



# Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish

Israel Guerrero-Cózar, Jessica Gomez-Garrido, Concha Berbel, Juan Martinez-Blanch, Tyler Alioto, M. Gonzalo Claros, Pierre-Alexandre Gagnaire, Manuel Manchado

## ► To cite this version:

Israel Guerrero-Cózar, Jessica Gomez-Garrido, Concha Berbel, Juan Martinez-Blanch, Tyler Alioto, et al.. Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish. *Scientific Reports*, 2021, 11 (1), 10.1038/s41598-021-92601-5 . hal-03443762

**HAL Id: hal-03443762**

**<https://hal.science/hal-03443762>**

Submitted on 23 Nov 2021

**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.

# **Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish**

Israel Guerrero-Cózar<sup>1</sup>; Jessica Gomez-Garrido<sup>2</sup>; Concha Berbel<sup>1</sup>; Juan F. Martinez-Blanch<sup>3</sup>; Tyler Alioto<sup>2,4</sup>; M. Gonzalo Claros<sup>5,6,7,8</sup>; Pierre-Alexandre Gagnaire<sup>9</sup>; Manuel Manchado<sup>1,10\*</sup>

<sup>1</sup>IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de Santa María, Cádiz, Spain

<sup>2</sup>CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, 08028, Spain.

<sup>3</sup>LifeSequencing, Parc Científic Universitat De Valencia, Edif. 2, C/ Catedrático Agustín Escardino Benlloch, 9, 46980 Paterna, Spain

<sup>4</sup>Universitat Pompeu Fabra (UPF). 08003 Barcelona, Spain

<sup>5</sup>Universidad de Málaga, Department of Molecular Biology and Biochemistry, Málaga, E-29071, Spain

<sup>6</sup>CIBER de Enfermedades Raras (CIBERER), Málaga, E-29071, Spain

<sup>7</sup>Institute of Biomedical Research in Málaga (IBIMA), IBIMA-RARE, Málaga, E-29010, Spain

<sup>8</sup>Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM-UMA-CSIC), Málaga, E-29010, Spain.

<sup>9</sup>ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

<sup>10</sup>“Crecimiento Azul”, Centro IFAPA El Toruño, Unidad Asociada al CSIC”

\*Corresponding author:

Manuel Manchado. IFAPA Centro *El Toruño*. Camino Tiro de Pichón s/n. 11500 El Puerto de Santa María (Cádiz), Spain. Tel: +34 671532088. Fax: +34 856102033. Email: [manuel.manchado@juntadeandalucia.es](mailto:manuel.manchado@juntadeandalucia.es)

## ABSTRACT

The integration of physical and high-density genetic maps is a very useful approach to achieve chromosome-level genome assemblies. Here, the genome of a male Senegalese sole (*Solea senegalensis*) was *de novo* assembled and the contigs were anchored to a high-quality genetic map for chromosome-level scaffolding. Hybrid assembled genome was 609.3 Mb long and contained 3,403 contigs with a N50 of 513 kb. The linkage map was constructed using 16,287 informative SNPs derived from ddRAD sequencing in 327 sole individuals from five families. Markers were assigned to 21 linkage groups with an average number of 21.9 markers per megabase. The anchoring of the physical to the genetic map positioned 1,563 contigs into 21 pseudo-chromosomes covering 548.6 Mb. Comparison of genetic and physical distances indicated that the average genome-wide recombination rate was 0.23 cM/Mb and the female-to-male ratio 1.49 (female map length: 2,698.4 cM, male: 2,036.6 cM). Genomic recombination landscapes were different between sexes with crossovers mainly concentrated toward the telomeres in males while they were more uniformly distributed in females. A GWAS analysis using seven families identified 30 significant sex-associated SNP markers located in linkage group 18. The follicle-stimulating hormone receptor appeared as the most promising locus associated with sex within a region with very low recombination rates. An incomplete penetrance of sex markers with males as the heterogametic sex was determined. An interspecific comparison with other Pleuronectiformes genomes identified a high sequence similarity between homologous chromosomes, and several chromosomal rearrangements including a lineage-specific Robertsonian fusion in *S. senegalensis*.

**Keywords:** Senegalese sole, linkage mapping, genome assembly, chromosome evolution, SNPs.

## INTRODUCTION

Genetic maps represent essential tools for genomic research in aquaculture. Originally, linkage mapping studies were mainly based on microsatellite (SSR) and AFLP markers<sup>1,2</sup>; nevertheless, they recently reached a milestone with the development of genotyping methods based on cost-effective massive parallel sequencing. The genomic revolution has made single-nucleotide polymorphisms (SNPs) very popular, opening up access to a simple biallelic marker with a wide distribution and high abundance across the genome. As consequence, an increasing number of high-density genetic maps is nowadays reported in non-model organisms including aquaculture fish<sup>3,4</sup>. These maps have proven to be useful to provide new clues on genome evolution and speciation between closely related lineages, and to unravel the genetic architecture of both simple Mendelian and complex quantitative traits in many fish species, thus facilitating marker-assisted selection in aquaculture<sup>5,6</sup>. More recently, a new application of high-density linkage maps as backbones to anchor *de novo* genome assemblies into pseudo-chromosomes has become more widespread<sup>7,8</sup>. Although long-read sequences have significantly enhanced the average size of scaffolds in *de novo* assembled genomes<sup>9</sup>, the total number of scaffolds are still far beyond the expected number of chromosomes. The large arrays of repeated sequences and the degree of conservation for some tandem repeats families widely distributed across the genome still remain a major obstacle for most *de novo* assembly algorithms, resulting in fragmented scaffolds or even misassembled sequences within chimeric contigs. Linkage maps thus provide highly valuable tools to anchor physical maps into pseudo-chromosomes, while enabling the identification of chimeric or misassembled contigs towards enhancing the quality of new genome assemblies<sup>7</sup>.

Flatfish (Pleuronectiformes) is an attractive group of fish that have long been investigated due to the drastic morphological, physiological and behavioural remodelling changes that occur during metamorphosis from a pelagic larva to a benthic juvenile stage. Several flatfish species are worldwide exploited in fisheries and aquaculture, thus representing an important resource for human consumption. This taxonomic group diverged from carangimorphs in the early Paleocene, and underwent a major diversification in the middle Paleocene<sup>10</sup>. Cytogenetic studies have suggested that the Pleuronectiformes ancestor should have  $2n = 48$  chromosomes in agreement with the most frequent number of chromosomes found in the sister clade Carangidae, and in the most deep-branching flatfish families (Pleuronectidae and Paralichthyidae)<sup>11</sup>. However, the number of chromosomes in flatfish encompasses a wide range varying from  $2n = 26$  to  $2n = 50$ <sup>11,12</sup>. An intense cascade of Robertsonian rearrangements and pericentromeric inversions seems to have shaped flatfish genome evolution, especially reducing the chromosome number in most recently diverged families of Soleidae, Cynoglossidae and Achiridae<sup>11</sup>. A recent comparison of the turbot genome with other fish assemblies clearly pointed out the high degree of conserved synteny across chromosomes in Pleuronectiformes, although with high rates of intrachromosomal reorganisations. Moreover, some chromosome fusions identified through comparative mapping are thought to have given arise to a new karyotype organization in turbot<sup>3</sup>. Hence, integrated genetic and physical maps are important genomic resources to understand chromosome evolution in flatfish.

The Senegalese sole is an important flatfish in aquaculture and fisheries. A genetic linkage map based on 129 SSRs grouped into 27 linkage groups (LG) was previously reported<sup>13</sup>. Moreover, an integrated map using BAC clones and repetitive DNA families was also developed using a multiple fluorescence in situ hybridization (mFISH) technique with at least one BAC mapped to each chromosome arm<sup>14</sup>. This cytogenetic study

evidenced a lack of heteromorphic sex chromosomes and identified the largest metacentric chromosome to result from a Robertsonian fusion of two acrocentric chromosomes during flatfish evolution<sup>15,16</sup>. Moreover, a preliminary draft genome sequence of a female Senegalese sole was reported (600.3 Mb, N50 of 85 kb), and then further improved with a hybrid assembly using Nanopore and Illumina reads (608 Mb long, N50 of 340 kb)<sup>17,18</sup>. This genome information was used to design whole-genome multiplex PCR and create a new integrated SSR map with 234 markers. Nevertheless, further efforts are required to better assemble and anchor scaffolds onto the 21 expected chromosomes, and to better understand the genomic architecture of sex-determination.

The aim of this study was to: 1) generate an improved *de novo* assembly of a male Senegalese sole based on a combination of long and short read sequencing; 2) build a high-density genetic map using ddRAD markers; 3) anchor the physical to the genetic map in order to 4) improve the scaffolding of the reference genome assembly; 5) estimate genome-wide variation in recombination rates; and 6) carry out GWAS analysis to identify sex-associated markers and intra- and interspecific comparative mapping to better understand the evolutionary history of chromosome rearrangements in flatfish.

## **2. Material and methods**

### **2.1 Animals**

Soles used for the preparation of ddRAD libraries and sequencing were selected from the genetic breeding program carried out by the IFAPA in collaboration with a commercial aquaculture company (CUPIMAR S.A.). Production of families used in this study, genotyping and parentage assignment were previously published<sup>19,20</sup>. Five families (three full-sib and two maternal half-sib families) containing between 48 and 96 individuals per

family (total n = 356) were selected to construct the genetic linkage map (Table 1). Moreover, seven families with sex ratios close to 1:1 were selected for genome-wide association analysis (GWAS). Average weight and length of each family are depicted in Table 1. As genotyping of parents was also required to build the genetic map, five fathers and three mothers involved in family production were sampled for blood by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at -20°C until use. To obtain high-molecular weight genomic DNA for genome sequencing, a wild male from the broodstock (weight higher than 2 kg; code Sse05\_10M) was sampled for blood as indicated above.

All procedures were authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101 by the National authorities for regulation of animal care and experimentation. The study was carried out in compliance with the ARRIVE guidelines and all procedures were performed in accordance with Spanish national (RD 53/2013) and European Union legislation for animal care and experimentation (Directive 86\609\EU).

## **2.2 Genome sequencing and assembly**

Methods for genome sequencing and assembly are fully described in Supplementary method. Briefly, high-molecular weight genomic DNA was prepared from heparinized whole blood using the MagAttract HMW DNA kit (Qiagen). Once confirmed quality, four libraries were prepared for sequencing using the Oxford nanopore Technology (ONT) MinION platform. Overall, 19.2 Gb of genome information was generated with an average read length of 4.3 kb. In parallel, the same sample was also sequenced in a NextSeq550 sequencer (Illumina, USA) that overall generated 43 Gb of sequence from 143 million reads (average length 147 nt). The main features of the libraries used during



the genome assembly are presented in Supplementary Table S1. The raw read data were deposited to the NCBI Sequence Read Archive (SRA) under accession number SAMN16809702. The hybrid genome assembly was carried out using MaSuRCAv3.2.3<sup>21,22</sup> with the Illumina libraries (57.3x coverage) and the error-corrected Nanopore reads (25.5x). The LR-hybrid assembly was characterized for completeness using Benchmarking Universal Single-Copy Orthologs (BUSCOv3.0.2)<sup>23,24</sup> containing 4,854 single-copy orthologs from *actinopterygii\_odb9*.

### **2.3 ddRAD-seq library preparation and sequencing**

Genomic DNA from the caudal fin (offspring) or whole blood (parents) were purified using the Isolate II Genomic DNA Kit (Bioline). DNA was sent to the company LifeSequencing S.L. (Valencia, Spain) and a total of 346 samples were selected for library construction (Table 1). Libraries were constructed based on the protocol described by Peterson et al.<sup>25</sup> using the EcoRI/NcoI enzyme combination that generated as average 24,874 SNPs per sample. Pools of libraries were loaded on a Novaseq 6000 sequencer (Illumina), following the manufacturer's instructions and the specifications mentioned above. The total number of reads generated for each library are indicated in Supplementary Table S2.

### **2.4 Genetic linkage map and scaffold anchoring**

Illumina reads were processed using Stacks v2.3e<sup>26</sup> as indicated in Supplementary methods. To construct the map, SNPs were filtered using Plink v1.9<sup>27</sup> to remove markers that segregated with Mendelian errors in more than 10% of individuals. Moreover, those individuals with more than 5% of markers with Mendelian errors were removed (Supplementary Fig. S1). The final SNP dataset contained 40,041 markers from 327 individuals (Table 1) and 8 parents that were imported in LepMap3<sup>7</sup>. The SNPs were

assigned to 21 linkage groups (named as SseLGs) corresponding to the expected number of chromosomes ( $2n = 42$ ) using the "SeparateChromosomes" module. A LOD threshold of 11 and a size limit of 200 were selected as the most adequate parameters to keep an optimal number of markers grouped in the expected number of SseLGs (Fig. 1A and B). Module JoinSingles2 was run to assign additional single SNPs to existing SseLG using decreasing LOD score iterations from 10 to 5 (Fig. 1B). Finally, the genetic distances between markers on each SseLG was calculated with the OrderMarkers2 module (male, female, sex average (SA)) using the Kosambi mapping function. The resulting genetic map was visualized using the software linkagemapview<sup>28</sup>. Scaffolds anchoring was carried out using the Lep-Anchor program following the author's recommendation<sup>29</sup> and indicated in Supplementary methods.

## **2.5 Genome annotation**

Genome annotation was performed by combining alignments of *Danio rerio*, *S. maximus* and *S. semilaevis* proteins, RNAseq from several tissues and developmental stages alignments and ab initio gene predictions. Annotation process is described in Supplementary methods with a higher detail. Functional annotation was performed on the male annotated proteins with Blast2GO<sup>30</sup>. After performing an alignment-based strategy to determine equivalences between female and male genomes (see Supplementary methods), the female proteins inherited the functional annotation of their male equivalences. Next, functional annotation was performed in the female genes that remained unannotated after this step. Gene Ontology (GO) enrichment was carried out with topGO in those genes that were unique to one of the genomes (Supplementary Table S3).

## **2.6 Recombination rates, association analyses and cross-species comparisons**

Recombination rate variation along the genome was evaluated by comparing the consensus linkage map for both sexes and SA and the physical map of each pseudo-chromosome using MareyMap<sup>31</sup>. The cumulative recombination frequency (RFm) along LGs was used to infer the chromosome type as previously described<sup>32</sup>. GWAS analysis were carried out with seven families (Table 1) using a logistic mixed model (multi-step) approach as implemented in the R package GENABEL (v1.8-0)<sup>33</sup> for binary traits (Female= 0 and Male= 1). A highly detailed analysis of synteny across flatfish is beyond the scope of this study, but a chromosome alignment analysis was carried out to identify chromosomal rearrangements in flatfish using D-Genies<sup>34</sup>. We then used the SatsumaSynteny to compute whole-genome synteny blocks<sup>35</sup> that were later represented using Shinycircos<sup>36</sup>.

### **3. Results**

#### **3.1 Male genome assembly and annotation**

A *de novo* hybrid genome for a male sole was assembled using a combination of Illumina and Nanopore long-reads. Main features about the total number of input reads used for each sequencing platform, the average read length and quality and total sequencing information used in the assembly are indicated in Supplementary Table S1. The hybrid assembly draft sequence was generated using MaSuRCA and later refined with Pilon to correct bases, mis-assemblies and filling gaps. Main statistics about the assembly are depicted in Supplementary Table S4. The new assembly consists of 3,403 contigs with a total length of 609,359,514 bp, and a N50 of 513 kb. Overall, 49.4% of contigs had a size longer than 50 kb and the largest fragment was 4.5 Mb long. The estimated gene integrity, as determined by BUSCO analysis, revealed 97.0% completeness. For comparison

purposes, the assembly statistics for a recent female genome draft of *S. senegalensis*<sup>20,37</sup> are also shown (Supplementary Table S4). Both genome assemblies had a similar size (608-610 Mb) although the newly assembled male genome had longer contigs with higher N50 values. A dot-plot alignment using the scaffolds of both genomes indicated that with 92.8% of genomic information highly similar (>75%) and only 5.3% had no similarity (average similarity 94%) (Fig. 2).

Assembly annotation statistics are depicted in Table 2. The number of protein-coding genes in the male assembly (27,175) was slightly lower than in the female (28,988) but with a longer mean length (7.4 vs 6.7 kb). The estimated percentages of annotated transcripts (69.4-72.1%) and gene density (45.03-47.68) were similar between both assemblies. Around 85% of the annotated genes in each assembly had an equivalent gene in the other assembly. However, a few genes were only present in one of the genomes (unique genes). Some of these might be due to genome heterozygosity and repeat content or even sex-specific genes. A GO enrichment analysis using these unique genes indicated that categories related to the cell-cycle regulation and regulation of transcription, involving canonical histones H3.2 and H4 and retinoid X receptor alpha (*rxra*), were highly significantly overrepresented in the female (p-value <10<sup>-3</sup>). Mapping of these two histone genes on female assembly showed that they were co-localized in five scaffolds (Sosen1\_s0284, Sosen1\_s0324, Sosen1\_s1454, Sosen1\_s1522, Sosen1\_s1726), four of which clustered in SseLG1 and one in SseLG16. In male, the most significant enriched categories for unique genes were skeletal system development and morphogenesis although with *P*-values >0.001 (Supplementary Table S3). Some short, single-exonic unique genes might be the result of scaffold splitting or annotation processes. The non-coding gene annotation resulted in 23,822 female and 21,123 male

transcripts, respectively. From these, 6,549 and 6,007 female and male transcripts were long non-coding RNAs (lncRNAs) and the rest short non-coding RNAs.

### **3.2. ddRAD sequencing and SNP detection for genetic linkage map**

Three full-sib and two half-sib families consisting of 47 to 95 individuals were used for ddRAD analysis (Table 1). The total number of paired-end reads generated for each family ranged between 280,609,738 (F5) and 398,313,256 (F2) with an average length of 150 nt (Table 3). The average number of reads per individual in each family varied between 6,444,752 (F1) and 11,692,072 (F5) (Table 3 and Supplementary Table S2). For parents, the average number of reads was 8,847,913.

The new assembled male genome was used as reference to map the ddRAD reads. The average fraction of primary alignments onto this reference genome ranged between 88.04 (F6) and 89.71% (F2). An average of 10.5% of reads had insufficient mapping qualities or excessively soft-clipped primary alignments while less than 0.34% were unmapped. A total of 199,188 ddRAD *loci* were reconstructed with an average number of loci per sample ranging between 23,828 (F1) and 30,550 (F7) and a mean insert length of 330.7 bp. The effective coverage per sample was  $193.3 \pm 110.4$  (ranging from 146 to 242 between families) and the estimated mean number of sites per locus was 242.8 (Table 3).

### **3.3. Construction of a linkage genetic map and anchoring to physical map**

To construct the genetic map, only those SNPs detectable in at least 80% of samples with a coverage of 10 reads per sample were considered. Moreover, SNPs with a significant deviation from Mendelian segregation were also removed (a total of 2,439 markers, 5.7% SNPs). By family, the number of markers with Mendelian errors ranged from 1.5 to 1.7% (Supplementary Fig. S1). Moreover, those animals with markers that had more than 5%

of Mendelian errors (19 specimens) were also removed. Overall, the final dataset contained 40,041 SNPs segregating in eight parents and their 327 offspring.

For linkage analysis, the ParentCall2 module retained only 16,287 informative markers after checking for segregation distortion ( $P < 0.05$ ). Markers grouped into 21 SseLGs (via the SeparateChromosomes2 module) with a LOD = 11 (Fig. 1), which is consistent with the number of chromosomes in *S. senegalensis*. Each SseLG contained between 530 and 1,337 markers with an average number of 21.9 markers per Mb (Fig. 3, Table 4 "Anchoring genetic map and physical map"). In total, the genetic map allowed the anchoring and positioning of 1,665 out of 3,403 total contigs, ranging between 50 to 129 contigs in each SseLG. The genome sequence positioned on the linkage map was larger (746.3 bp) than the assembly size, mainly due to the presence of chimeric contigs (n = 133) positioned in various chromosomes.

#### **3.4. Rescaffolding of reference genome with the genetic map**

SNP marker information was further used for fine-scale correction of genome contigs to build 21 pseudo-chromosomes. After masking the repetitive sequences, the contigs were orientated and sorted within each SseLG (Table 4 "Genome re-scaffolding"). The total number of positioned contigs reduced from 1,665 to 1,563. Lep-anchor corrected the contig errors removing six contigs, splitting another 105 into two fragments, 20 in three fragments, and two in more than four fragments. After these corrections, the total number of markers assigned to the SseLGs decreased by 1.3% (16,075 SNPs) and 212 markers were moved to unplaced with an average density of 10.3 markers per contig. After these corrections, 548.6 Mb out of the 610.4 Mb total assembly length (89.9%) were assigned to the 21 SseLGs and only 61.9 Mb remained as unanchored (Table 4). The total map length was 2,408.1 cM, SseLG1 was the largest group (42,924,012 bp and 147.3 cM) and

SseLG4 showed the highest marker density per megabase (33.1). The average marker interval reached 0.155 cM. A further refining of anchored markers was carried out through the comparison of physical and genetic distance in MareyMap. The average genome-wide recombination rate (RR) was 4.35 cM/Mb (ranging between 3.45 and 5.26 cM/Mb among chromosomes) (Table 4 "Marker refining"). An alignment of the anchored and refined reference male genome with the scaffolds of the female assembly (Fig. 2B) slightly increased to 93.2% the regions with more than 75% similarity and provided a clear sequence alignment in the diagonal with only dispersion in unplaced scaffolds.

### **3.5 Analysis of recombination rates**

Consensus genetic maps for female and male were 2,698.4 cM (15,022 markers) and 2,036.6 cM (15,390 markers), respectively. These differences in map size were observable for the 21 SseLGs (Fig. 4A and Table 5). Overall, the female-to-male ratio (F:M) for genetic distances was 1.32, ranging from 1.08 (SseLG15) to 1.77 (SseLG5) (Table 5). The genetic map length of chromosomes was highly positively correlated with their physical length in both males ( $r = 0.43$ ) and females ( $r = 0.60$ ) (Fig. 4B). The average genome-wide RR was estimated  $3.02 \pm 0.37$  cM/Mb in males and  $4.51 \pm 0.57$  cM/Mb in females (Table 5). The overall female-to-male ratio (F: M) for RR was 1.49, ranging from 1.43 to 1.90 across chromosomes. In the case of males, SseLG12 showed the lowest (2.47 cM/Mb) and SseLG16 the highest (3.60) mean RR values. In females, SseLG4 had the lowest (3.57 cM/Mb) and SseLG5 the highest (5.65 cM/Mb) mean RR values.

The local RR value as estimated by the relative distance to the nearest telomere was clearly different between males and females. High RR values were mainly concentrated close to the telomeres in males (Fig. 5A), while they were more uniformly distributed in females with higher RR being found around 15% of the distance to the nearest telomere

(Fig. 5B). This was illustrated by contrasted chromosomal RR landscapes between males and females, as shown Fig. 5 C&D for SseLG1 (landscape for all SseLGs are represented in the Supplementary Fig. S2 for males and Supplementary Fig. S3 for females). We detected some regions within SseLGs (i.e 5, 11, 13, 14, 15, 18) with very low RR. In the case of SseLG18, partially restricted male or female RR was detected in the region 9.5-10.9 Mb. This region had very low RR in males (1.2) and females (0.6) compared with average SseLG18 (3.0 and 4.9 RR, respectively). Cumulative RR crossed between both sexes around chromosomal position 10 Mb with female RR closed to zero in 10.8-10.9 Mb (Fig. 6, Supplementary Fig. S2 and S3). Moreover, recombination frequencies were used to describe and classify chromosome morphologies. Fig. 7 depicts the typical RFm plots for an acrocentric (SseLG20) and a metacentric (SseLG1) chromosome (for all SseLG see Supplementary Fig. S4).

### **3.6 Association analyses for sex**

To identify genome regions associated with sex, a GWAS analysis was carried using seven families (Table 1) and a total of 10 426 markers. Data for RAD-seq data and markers are indicated in Table 3. The results showed 30 markers significantly associated with sex after bonferroni correction using seven families ( $P \leq 4.8 \times 10^{-6}$ ; Fig. 6A and Supplementary Table S5). When the association analysis was repeated separately by family, five families provided some new 36 significant markers (Supplementary Table S5). All of them (66 SNPs including the whole-population and families) were spread in the SseLG18 with a hot region around 9.5-10.9 Mb (Fig. 6B). RR in this region was low (see above) with partially restricted RR associated with sex. Overall, 80.7% of significant markers using the whole population were preferentially heterozygous in males although penetrance was incomplete in most of them. This model is compatible with a nascent XY



system. It should be noted that specific markers in family 4 had an expected high number of heterozygous *loci* in females.

To detect candidate sex-related genes, the full-length transcriptome<sup>38</sup> was blasted onto the SseLG18 (Supplementary Fig. S5) and a total of 229 genes were positioned. The significant SNPs were highly distributed through the pseudo-chromosome, but the follicle stimulating hormone receptor (*fshr*) gene just appeared located in the hot region revealing as a clear candidate gene for sex determination.

### 3.7 Interspecific chromosome rearrangements

An alignment of SseLGs pseudo-chromosomes with the chromosomes of three other Pleuronectiformes genomes (*Cynoglossus semilaevis*, *Scophthalmus maximus*, *Paralichthys olivaceus*) showed high similarity rates of and conserved macrosynteny level for fifteen out of 21 SseLGs (Fig. 8 and Supplementary Table S6). However, deviations from diagonal in the dot plot alignment indicated extensive intrachromosomal rearrangements among species. The three largest SseLGs appeared to be the result of total or partial chromosome fusions when compared with other flatfish genomes (Supplementary Fig. S6 and S7), and *S. maximus* seemed to be the flatfish species with the highest number of chromosome rearrangements between the four species compared. Genome comparisons using D-Genies<sup>34</sup> indicated that the highest similarity was with *P. olivaceus* (no match 57.3%), followed by *S. maximus* (no match 59.6%), and *C. semilaevis* (no match 78.4%).

When the reduction of the number of chromosomes was explored three main Robertsonian fusions in the SseLG1 (Chr18-Chr11), SseLG2 (Chr14-Chr15) and SseLG3 (Chr9-Chr16) could explain the reduction from  $n = 24$  in *P. olivaceus* to  $n = 21$  in *S.*

*senegalensis* (Fig. 7, Supplementary Fig. S6 and S7 and Supplementary Table S6). When compared to *S. maximus* (n = 22), the SseLG1 appeared as a fusion of Chr7 and Chr21. Moreover, translocations of regions from Chr1, Chr4, Chr7, Chr14 and Chr16 were also observed. In the case of *C. semilaevis* with sexual chromosomes (ZW) and the same number of chromosome than *S. senegalensis*, a Robertsonian fusion in SseLG1 between Chr3-Chr20 was observed. Moreover, the SseLG3 appeared as a new chromosome resulting of the fission of Chr1 (mainly located in SseLG16) and Chr8 (mainly located in SseLG18). Two other major features in this species with respect to *S. senegalensis* were: i) a translocation of a Chr14 region to Chr16 to create the SseLG2; and ii) sexual ZW chromosomes appear concentrated in SseLG5 although high similar sequences are widely distributed throughout the genome. Comparison among all flatfish species (Fig. 7, Supplementary Fig. S6 and S7, Supplementary Table S6) indicated that those chromosomal regions associated with SseLG2 and SseLG3 were mainly involved in the changes of karyotypes of the four Pleuronectiformes species whereas the SseLG1 arose as a lineage-specific fusion event.

## Discussion

Genome assemblies and genetic linkage maps provide complementary information that can be integrated to produce high-quality physical maps. The resulting accurate chromosome assemblies are suitable to investigate genome evolution and species diversification, the genetic architecture of QTLs and the regulation of targeted genome regions. In this study, a *de novo* hybrid assembly for a male sole and a high-density SNP map were generated and combined to provide a polished draft assembly of 21 pseudo-chromosomes. A genome for a female sole was previously reported <sup>17</sup> although it was highly fragmented (N50=85 kb, 600.3 MB long). Later, this assembly was improved by

integrating Nanopore and Illumina reads, resulting in 5,748 contigs with N50 = 339.9 kb and 608 Mb long<sup>20</sup> (Supplementary Table S4). In this study, the newly obtained male assembly has a lower number of contigs (3,403) and higher N50 (512.7 kb) and confirmed that the genome size of sole is around 609 Mb. This genome size is similar or even a bit larger than other flatfish<sup>39-42</sup>. A dot-plot alignment analysis indicated a high similarity between male and female genome assemblies perfectly aligned along the diagonal (Fig. 2) with a completeness similar to other high-quality fish assemblies (>95.5% complete genes)<sup>40,43,44</sup>.

Male genome characterization identified 50,133 transcripts and 27,175 protein-coding that agrees with the number of predicted transcripts in a recently assembled informative transcriptome<sup>38</sup>. Moreover, a small subset of unique genes was identified in both sexes with a high overrepresentation of cell-cycle regulation and regulation of transcription categories (including mainly the histones H3.2 and H4) in the female. In mammals, unique histone variants are specifically expressed in spermatogenic cells<sup>45</sup>. Moreover, expansion of histone multigene clusters in scleractinians was associated with sexually dimorphic expression of some variants playing a role in the control of gene expression in female and male germ cells during gametogenesis<sup>46</sup>. In sole, at least two *loci* of canonical histones in the largest metacentric chromosome SseLG1 linked to *dmrt1*, a key determination gene in other flatfish, were reported in sole<sup>16,39,47</sup>. This chromosome arose after a Robertsonian fusion and intense reorganization events<sup>12</sup> that could have birth to new histone clusters under purifying selection<sup>48</sup>. Although we cannot exclude that some differences in the number of histone copies between both genomes could be attributed to individual variation, one plausible hypothesis is that some of these histone clusters could have subfunctionalized and acquired a role in gametogenesis in a sex-specific manner. This hypothesis is supported by the identification of a *rxra*-like receptor also represented

in such GO categories able to mediate the masculinizing effects of females mediated by its ligand TBT in rockfish females <sup>49</sup>.

*De novo* assembled male genome was used as reference to map the ddRAD sequences and construct a high-density genetic map. The sole consensus map size and the number of high-quality markers used (Fig. 3; Table 4) were similar to those reported for turbot (2,622.09 cM)<sup>6</sup> and flounder (3,497.29 cM)<sup>50</sup> although with a higher density of markers (only 6,647 and 12,712 SNPs in turbot and flounder, respectively). Most importantly, markers were distributed into 21 SseLGs that match with the haploid karyotype (2n = 42) of the species<sup>51</sup>. Until now, two genetic maps with 129-229 microsatellites were reported in Senegalese sole <sup>13,20</sup> Moreover, a cytogenetic map was also published although the number of BACs did not still cover all chromosomes<sup>14,16</sup>. This new high-density SNP map (Fig. 3) thus represents a key step forward for future genomic studies and QTL identification with respect the current information available until now in this species.

Although hybrid assemblies using long and short sequences reads reduce genome fragmentation and increase the average scaffold sizes as observed in this study, most of *de novo* genome assemblies still do not reach chromosome-level with the expected number of chromosomes due to, among other factors, the repetitive fraction of the genome. To get around this limitation, information of genome-wide physical maps and dense genetic linkage maps can be integrated to assign chromosomal locations to sequence contigs<sup>52</sup>. This anchoring can also remove assembly artifacts and position misplaced scaffolds to increase the contiguity of the assembled scaffolds. In this study, the high-density SNP genetic map was used to anchor, sort and refine the assembled contigs. Overall, 89.9% of the genome assembly could be anchored to 21 pseudo-chromosomes and a total of 102 contigs were removed or split to separate positions in SseLGs. A similar strategy was followed in turbot using 31 families that allowed for the

rearrangement of 20% of the genome assembly<sup>3</sup>. A comparison between male and female demonstrated a high co-linearity between our physical map and female scaffolds (only 5.53% mismatch). Although 10.1% of genome information remained as unplaced, the anchored physical map is essential for gene association analysis, synteny and cross-species studies and targeted genome resequencing. Further studies will be required to accurately anchor the remaining 61.9 Mb unanchored regions to their position in the genome.

It is well-known that the genome-wide RR differs between males and females (heterochiasmy) and that the recombination landscape also varies along chromosomes. In animals and plants, females tend to have higher RR than males, which in turn result in larger map lengths<sup>53-55</sup>. In our study, map was longer in the female than in the male (2,698.4 vs 2,036.6 cM; ratio 1.32). Assessment of sex-specific RR indicated a female-biased heterochiasmy across all SseLGs, with an average RR of 3.02 in male vs 4.51 cM/Mb in female. Four species of Pleuronectidae also exhibited wide heterochiasmy through all chromosomes similarly to sole with some intervals of male- and female-restricted meiotic recombination<sup>56</sup>. However, such differences in RR between males and females are not fully conserved in flatfish when map size is considered. Female maps are larger in turbot (1.36 times) and halibut (1.07 times)<sup>1,2,57</sup>, this is not the case of flounder or tongue sole with slightly larger maps in males (1.03-1.09 times)<sup>50,58,59</sup>. *C. semilaevis* is the only flatfish known with heteromorphic sex chromosomes (ZZ/ZW) that has been described in several mammals, birds and insects as a cause for an arrest of recombination in the heterogametic sex (XY males or ZW females). This could explain a shift in the direction of heterochiasmy<sup>53</sup>.

In addition to such differences in overall RR between sexes, the chromosomal recombination landscapes also differed between male and female according to typical patterns. In fish, it has been shown that recombination occurs at higher frequencies near telomeres in males while the distribution is quite more uniform or elevated near centromeres in females<sup>54</sup>. In stickleback fish, it has been demonstrated that centromeres and telomeres have little or no effect on recombination in females, however, in males, the recombination rates are suppressed near the centromeres and hence crossovers localize mainly at the ends of long arms in acrocentric chromosomes<sup>55</sup>. This feature seems to be conserved in sole since RR were also more frequent toward the end of males SseLGs compared to females (Fig. 5).

Heterochiasmy is considered a major force that guides the evolution of genetic sex determination systems and speciation<sup>56,60</sup>. Normally, genome regions with very low RR are associated with sex-determining regions in young sex chromosome systems and sex-linked traits such as pigmentation<sup>61</sup>. In Atlantic halibut, the sex determining gene *gsdf* is located in a region of chromosome 13 with restricted male and female RR<sup>56</sup>. In *S. senegalensis*, 30 significant sex-associated SNPs (66 if we consider the SNPs of separated families) were distributed throughout the SseLG18 with very low RR hot region (Fig. 6 and Supplementary Fig. S2 and S3). The shift and crossing between male and female RR suggest sex-specific restricted meiotic recombination events and that heterochiasmy might be involved in nascent sex chromosome system.

Most of SNP markers in the whole-population were heterozygous in males suggesting an XX/XY system. However, it should be noted high levels of incomplete penetrance in the families analysed (Supplementary Table S5). The fact that this proportion was even inverted in specific markers of F4 indicates a high effect of environmental factors on sex determination. The temperature seems to be a major factor that modifies sex ratios during

larval development generating skewed populations of neomales and neofemales <sup>62,63</sup>. Familial sex ratios in sole were reported to oscillate from 16 up to 90% males supporting a high impact of environmental factors to modulate sex differentiation and sex population ratios <sup>19</sup>.

After analyzing the hot region in SseLG18, the *fshr* appeared as a putative candidate for sex determination. The *fshr* locus was recently associated with male sex in flathead grey mullet with an incomplete penetrance as observed in sole <sup>64</sup>. These authors proposed that *fshr* might act as a proxy for the genetic transduction of environmental factors such as temperature. Under this hypothesis, sex determination would not rely on a single genetic cascade but a continuum of environmental and genetic factors. In sole, *fshr* was mainly expressed in testis<sup>65</sup>. The *Fshr* together with *StAR* are expressed in the steroidogenic Leydig cells and *Fshr* act as a promiscuous receptor that mediates the steroidogenic activity induced by both FSH and LH<sup>66,67</sup>. This double action supports a prolonged spermatogenesis and spermatid availability within the testis throughout the year mediated by FSH and the differentiation of spermatids into spermatozoa and subsequent spermiation mediated by LH <sup>66</sup>. Functional studies are needed to validate this putative candidate.

A synteny comparison of SseLGs with different flatfish genomes indicated that there was a one-to-one correspondence for 15 chromosomes, with some lineage-specific rearrangements (Fig. 8 and Supplementary Table S6). This correspondence between chromosomes was also confirmed when genome of turbot was compared with other flatfish demonstrating intrachromosomal rearrangements that shaped chromosome synteny and gene organization<sup>3</sup>. In our data, deviations from diagonal unlike in the comparison between male and female are indicative of this intense internal reorganization across species. The three SseLGs (SseLG1, SseLG2 and SseLG3) deserve special

attention as they can provide an evolutionary framework to understand the history of chromosome fusions and fissions that shaped the karyotypes in flatfish. The SseLG1, predicted as a metacentric chromosome by the analysis of recombination frequency (Fig. 6), was previously identified by cross-species genomic comparison as the largest metacentric chromosome in Senegalese sole suggesting it may be a proto-sexual chromosome<sup>12,16</sup>. Our data support the hypothesis that this chromosome has primarily emerged by a lineage-specific Robertsonian fusion, since the homologs in other flatfish maintained their integrity across evolution (Supplementary Fig. S7). A complex series of events including small chromosomal translocations and rearrangements, fusions, and pericentric inversions would explain the current gene content and organization<sup>12</sup>. Unlike SseLG1, the SseLG2 and SseLG3 contain those chromosomes whose remodeling have shaped the karyotypes in flatfish from n=24 in *P. olivaceus* to 22 *S. maximus* and 21 in *S. senegalensis* and *C. semilaevis*. A fusion model envisaged suggests a small number of chromosomes in the older lineage Paralichthyidae (9,14 and 16)<sup>10</sup> that combined with other chromosomes in a lineage-specific way could explain the major rearrangement events that shaped the karyotype in this species.

In conclusion, this study reports a new genome assembly for a male sole and a high-density SNP genetic map with 15,511 high-quality markers distributed in 21 linkage groups. The physical map was anchored to the consensus genetic map to generate 21 pseudo-chromosomes, in agreement with the number of chromosomes in this species. The larger map in females was the result of higher RR with distinct recombination landscape between sexes. Recombination frequencies were used to assess the putative morphology of SseLGs that will have to be validated by cytogenetic studies. A GWAS analysis identified 30 sex-associated markers, all located in SseLG18. A low recombining hot region hosted the putative candidate gene *fshr*. *In silico* comparison with other



Pleuronectiformes genomes demonstrated a high conservation of chromosome synteny, although with much intrachromosomal reorganization. Moreover, these changes in karyotype chromosome number were associated with lineage-specific Robertsonian fusions (i.e SseLG1 in *S. senegalensis*) and several other rearrangements that involved mainly three chromosomes in the ancestral lineage. The consistent physical and genetic maps reported in Senegalese sole represent a valuable genomic resource for functional and genome-wide association studies, and the identification of genomic processes involved in speciation.

## **Acknowledgements**

This study was funded by project RTA2017-00054-C03-01 and RTA2017-00054-C03-funded from MCIU/AEI/FEDER, UE and cofunded 80% by Programa Operativo FEDER de Andalucía 2014-2020, project PP.AVA.AVA201601.9 SOLEALGAE. Moreover, the study has received funding from EU H2020 research and innovation program under grant agreement 817992 ERANET-BLUEBIO COFUND project PCI2020-111994 BestBrood/AEI/10.13039/501100011033. IGC is funded by a predoctoral fellowship from INIA. This work would not have been possible without the computer resources and the technical support provided by the Plataforma Andaluza de Bioinformática of the University of Málaga and CNAG. We acknowledge the support of the Spanish Ministry of Science, Innovation and Universities to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme / Generalitat de Catalunya, the Spanish Ministry of Science and Innovation through the Instituto de Salud Carlos III, Generalitat de Catalunya through Departament de Salut and Departament d'Empresa i Coneixement and co-financing with funds from the European Regional Development

Fund by the Spanish Ministry of Science and Innovation corresponding to the Programa Operativo FEDER Plurirregional de España (POPE) 2014-2020 and by the Secretaria d'Universitats i Recerca, Departament d'Empresa i Coneixement of the Generalitat de Catalunya corresponding to the Programa Operatiu FEDER de Catalunya 2014-2020

## **Ethics declarations**

## **Competing Interests**

The authors declare no competing interests.

## **Contributions**

**IGC:** Investigation, Data analysis, Data curation, Writing- Original draft preparation. In silico analysis.

**JGG:** Investigation, genome annotation

**CB:** Methodology, fish production

**JMB:** Methodology. Genome sequencing

**TA:** Investigation, genome annotation. Writing - Review & Editing

**MGC:** Conceptualization, Funding acquisition, Writing - Review & Editing

**PAG:** Conceptualization, Investigation, Data analysis, Data curation, Writing - Review & Editing

**MM:** Resources, Conceptualization, Funding acquisition, Writing- Original draft preparation, Writing - Review & Editing.

## Captions

### **Figure 1. Selection of LOD score limit (Lod) to construct genetic map in LepMap3.**

A. The average of number of markers (nMarkers) positioned in linkage groups (left Y axis) and the number of linkage groups (nLG; right Y axis) for Lod values from 1 to 15 as implemented in the "SeparateChromosomes" module. Lod11 (shaded) indicates the value selected that grouped the markers in 21 LGs. B) Average number of markers recovered and added to the 21 LGs using decreasing LOD score iterations from 10 to 5 in the JoinSingles2 module.

**Figure 2. Dot plot comparison of scaffolds (SCF) assembled (A) or 21 pseudo-chromosomes (B) in the male with respect to SCF in the female.** Scale is indicated below.

**Figure 3. Genetic distance (cM) and SNP distribution across 21 linkage groups (SseLG) of the Senegalese sole.**

**Figure 4. Comparison of male and female genetic maps.** (A) Male *vs* female linkage groups lengths (cM) for the 21 Senegalese sole chromosomes. All chromosomes exhibit female-biased recombination. (B) Correlation between recombination map and physical map lengths in both males (blue) and females (orange). The determination coefficient  $R^2$  is shown separately for each sex.

**Figure 5. Recombination landscape averaged across linkage groups for (A) male and (B) female.** The recombination rates (cM/Mb) and the relative distance from the nearest telomere scaled by the chromosome length ( $f$ ) is represented. The red dashed line indicates the observed tendency. Panels (C) and (D) show the relationship between physical and genetic distances for SseLG1 in male and female, respectively. The square

inside the panels C and D show the specific recombination landscape. The complete information for all SseLGs is shown in Supplementary Fig. S2 and S3.

**Figure 6. Sex-associated SNPs and RR landscape for males and females in SseLG18.**

A) Manhattan plot of GWAS results for sex-associated SNPs using seven families. Significant markers are indicated in green. The horizontal red line represents the Bonferroni significance threshold. B) Distribution of all 66 sex-associated significant markers using seven families and by family (in red, Supplementary Table S5) and RR (cM/Mb) landscape of males and females. A hot region from 9.5 to 10.9 Mb containing the candidate gene *fshr* is indicated on the right side. Physical positions of SseLG18 in Mb are indicated in black. Black lines indicate non-significant markers in SseLG18.

**Figure 7. Plots illustrating the recombination frequency estimates (RFm) for intervals between markers along SseLG1 and SseLG20 in the male and female.** For each LG, RFm was calculated from both chromosomal extremities (right: red circles; left: blue circles), using each of the two terminal markers as a reference starting point. The RFm plots of SseLG1 and SseLG20 show a classical metacentric and acrocentric pattern, respectively. The RFm plots of all SseLGs are illustrated in Supplementary Fig. S4.

**Figure 8. Chromosomal alignment and synteny analysis between flatfish genomes.**

Top panel, Dot plot comparison of 21 pseudo-chromosomes of *S. senegalensis* with the genomes of the flatfish *C. semilaevis* (left), *S. maximus* (center) and *P. olivaceus* (right). Chromosome numbers or SseLGs are indicated. The chromosome fusions are boxed. Identity scale is indicated below. Bottom panel, syntenic comparison between flatfish genomes.

**Supplementary Fig. S1. Data filtering of markers with Mendelian errors.** (A) Data by family. The % markers with Mendelian errors in each family are indicated on the bars. (B) Relationship between number of markers with Mendelian errors and % individuals identified. Markers with an error % higher than the threshold (10% in dashed red line) were removed; (C) Relationship between the number of individuals and number of markers with Mendelian errors. Individuals with more than 5% of markers with Mendelian were removed (dashed red line).

**Supplementary Fig. S2. Relationship between physical and genetic distances for linkage groups (SseLG) in male.** The square below indicates the specific recombination landscape.

**Supplementary Fig. S3. Relationship between physical and genetic distances for linkage groups (SseLG) in female.** The square below indicates the specific recombination landscape.

**Supplementary Fig. S4. Plots illustrating the recombination frequency estimates (RFm) for intervals between markers for SseLG using the sex-average genetic map.** RFm values were calculated from both chromosomal extremities (right: red circles; left: blue circles), using each of the two terminal markers as a reference starting point.

**Supplementary Fig. S5. Sex-associated SNPs and gene distribution in SseLG18.** Significant SNPs using seven families are indicated in red and those significant by family are indicated in green. The gene names and distribution through the SseLG18 are shown. The hot region and the gene candidate *fshr* are indicated

**Supplementary Fig. S6. Dot plot alignment of *S. maximus* vs *C. semilaevis* (left), *S. maximus* vs *P. olivaceus* (center) and *P. olivaceus* vs *C. semilaevis* (right).**

Chromosome numbers are indicated. The chromosome fusions are boxed. Scale is indicated below.

**Supplementary Fig. S7. Synteny between *S. senegalensis* and the genomes of the flatfish *C. semilaevis*, *S. maximus* and *P. olivaceus*.** A) Whole genome synteny of *S. senegalensis* with respect to the other flatfish species; B) Synteny of SseLG1, SseLG2 and SseLG3 as visualized using Mizbee

## References

- 1 Bouza, C. *et al.* A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics* **177**, 2457-2467, <https://doi.org/10.1534/genetics.107.075416> (2007).
- 2 Reid, D. P. *et al.* A Genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Genetics* **177**, 1193-1205, <https://doi.org/10.1534/genetics.107.075374> (2007).
- 3 Maroso, F. *et al.* Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. *DNA Res* **25**, 439-450, <https://doi.org/10.1093/dnares/dsy015> (2018).
- 4 Nguyen, N. H., Rastas, P. M. A., Premachandra, H. K. A. & Knibb, W. First high-density linkage map and single nucleotide polymorphisms significantly associated with traits of economic importance in Yellowtail Kingfish *Seriola lalandi*. *Front Genet* **9**, 127, <https://doi.org/10.3389/fgene.2018.00127> (2018).
- 5 Liu, Z. J. & Cordes, F. J. DNA marker technology and their applications in aquaculture genetics. *Aquaculture* **238**, , 1-37, <http://dx.doi.org/10.1016/j.aquaculture.2004.05.027> (2004).
- 6 Wang, W. *et al.* High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. *PloS one* **10**, e0120410, <https://doi.org/10.1371/journal.pone.0120410> (2015).
- 7 Rastas, P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33**, 3726-3732, <https://doi.org/10.1093/bioinformatics/btx494> (2017).

- 8 Catchen, J., Amores, A. & Bassham, S. Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3 (Bethesda)*, <https://doi.org/10.1534/g3.120.401485> (2020).
- 9 Goodwin, S. *et al.* Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* **25**, 1750-1756, <https://doi.org/10.1101/gr.191395.115> (2015).
- 10 Shi, W. *et al.* Flatfish monophyly refereed by the relationship of Psettodes in Carangimorphariae. *BMC genomics* **19**, 400, <https://doi.org/10.1186/s12864-018-4788-5> (2018).
- 11 Azevedo, M. F. C., Oliveira, C., Pardo, B. G., Martinez, P. & Foresti, F. Cytogenetic characterization of six species of flatfishes with comments to karyotype differentiation patterns in Pleuronectiformes (Teleostei). *J. Fish Biol.* **70**, 1-15, <https://doi.org/10.1111/j.1095-8649.2006.01287.x> (2007).
- 12 Garcia-Angulo, A. *et al.* Evidence for a Robertsonian fusion in *Solea senegalensis* (Kaup, 1858) revealed by zoo-FISH and comparative genome analysis. *BMC genomics* **19**, 818, <https://doi.org/10.1186/s12864-018-5216-6> (2018).
- 13 Molina-Luzon, M. J. *et al.* First haploid genetic map based on microsatellite markers in Senegalese sole (*Solea senegalensis*, Kaup 1858). *Mar. Biotechnol (NY)* **17**, 8-22, <https://doi.org/10.1007/s10126-014-9589-5> (2015).
- 14 Garcia, E. *et al.* Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a Rex transposon located in a proto-sex chromosome. *Sci Rep* **9**, 17146, <https://doi.org/10.1038/s41598-019-53673-6> (2019).
- 15 Rodriguez, M. E. *et al.* Evolution of the Proto Sex-Chromosome in *Solea senegalensis*. *Int J Mol Sci* **20**, <https://doi.org/10.3390/ijms20205111> (2019).
- 16 Portela-Bens, S. *et al.* Integrated gene mapping and synteny studies give insights into the evolution of a sex proto-chromosome in *Solea senegalensis*. *Chromosoma* **126**, 261-277, <https://doi.org/10.1007/s00412-016-0589-2> (2017).
- 17 Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L. & Claros, M. G. in *Genomics in Aquaculture* (eds S. Mackenzie & S. Jentoft) 195–221 (Elsevier, 2016).
- 18 Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L. & Claros, M. G. in *The Biology of Sole* (eds J. Muñoz-Cueto, E. Mañanós-Sánchez, & F. Sánchez-Vázquez) Ch. B-6.1, 375-394 (CDC Press, 2019).
- 19 Guerrero-Cozar, I. *et al.* Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole. *Aquaculture* **539**, 736665, <https://doi.org/10.1016/j.aquaculture.2021.736665> (2021).

- 20 Guerrero-Cozar, I. *et al.* Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. *Sci. Reports* **10**, 21905, <https://doi.org/10.1038/s41598-020-78397-w> (2020).
- 21 Zimin, A. V. *et al.* Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787-792, <https://doi.org/10.1101/gr.213405.116> (2017).
- 22 Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677, <https://doi.org/10.1093/bioinformatics/btt476> (2013).
- 23 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, <https://doi.org/10.1093/bioinformatics/btv351> (2015).
- 24 Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* **35**, 543-548, <https://doi.org/10.1093/molbev/msx319> (2018).
- 25 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* **7**, e37135, <https://doi.org/10.1371/journal.pone.0037135> (2012).
- 26 Rochette, N. C. & Catchen, J. M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat Protoc* **12**, 2640-2659, <https://doi.org/10.1038/nprot.2017.123> (2017).
- 27 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, <https://doi.org/10.1086/519795> (2007).
- 28 Ouellette, L. A., Reid, R. W., Blanchard, S. G. & Brouwer, C. R. LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics* **34**, 306-307, <https://doi.org/10.1093/bioinformatics/btx576> (2018).
- 29 Rastas, P. Lep-Anchor: automated construction of linkage map anchored haploid genomes. *Bioinformatics* **36**, 2359-2364, <https://doi.org/10.1093/bioinformatics/btz978> (2020).
- 30 Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676, <https://doi.org/10.1093/bioinformatics/bti610> (2005).
- 31 Rezvoy, C., Charif, D., Gueguen, L. & Marais, G. A. MareyMap: an R-based tool with graphical interface for estimating recombination rates. *Bioinformatics* **23**, 2188-2189, <https://doi.org/10.1093/bioinformatics/btm315> (2007).



- 32 Limborg, M. T., McKinney, G. J., Seeb, L. W. & Seeb, J. E. Recombination patterns reveal information about centromere location on linkage maps. *Mol. Ecol. Resour.* **16**, 655-661, <https://doi.org/10.1111/1755-0998.12484> (2016).
- 33 Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296, <https://doi.org/10.1093/bioinformatics/btm108> (2007).
- 34 Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958, <https://doi.org/10.7717/peerj.4958> (2018).
- 35 Grabherr, M. G. *et al.* Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145-1151, <https://doi.org/10.1093/bioinformatics/btq102> (2010).
- 36 Yu, Y., Ouyang, Y. & Yao, W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* **34**, 1229-1231, <https://doi.org/10.1093/bioinformatics/btx763> (2018).
- 37 Claros, M. G., Seoane, P. & Manchado, M. Sequences and annotations of a provisional genome draft of a Senegalese sole female. figshare <https://doi.org/10.6084/m9.figshare.12472100.v1>. (2020).
- 38 Cordoba-Caballero, J. *et al.* An improved de novo assembling and polishing of Solea senegalensis transcriptome shed light on retinoic acid signalling in larvae. *Sci Rep* **10**, 20654, <https://doi.org/10.1038/s41598-020-77201-z> (2020).
- 39 Chen, S. *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* **46**, 253-260, <https://doi.org/10.1038/ng.2890> (2014).
- 40 Xu, X. W. *et al.* Draft genomes of female and male turbot *Scophthalmus maximus*. *Sci Data* **7**, 90, <https://doi.org/10.1038/s41597-020-0426-6> (2020).
- 41 Figueras, A. *et al.* Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA Res* **23**, 181-192, <https://doi.org/10.1093/dnares/dsw007> (2016).
- 42 Shao, C. *et al.* The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* **49**, 119-124, <https://doi.org/10.1038/ng.3732> (2017).
- 43 Ge, H. *et al.* De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Mol. Ecol. Resour.* **19**, 1461-1469, <https://doi.org/10.1111/1755-0998.13064> (2019).
- 44 Yang, L. *et al.* A Chromosome-Scale Reference Assembly of a Tibetan Loach, *Triplophysa siluroides*. *Front Genet* **10**, 991, 10.3389/fgene.2019.00991 (2019).

- 45 Hoghoughi, N., Barral, S., Vargas, A., Rousseaux, S. & Khochbin, S. Histone variants: essential actors in male genome programming. *J Biochem* **163**, 97-103, 10.1093/jb/mvx079 (2018).
- 46 Chiu, Y. L., Shikina, S., Yoshioka, Y., Shinzato, C. & Chang, C. F. De novo transcriptome assembly from the gonads of a scleractinian coral, *Euphyllia ancora*: molecular mechanisms underlying scleractinian gametogenesis. *BMC genomics* **21**, 732, 10.1186/s12864-020-07113-9 (2020).
- 47 Merlo, M. A. *et al.* Analysis of the histone cluster in Senegalese sole (*Solea senegalensis*): evidence for a divergent evolution of two canonical histone clusters. *Genome* **60**, 441-453, 10.1139/gen-2016-0143 (2017).
- 48 Rooney, A. P., Piontkivska, H. & Nei, M. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol Biol Evol* **19**, 68-75, 10.1093/oxfordjournals.molbev.a003983 (2002).
- 49 Zhang, J., Zuo, Z., Zhu, W., Sun, P. & Wang, C. Sex-different effects of tributyltin on brain aromatase, estrogen receptor and retinoid X receptor gene expression in rockfish (*Sebastes marmoratus*). *Mar Environ Res* **90**, 113-118, 10.1016/j.marenvres.2013.06.004 (2013).
- 50 Shao, C. *et al.* Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. *DNA Res* **22**, 161-170, <https://doi.org/10.1093/dnares/dsv001> (2015).
- 51 Vega, L., Díaz, E., Cross, I. & Rebordinos, L. Caracterizaciones citogenética e isoenzimática del lenguado *Solea senegalensis* Kaup, 1858. *Boletín Inst Español Oceanogr.* **18**, 1-6 (2002).
- 52 Mascher, M. & Stein, N. Genetic anchoring of whole-genome shotgun assemblies. *Front Genet* **5**, 208, <https://doi.org/10.3389/fgene.2014.00208> (2014).
- 53 Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W. & Smadja, C. M. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc Lond B Biol Sci* **372**, <https://doi.org/10.1098/rstb.2016.0455> (2017).
- 54 Sardell, J. M. & Kirkpatrick, M. Sex differences in the recombination landscape. *Am Nat* **195**, 361-379, <https://doi.org/10.1086/704943> (2020).
- 55 Sardell, J. M. *et al.* Sex differences in recombination in Sticklebacks. *G3 (Bethesda)* **8**, 1971-1983, <https://doi.org/10.1534/g3.118.200166> (2018).
- 56 Edvardsen, R. *et al.* Heterochiasmy facilitated the establishment of *gsdf* as a novel sex determining gene in Atlantic halibut. <https://doi.org/10.1101/2020.11.24.396218> (2020).

- 57 Ruan, X., Wang, W., Kong, J., Yu, F. & Huang, X. Genetic linkage mapping of turbot (*Scophthalmus maximus* L.) using microsatellitemarkers and its application in QTL analysis. *Aquaculture* **308**, 89-100, <https://doi.org/10.1016/j.aquaculture.2010.08.010> (2010).
- 58 Song, W. *et al.* Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). *PloS one* **7**, e52097, <https://doi.org/10.1371/journal.pone.0052097> (2012).
- 59 Castaño-Sanchez, C. *et al.* A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *BMC genomics* **11**, 554, <https://doi.org/10.1186/1471-2164-11-554> (2010).
- 60 Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079-1083, <https://doi.org/10.1038/nature08441> (2009).
- 61 Wright, A. E. *et al.* Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat Commun* **8**, 14251, <https://doi.org/10.1038/ncomms14251> (2017).
- 62 Viñas, J., Asensio, E., J.P., C. & Piferrer, F. Gonadal sex differentiation in the Senegalese sole (*Solea senegalensis*) and first data on the experimental manipulation of its sex ratios. *Aquaculture*, 384-386, <https://doi.org/10.1016/j.aquaculture.2012.12.012> (2012).
- 63 Blanco-Vives, B. *et al.* Exposure of larvae to daily thermocycles affects gonad development, sex ratio, and sexual steroids in *Solea senegalensis*, kaup. *J Exp Zool A Ecol Genet Physiol* **315**, 162-169, <https://doi.org/10.1002/jez.664> (2011).
- 64 Ferraresso, S. *et al.* fshr: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations. *iScience* **24**, 10186, <https://doi.org/10.1016/j.isci.2020.101886> (2021).
- 65 Chauvigne, F. *et al.* Functional and evolutionary analysis of flatfish gonadotropin receptors reveals cladal- and lineage-level divergence of the teleost glycoprotein receptor family. *Biol Reprod* **82**, 1088-1102, <https://doi.org/10.1095/biolreprod.109.082289> (2010).
- 66 Chauvigne, F. *et al.* Follicle-stimulating hormone and luteinizing hormone mediate the androgenic pathway in Leydig cells of an evolutionary advanced teleost. *Biol Reprod* **87**, 35, <https://doi.org/10.1095/biolreprod.112.100784> (2012).
- 67 Chauvigne, F., Zapater, C., Gasol, J. M. & Cerda, J. Germ-line activation of the luteinizing hormone receptor directly drives spermiogenesis in a nonmammalian vertebrate. *Proc Natl Acad Sci U S A* **111**, 1427-1432, <https://doi.org/10.1073/pnas.1317838111> (2014).