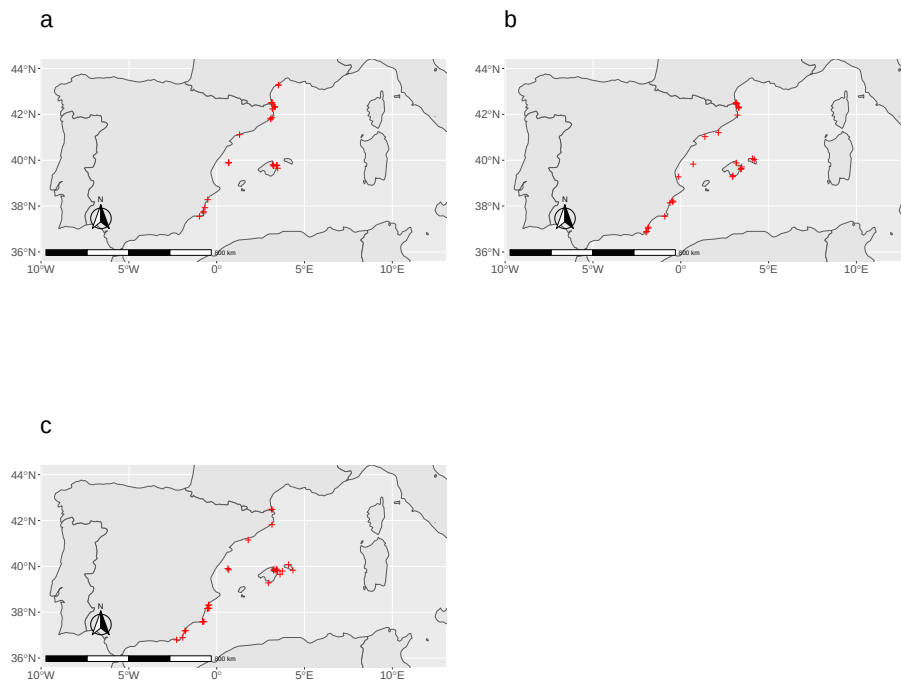


Figure S3. Map of the sampling sites. a) *Diplodus sargus* b) *Mullus surmuletus* c) *Serranus cabrilla*

References

- AlMomin, S., Kumar, V., Al-Amad, S., Al-Hussaini, M., Dashti, T., Al-Enezi, K., & Akbar, A. (2016). Draft genome sequence of the silver pomfret fish, *Pampus argenteus*. *Genome*, *59*(1), 51-58. doi:10.1139/gen-2015-0056
- Domingos, J. A., Zenger, K. R., & Jerry, D. R. (2015). Whole-genome shotgun sequence assembly enables rapid gene characterization in the tropical fish barramundi, *Lates calcarifer*. *Animal Genetics*, *46*(4), 468-469. doi:10.1111/age.12312
- Nakamura, Y., Mori, K., Saitoh, K., Oshima, K., Mekuchi, M., Sugaya, T., . . . Inouye, K. (2013). Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(27), 11061-11066. doi:10.1073/pnas.1302051110
- Shin, S. C., Ahn, D. H., Kim, S. J., Pyo, C. W., Lee, H., Kim, M. K., . . . Park, H. (2014). The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biology*, *15*(9). doi:10.1186/s13059-014-0468-1
- Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S. T., . . . Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*. doi:10.1038/ncomms6770
- Xu, T. J., Xu, G. L., Che, R. B., Wang, R. X., Wang, Y. J., Li, J. R., . . . Yang, Q. (2016). The genome of the miiuy croaker reveals well-developed innate immune and sensory systems. *Scientific Reports*, *6*. doi:10.1038/srep21902