A chromosome-level genome assembly enables the identification of the follicule stimulating hormone receptor as the master sex determining gene in the flatfish Solea senegalensis

Title: A chromosome-level genome assembly enables the identification of the follicule stimulating hormone receptor as the master sex determining gene in *Solea senegalensis*

Authors: de la Herrán R. [1]*, Hermida M. [2]*, Rubiolo J. [2*], Gómez-Garrido J. [3], Cruz F.[3], Robles F. [1], Navajas-Pérez R. [1], Blanco, A. [2], Villamayor P. R. [2], Torres D. [2], Sánchez-Quinteiro P. [4], Ramirez D. [5], Rodríguez M. E. [5], Arias-Pérez A. [5], Cross I. [5], Duncan N. [6], Martínez-Peña T. [7], Riaza A. [7], Millán A. [8], De Rosa M. C.[9], Pirolli D[9]., Gut M. [3], Bouza C. [2], Robledo D. [10], Rebordinos L. [5], Alioto T. [3,11], Ruíz-Rejón C. [1], and Martínez P. [2]

[1] Departamento de Genética, Facultad de Ciencias, Universidad de Granada, 18071, Granada, Spain.

[2] Departamento de Zoología, Genética y Antropología Física; Facultad de Veterinaria. Universidade de Santiago de Compostela. Campus de Lugo. 27002 Lugo, Spain.

[3] Centre Nacional d'Anàlisi Genòmica (CNAG-CRG), Centre de Regulació Genómica, Parc Científic de Barcelona, Barcelona, 08028, Spain.

[4] Departamento de Anatomía, Producción Animal y Ciencias Clínicas Veterinarias Facultad de Veterinaria. Universidade de Santiago de Compostela. Campus of Lugo. 27002 Lugo, Spain.

[5] Departamento de Biomedicina, Biotecnología y Salud Pública CASEM - Facultad de Ciencias del Mar y Ambientales, Universidad de Cádiz, 11519 Puerto Real, Cádiz Spain

[6] IRTA Sant Carles de la Rapita, AP200, 43540 Sant Carles de la Rapita, Tarragona, Spain.

[7] Stolt Sea Farm SA, Departamento I+D, Lira, 15292 Carnota, A Coruña, Spain

[8] Geneaqua SL, 27002, Lugo, Spain

[9]Institute of Chemical Sciences and Technologies "Giulio Natta" (SCITEC) – CNR c/o Catholic University of Rome, 00168, Rome, Italy

[10] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK

[11] Universitat Pompeu Fabra (UPF), Barcelona (Spain).

* These authors contributed equally to this work

Corresponding author

Professor Paulino Martínez

Departamento de Zoología, Genética y Antropología Física; Facultad de Veterinaria. Universidade de Santiago de Compostela. Campus de Lugo. 27002 Lugo, Spain.

TLF +34 982 82 24 25

Fax +34 982 82 20 01

e-mail: paulino.martinez@usc.es

Running title: Master sex determining gene of *Solea senegalensis*

Key words: Senegalese sole, whole genome sequencing, genetic map, sex determination, follicule stimulating hormone receptor

Abstract

Sex determination (SD) shows huge variation among fish and a high evolutionary rate, as illustrated by the Pleuronectiformes (flatfishes). ~~, an~~ This order is characterized by its adaptation to demersal life, compact genomes and diversity of SD mechanisms. Here, we assembled the *Solea senegalensis* genome, a flatfish of great commercial value, into 82 contigs (614 Mb) combining long- and short-read sequencing, which were next scaffolded using a highly dense genetic map (28,838 markers, 21 linkage groups), representing 98.9% of the ~~whole~~ assembly. Further, we established the correspondence between the assembly and the 21 chromosomes ~~of its haploid karyotype~~ by using BAC-FISH. Whole genome resequencing of six males and six females enabled the identification of 41 SNP variants in the follicle stimulating hormone receptor (*fshr*) consistent with an XX / XY SD system. The observed sex association was validated in a broader independent sample, providing a novel molecular sexing tool. *Fshr* displayed differential gene expression between male and female gonads from 86 days post-fertilization, when the gonad is still an undifferentiated primordium, concomitant with the activation of *amh* and *cyp19a1a*, testis and ovary marker genes, respectively, in males and females. The Y-linked *fshr* allele, which included 24 non-synonymous variants and showed a highly divergent 3D protein structure, was overexpressed in males compared to the X-linked allele at all stages of gonadal differentiation. We hypothesize a mechanism hampering the action of the follicle stimulating hormone driving the undifferentiated gonad toward testis.

Introduction

Sex determination (SD) refers to the mechanism controlling the fate of the gonadal primordium at the initial stages of development responsible for the sex of a mature individual. While highly conserved and evolutionary mature SD systems have been reported in mammals and birds, increasing data from ectothermic vertebrates provide a sharply different picture (Martínez et al., 2014; Guiguen et al., 2019). Fish display highly diverse chromosome SD systems (Cioffi et al., 2017) and several master SD genes have been reported in this group (Martínez et al., 2014). Among them, classical transcription factors such as *dmy* (Matsuda et al., 2002; Wang et al., 2022), *sox3* (Takehana et al., 2014) or *sox2* (Martínez et al., 2021); transforming growth factor ß-related genes such as *gsdf* (Myosho et al., 2012; Herpin et al., 2021) or *amh* (Hattori et al., 2012; Pan et al., 2019) and its receptor *amhr2* (Kamiya et al., 2012; Feron et al., 2020; Nakamoto et al., 2021; Wen et al., 2022; Nacif et al., 2022; Zheng et al., 2022); genes related to the steroidogenic pathway such as *bcar1* (Bao et al., 2019) or *hsd17b1* (Koyama et al., 2019); and finally, some unexpected, such as the interferon-related *sdY* gene in salmonids (Yano et al., 2013). The recent identification of several master SD genes in this group has been associated with the highly contiguous and reliable chromosome-level genome assemblies achieved via improvements in long-read sequencing technologies, scaffolding and bioinformatic approaches (Ramos & Antunes, 2022).

Flatfish (Pleuronectiformes) represent a fish order with notable adaptations to demersal life (Chen et al., 2014; Figueras et al., 2016; Robledo et al., 2016; Shao et al., 2017; Lü et al., 2021). They experience a remarkable metamorphosis from the bilateral morphology of pelagic larvae to the flat morphology typical of this group (Shao et al., 2017; Lü et al., 2021). The adaptation to demersal lifestyle promoted a quick diversification, reflected by a higher molecular evolutionary rate than their bilateral counterparts (Lü et al., 2021). This rapid radiation also led

to a large variety of SD mechanisms, involving different master SD genes and non-orthologous SD regions in all species analysed to date (Luckenbach et al., 2009; Martínez et al., 2021). Flatfish genomes are compact (~500-700 Mb; Robledo et al., 2017; Lü et al., 2021), which has facilitated the achievement of chromosome-level genome assemblies in species from several families of the order (Martínez et al., 2021; Guerrero-Cózar et al., 2021; Lü et al., 2021).

S. senegalensis is a valuable commercial flatfish which lives on sandy or muddy bottoms, from brackish lagoons and shallow waters to deeper coastal regions (Cabral, 2000, Díaz-Ferguson et al., 2012). Further, it is a very promising aquaculture species (currently ~2000 tons; Ana Riaza, pers. comm.), which has promoted the development of a number of genomic resources and tools in the last decade (Robledo et al., 2017), including a recent whole genome assembly (Guerrero-Cózar et al., 2021). The species shows high larval survival rates as compared to other flatfish, such as turbot, and great capacity for adjusting to intensive production, the commercial size being reached at the age of 1 year (Morais et al., 2016). Females have a higher growth rate and mature at three years of age, while males do at two years (Viñas et al., 2013), so obtaining all-female populations is an appealing strategy to increase growth rate at farms. As other Pleuronectiformes, the Senegalese sole undergoes a process of metamorphosis at 10-12 dph (days post-hatching) and for about 7 days to acquire the asymmetry adapted to benthic lifestyle (Martín et al., 2019). ~~As in other flatfish, *S. senegalensis* females outgrow males (Viñas et al., 2013), so obtaining all-female populations is an appealing strategy to increase growth rate at farms.~~ An important limitation for the expansion of *S. senegalensis* aquaculture has to do with the low performance of F1 males (born and reared in captivity) compared to wild specimens reared in captivity, hampering the development of breeding programs (Martín et al., 2019). Recently, Guerrero-Cózar et al. (2022) identified the follicule stimulating hormone receptor as a candidate SD determining gene using RADseq but highlighted the need for further validation

Formatted: Font: (Default) Times New Roman, 12 pt, English (United States)

the SD mechanism, and how external cues are connected to gonad development via neural communication, will be essential to boost *S. senegalensis* aquaculture. new highly contiguous genome assembly in *S. senegalensis* and integrated all mapping resources (physical, genetic and cytogenetic), thus providing a robust framework for evolutionary genomics in pleuronectiforms and teleosts. This assembly was used to identify the putative master SD gene of the species, the follicle stimulating hormone receptor (*fshr*), by re-sequencing a sample of males and females. The pattern of variation observed confirmed an XX / XY system and enabled the development of a molecular tool for sex determination in this species. Functional differences between male and female gonads were evaluated throughout gonad development, from the undifferentiated primordium to maturation. Our study, supported by gene expression data, three-dimensional protein structure predictions and whole genome re-sequence analysis of males and females, suggests that a functionally divergent, overexpressed Y-linked allele of the follicle stimulating hormone receptor (*fshry*) may determine the testis fate of the undifferentiated primordium in *S. senegalensis*.

Materials and Methods

Samples

Biological and genomic metadata of *S. senegalensis* samples analysed in this study are recorded in Supplemental Table Metadata.

DNA sequencing

Long-read Whole Genome Sequencing

High molecular weight DNA was purified from one female of *S. senegalensis* whole blood using the Nanobind CBB Big DNA Kit (Circulomics) following the manufacturer´s instructions

and eluted in EB buffer (Qiagen). The sequencing libraries were prepared using the Ligation sequencing kit SQK-LSK109 from Oxford Nanopore Technologies (ONT). Briefly, 4.0 µg of the DNA was DNA-repaired and DNA-end-repaired using NEBNext FFPE DNA Repair Mix (NEB) and the NEBNext UltraII End Repair/dA-Tailing Module NEB. ~~and followed by~~ Then, ~~the~~ sequencing adaptor ligation, purification by 0.4X AMPure XP Beads and elution in Elution Buffer (SQK-LSK109) was accomplished. The sequencing runs were performed on GridION Mk1 (ONT) using a Flowcell R9.4.1 FLO-MIN106D (ONT) and the sequencing data was collected for 110 hours. The quality parameters of the sequencing runs were monitored by the MinKNOW platform version 4.1.2 in real time and base-called with Guppy version 4.2.3.

Short-read whole genome sequencing

The short-insert paired-end libraries for the whole genome sequencing (WGS) were prepared from DNA of the same female used for long-read sequencing with PCR free protocol using KAPA HyperPrep kit (Roche) with some modifications. In short, 1.0 µg of genomic DNA was sheared on a Covaris™ LE220-Plus (Covaris) and size-selected for the fragment size of 220-550 bp with AMPure XP beads (Agencourt, Beckman Coulter). The genomic DNA fragments were then end-repaired and~~-~~ adenylated. Next, ~~and Illumina platform~~ compatible adaptors for Illumina platforms with unique dual indexes ~~and~~ including unique molecular identifiers (Integrated DNA Technologies) were ligated. The libraries were quality controlled on an Agilent 2100 Bioanalyser with the DNA 7500 assay (Agilent) for size and quantified by Kapa Library Quantification Kit for Illumina platforms (Roche).

RNA-Seq

Two different projects fed RNA data into our study, which included mRNA and small RNA sequencing. In the first project RNA-Seq was performed on brain, liver and head kidney using

pools of two females for each tissue. In the second one, RNA-Seq and small RNA-Seq were ~~performed on~~ obtained from vertebral bone, muscle, fin and 31 days after hatching (dah) postlarvae using several juveniles (123 dah) or postlarvae samples. Total RNA extraction was performed in both projects using the RNeasy mini kit (Qiagen) with DNase treatment. ~~and~~ RNA quantity and quality were evaluated with the Qubit® RNA BR Assay kit (Thermo Fisher Scientific) and the RNA integrity estimated by using RNA 6000 Nano Bioanalyser 2100 Assay (Agilent). Next, single individuals or equimolar RNA pools of several individuals were used for library construction of each tissue after evaluation of individual RNA extractions.

The RNA-Seq libraries of the first project were prepared with KAPA Stranded mRNA-Seq Illumina® Platforms Kit (Roche) following the manufacturer´s recommendations. Briefly, 500 ng of total RNA was used for the poly-A fraction enrichment with oligo-dT magnetic beads, following the mRNA fragmentation. The strand specificity was achieved during the second strand synthesis performed in the presence of dUTP instead of dTTP. The blunt-ended double stranded cDNA was 3´adenylated before Illumina platform compatible adaptors with unique dual indexes and unique molecular identifiers (Integrated DNA Technologies) were ligated. The ligation product was enriched with 15 PCR cycles and the final library was validated on an Agilent 2100 Bioanalyser with the DNA 7500 assay.

The short-read -WGS and RNA-Seq libraries were sequenced on NovaSeq 6000 (Illumina) in paired-end, with a read length of 151 bp for the WGS and 100 bp for the RNA-Seq following the manufacturer's protocol for dual indexing. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 3.4.4) and followed by generation of FASTQ sequence files.

A similar methodology was applied for the second RNA-Seq project with minor technical modifications: total RNA was extracted using TRIzol Reagent (Life Technologies) and the

quantity and purity of RNA was performed in a NanoDrop® ND-1000 spectrophotometer (NanoDrop® Technologies Inc.). Samples of total RNA of the different stages and tissues were delivered to Novogene-Europe, UK, for constructing poly-A enriched mRNA and small RNA libraries to be sequenced in 150-bp pair-end and 50-bp single-end, respectively, using an Illumina NovaSeq 6000 platform.

Genome assembly

Before assembly, reads were preprocessed as follows: the Illumina reads were trimmed using Trim-galore v0.6.6 (with options *--gzip -q 20 --paired --retain_unpaired*) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and the nanopore reads were filtered using FiltLong v0.2.0 (with options *--min_length* 5000 *--target_bases* 40,000,000,000) (FiltLong: https://github.com/rrwick/Filtlong). The filtering of nanopore data ensured having reads of at least 5 kb while optimizing for both length and higher mean base qualities, keeping 40 Gb (~ 65x coverage).

We assembled the filtered ONT reads with NextDenovo v2.4.0 (https://github.com/Nextomics/NextDenovo) applying the options: minimap2_options_raw = -x ava-ont, minimap2_options_cns = -x ava-ont -k17 –w17 and seed_cutoff=10k. The resulting contigs were polished with Nextpolish v1.3.1 (Hu et al., 2020) using two rounds of long-read polishing and two rounds of short-read polishing. This assembly (fSolSen1.1) was evaluated with BUSCO v 5.4.0 (Simão et al., 2015) using actinopterygii_odb10 in genome mode, Merqury v 1.1 (Rhie et al., 2020) for consensus quality (QV) and k-mer completeness. Note that the merqury QV is computed as the Phred quality score treating E as the base error probability $QV = -10 \times \log_{10} E$. Finally, to compute the contiguity we used our in-house script Nseries.pl (https://github.com/cnag-aat/assembly_pipeline/blob/v2.0.0/scripts/Nseries.pl).

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Field Code Changed

Formatted: English (United States)

Genome annotation

Repeats present in the fSolSen1 genome assembly were annotated with RepeatMasker v4-0-7 (http://www.repeatmasker.org) using the custom repeat library available for *Danio rerio*. Moreover, a new repeat library specific for our assembly was made with RepeatModeler v1.0.11. After excluding those repeats that were part of repetitive protein families (performing a BLAST search against UniProt) from the resulting library, RepeatMasker was run again with this new library in order to annotate the specific repeats.

The gene annotation of the assembly was obtained by combining transcript alignments, protein alignments and *ab initio* gene predictions. Firstly, RNA-Seq reads obtained from several tissues and developmental stages, either sequenced specifically in this study (brain, liver, bone, muscle, head kidney, fin, and 31 dah postlarvae) or existing in public databases, were aligned to the genome with STAR (Dobin et al., 2013) (v-2.7.2a). Transcript models were subsequently generated using Stringtie (Pertea et al., 2015) (v2.0.1) on each BAM file and then all the models produced were combined using TACO v0.6.2. ~~Finally,~~ PASA assemblies were produced with PASA (Haas et al., 2008) (v2.4.1) and~~.~~ t~~T~~he TransDecoder program, which is part of the PASA package, was run on the PASA assemblies to detect coding regions in the transcripts. Secondly, the complete *D. rerio*, *Scophthalmus maximus* and *Cynoglossus semilaevis* proteomes were downloaded from UniProt in April 2020 and aligned to the genome using SPALN (Slater & Birney, 2005) (v2.4.03). *Ab initio* gene predictions were performed on the repeat-masked fSolSen1 assembly with three different programs: GeneID (Parra et al., 2000) v1.4, Augustus (Stanke et al., 2006) v3.3.4 and Genemark-ES (Lomsadze et al., 2014) v2.3e with and without incorporating evidence from the RNA-Seq data. The gene predictors were run with trained parameters for human, except Genemark that runs in a self-trained manner. Finally, all the data were combined into consensus CDS models using EvidenceModeler-1.1.1 (EVM, Haas et al.,

2008). Additionally, UTRs were identified, and alternative splicing forms annotated through two rounds of PASA annotation updates. Functional annotation was performed on the annotated proteins with Blast2go (Conesa et al., 2005). First, a Diamond blastp (Buchfink et al., 2021) search was made against the nr (last accessed May 2021) and Uniprot (last accessed August 2021) databases. ~~Furthermore,~~Then, Interproscan (Jones et al., 2014) was run to detect protein domains on the annotated proteins. All these data were combined by Blast2go, which produced the final functional annotation results.

The non-coding RNA annotation required several steps. First, those expressed transcripts that had been assembled by PASA but that had not been annotated as Protein-Coding genes were tagged as long-non-coding RNAs. The reason for this step is that it helps to have putative lncRNAs annotated before using annotation for downstream analysis. However, due to lack of lncRNAs conservation between species, no function was assigned to these lncRNA genes. Moreover, in order to remove false positives, transcripts overlapping with other Protein-coding genes or repeats were not included into the lncRNA annotation. Finally, only transcripts longer than 200 bp were considered lncRNAs.

We also used small RNAs data from vertebral bone, muscle, fin and 31 dah postlarvae to facilitate their annotation in the *S. senegalensis* genome. The corresponding reads were aligned with STAR (Dobin et al., 2013) (v-2.7.2a) with parameters (-outFilterMultimapNmax 25 --alignIntronMax 1 --alignMatesGapMax 1000000 --outFilterMismatchNoverLmax 0.05 --outFilterMatchNmin 16 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0). The resulting mappings were processed to produce the annotation of small non-coding RNAs. First, TACO was run to assemble the reads into transcripts. Transcripts overlapping exons from the protein-coding or lncRNA annotations were removed from the set of small non-coding

RNAs (sncRNAs). Finally, the program cmsearch (Cui et al., 2016) (v1.1.4) that includes the embedded tool Infernal (Nawrocki & Eddy, 2013) was run on the sncRNAs against the RFAM (Nawrocki et al., 2015) database of RNA families (v14.6) in order to annotate products of those genes.

The final non-coding annotation contains the lncRNAs and the sncRNAs. The resulting transcripts were clustered into genes using shared splice sites or substantial sequence overlap as criteria for designation as the same gene.

Genetic map construction and genome reassembly

Six *S. senegalensis* breeders, three males and three females, were used for producing three full-sib families using *in vitro* fertilisations as described for the pProof-of-concept experiment by Ramos-Júdez et al. (2021). Briefly, females were selected by maturity status and induced to ovulate with an injection of 5 µg kg^{-1} of GnRHa (Sigma code L4513, Sigma, Spain), whilst no hormones were used on males. Gametes were extracted from females and males using gentle abdominal pressure, fertilised (1 male x 1 female) and incubated in 30 L incubators. Larvae were randomly sampled from the incubators and placed in 70% ethanol 8 dpf. All this work was performed at the IRTA Sant Carles de la Rápita Center (Catalonia, Spain).

Library preparation for SNP calling and genotyping followed the 2b-RAD protocol (Wang et al., 2012) with slight modifications (Maroso et al., 2018). The use of this methodology for genotyping was successfully applied for high density genetic map construction in other flatfish such as *S. maximus* (Maroso et al., 2018). Briefly, DNA samples were adjusted to 80 ng/µL and digested using the IIb type restriction enzyme AlfI (Thermo Fisher). Specific adaptors and individual sample barcodes were ligated and the resulting fragments amplified. After PCR purification, samples were quantified and equimolarly pooled. The pools were sequenced on a

NextSeq 500 Illumina sequencer in the FISABIO facilities (Valencia, Spain). A total of 81, 77 and 71 offspring from the three full-sib families founded (Fam1, Fam2 and Fam3, respectively) were evenly mixed within each family and sequenced in three independent runs with parents also included at double concentration.

Demultiplexed reads according to the sample barcodes were first trimmed to 36 nucleotides and a custom perl script was used to remove reads without the AlfI recognition site in the correct position (https://github.com/abhortas/USC-RAD-seq-scripts). Then, reads were processed using the *process_radtags* module in STACKS v2.0 (Catchen et al., 2013), removing reads with uncalled nucleotides or a mean quality score below 20 in a sliding window of 9 nucleotides. Bowtie 1.1.2 (Langmead et al. 2009) was used to align the filtered reads against the assembled genome (see above), allowing a maximum of three mismatches and a unique valid alignment. Finally, the output files were used to feed the *gstacks* module in STACKS, using the *marukilow* model to call variants and genotypes.

To build the genetic map, SNPs with extreme deviations from Mendelian segregation (chi-square; $P < 0.001$) were removed, and only informative SNPs genotyped in at least 60% offspring were used. The *grouping* function of JoinMap 4.1 (Stam, 1993) was used to build linkage groups (LG), based on an increasing series of LOD scores from 7.0 to 10.0 to accommodate to the 21 chromosomes (C) of the *S. senegalensis* karyotype (Vega et al., 2002). Marker ordering was performed using the Maximum Likelihood (ML) algorithm with default parameters with the Kosambi mapping function used to compute centi-Morgans (cM) map distances. Consensus maps were built using MergeMap (Wu et al., 2011) and visualized with MapChart 2.3 (Voorrips 2002) and Circos software (Krzywinski et al., 2009).

Chromonomer 1.13 (Catchen et al., 2020) was used with default parameters to anchor and orient the contigs of the genome to the genetic map. Chromonomer assigns each scaffold to a linkage

Formatted: Font: (Default) Times New Roman

303 group and find the maximum set of non-conflicting markers to assign a forward or reverse
304 orientation to the scaffold. Further, the correspondence between genetic map and contigs
305 enabled the manual curation of original contigs that mapped to different linkage groups (LG).
306 This occurred with the longest contig (see Results), which was split into two fragments at a
307 specific position following the method by Maroso et al. (2018). Briefly, we first considered the
308 middle of the gap between the two markers that flanked the two fragments mapping on different
309 LGs according to Chromonomer information, and then refined the breakpoint by comparing the
310 sequence between both markers with orthologous regions of other flatfish chromosome-level
311 genomes available in Ensembl. The program RIdeogram was used to visualize the
312 correspondence between genome contigs and chromosomes (Hao et al., 2020).

313 *Cytogenetic map and mapping integration*

314 The BAC clones used for mapping came from a *S. senegalensis* BAC library including 29,184
315 clones (García-Cegarra et al., 2013). To establish the correspondence between LGs / genome
316 scaffolds and the chromosomes for mapping integration, 141 BAC clones previously positioned
317 on chromosomes using BAC-FISH were located in the 21 scaffolds of the genome by a
318 megablast search tool from blast algorithm (Altschul et al., 1990) using the following
319 parameters: Evalue < E-20; max_hsps = 10; sequence overlap > 5 kb. Alignments were
320 visualized with the Integrative Genomics Viewer (IGV) program (Robinson et al., 2011) and
321 manually explored.

322 *Sex determining (SD) gene candidates*

323 DNA extracted from fin-clips of six adult males and six adult females were re-sequenced using
324 150 bp PE reads on an Illumina NovaSeq 6000 System to 20x coverage in the Centre Nacional
325 d'Anàlisi Genòmica (CNAG, Barcelona, Spain) Platform following the outlined short-read

WGS protocol. The high throughput SNP screening of the genome was expected to identify a narrow sex associated region to be further validated on specific markers in a broad sample of males and females. The reads were filtered using fastp v.0.19.7 (Chen et al., 2018), trimming bases with Phred quality <15 and reads with length <30 bp, and then each sample was aligned independently against the newly assembled reference *S. senegalensis* genome using Burrows-Wheeler Aligner v.0.7.17 (Li and Durbin, 2009) with default parameters. A broad SNP dataset was identified and genotyped in those ~~twelve~~ six males and six females using SAMtools v.1.10 (Li et al., 2009) and SNPs showing quality scores below 20 were removed. This SNP dataset was used to estimate the relative component of genetic differentiation between males and females (F_{ST}) and the intra-sex fixation index (F_{IS}) across the whole genome using GENEPOP 4.7.5 (Raymond & Rousset, 1995). F_{ST} and F_{IS} values were averaged over 50 consecutive SNPs and explored using sliding windows across each chromosome of the genome to look for deviation from the null hypothesis (F_{ST} and $F_{IS} = 0$).

Developing a tool for sexing

As shown in Results, *fshr* was the most consistent sex determining (SD) candidate gene. It included large number of diagnostic markers across its whole length, homozygous in females and heterozygous in males, which agrees with the XX / XY system reported for *S. senegalensis* (Molina-Luzón et al., 2015). Several of these markers were used to develop a molecular tool to identify sex using a non-invasive method (e. g. from a fin-clip). This tool was valuable to assess gene expression across gonad development from the undifferentiated germinal primordium, especially at those stages where the gonad was still undifferentiated (see below). A SNaPshot assay used for SNP genotyping was developed for three diagnostic markers, chosen by their technical feasibility, i. e. with no other polymorphism within ±100 bp from the SNP, according to the resequencing information of six males and six females, and using three different regions

(exons 12 and 14, and 3´UTR). The SNaPshot assay consists of two consecutive reactions; the first step involves the PCR amplification of the region where the target SNP is located and the second a mini-sequencing reaction from a primer adjacent to the SNP site using dideoxy nucleotides. Thus, two flanking PCR primers and one internal primer adjacent to the variable site were designed for genotyping each SNP using Primer 3 software (Rozen and Skaletsky, 2000) taking flanking sequences from our assembled genome. SNaPshot products were separated in an ABI 3730xl Genetic Analyser (Applied Biosystems) and results were analysed with GeneMapper 4.0 software (Applied Biosystems). PCR was performed on a Verity ™ 96-Well Thermal Cycler (Applied Biosystems) as follows: initial denaturation at 95 °C for 5 min, 30 cycles of denaturation at 94 °C for 45 s, annealing temperature at 58°C for 50 s, and extension at 72 °C for 50 s; a final extension step was done at 72 °C for 10 min. Subsequently, 1 µL of the PCR product was purified by incubation with 0.5 µL illustra™ ExoProStar™ 1-STEP Kit at 37 °C for 15 min followed by 85 °C for 15 min to eliminate unincorporated primers and dNTPs. The SNaPshot mini-sequencing reaction was carried out using the SNaPshot® Multiplex Kit (Applied Biosystems) in an ABI Prism 3730xl DNA sequencer. For each reaction, 1.5 µL of purified PCR product, 0.5 µL (2 µM) of the internal primer and 2 µL of SNaPshot™ Multiplex Ready Reaction Mix were used in a final volume of 5 µL. The reaction profile consisted of initial denaturation at 96 °C for 1 min, 30 cycles at 96 °C for 10 s, 55 °C for 5 s, and 60 °C for 30 s. The extension product was incubated with 1 µL of shrimp alkaline phosphatase at 37 °C for 60 min followed by 85 °C for 15 min to remove unincorporated dideoxinucleotides (ddNTPs) after thermal cycling. The three sex-associated candidate SNPs were checked in a small sample of adult males and females for SNaPshot performance, and then, the best one was validated in a large sample of 48 male and 48 female *S. senegalensis* adults provided by a farm company, where they are routinely used for breeding. Additionally,

38 individuals where information for gonadal sex and genotyping for the sex marker was available, were also used to check for the association between genotypic and the phenotypic sex, thus totaling 134 individuals: 12 whole genome re-sequenced individuals (6 males and 6 females); 6 individuals used for anatomy/histology analysis (126 dpf); and 10 juveniles and 10 fry (126 dpf) from the qPCR / RNA-Seq assays.

Gonad differentiation: histological evaluation

Three fish per sex and stage were collected at 84, 98 and 126 dpf, juveniles (315 dpf) and mature adults (810 dpf) at Stolt Sea Farm SL facilities (Ribeira / Cervo, Spain) and sacrificed by decapitation. All fishes were maintained at the same standard temperature, air-flow and feeding conditions of the usual production protocol of the company until sacrifice. Gonads were dissected fresh for macroscopic evaluation and classified as testes or ovaries by visual inspection in juveniles and adults. At the three initial stages, the molecular tool outlined before was used for sexing. Then, the gonads were fixed by immersion in 4% paraformaldehyde and later embedded in paraffin wax to be cut into sagittal sections 3-6 µm and stained with haematoxylin-eosin for optical microscopy evaluation. Experimental procedures for the use of farm animals were carried out following the regulations of the University of Santiago de Compostela and Stolt Sea Farm SA company (Spain) and the Guidelines of the European Union Council (86/609/EU).

Gonad differentiation: gene expression

Gene expression of the *fshr* gene along with several marker genes for the initial stages of gonadal differentiation were evaluated through qPCR on five male and five female gonads at the five developmental stages considered: 84, 98 and 126 dpf, juveniles and adults. Stages were chosen considering histological data on *S. senegalensis* (this study; Viñas et al., 2012) and

previous information in *S. maximus*, a species with similar gonadal developmental pattern (Robledo et al., 2015). RNA extraction was performed using the RNeasy mini kit (Qiagen) with DNase treatment and RNA quality and quantity were evaluated in a Bioanalyser (Bonsai Technologies) and in a NanoDrop® ND-1000 spectrophotometer (NanoDrop® Technologies Inc), respectively. Primers for qPCR of the candidate gene (*fshr*, see Results), and for ovary (*cyp19a1*, aromatase) and testes (*amh*, anti-müllerian hormone; *sox9*, SRY-Box Transcription Factor 9) markers (Robledo et al., 2015), along with those related to germinal cell proliferation (*gsdf,* gonadal soma-derived factor, and *vasa*, ATP-dependent RNA helicase) were designed using the Primer 3 software using the annotation of the *S. senegalensis* genome here presented. Reactions were performed using a qPCR Master Mix Plus for SYBR Green I No ROX (Eurogenetec) following the manufacturer instructions, and qPCR was carried out on a MX3005P (Agilent Technologies). Analyses were performed using the MxPro software (Agilent). The ribosomal protein S4 (*rps4*) and L17 (*rpl17*) genes, and ubiquitin (*ubq*), previously validated for qPCR in turbot gonads by Robledo et al. (2014), were used as reference genes. Two technical replicates were included for each sample. The ΔΔCT method (Kubista et al., 2007) was used to estimate gene expression; briefly, Ct values were normalized using the reference genes, log transformed and finally mean centered to obtain mean centered fold change values which were used for statistical analysis. An unpaired Student's t-test was used to determine significant differences between sex at each stage.

Additionally, data from an ongoing RNA-Seq study on gonad differentiation was used to ~~identify~~ validate diagnostic SNPs on the exons of the SD candidate *fshr* gene and especially, to estimate gene expression of the X-linked and Y-linked alleles (see Results). Samples of total RNA were delivered to Novogene-Europe, UK, for constructing poly-A enriched mRNA to be sequenced in 150-bp pair-end using an Illumina NovaSeq 6000 platform. Raw RNA sequencing

reads were filtered using fastp v.0.19.7 (Chen et al., 2018), trimming bases with Phred quality <15 and reads with length <30 bp, and then each sample was aligned independently against the *S. senegalensis* genome using STAR v.2.7.9a (Dobin et al. 2013) using default parameters. SNPs were identified using mpileup and the variant calling command of bcftools (Li 2011). The resulting vcf file was used to obtain read counts corresponding to each SNP. Then, for each SNP (0: reference allele, 1: alternative allele in the *S. senegalensis* genome), homozygous females (0/0 or 1/1) and heterozygous males (0/1) were identified using a read counting ≥ 8 for consistent genotyping to avoid misclassification of heterozygotes as false homozygotes. Using genotyping across all exons, females (0/0, 1/1) and males (0/1) were consistently identified, confirming the gonadal sex, when gonads were histologically differentiated or when the genetic sex was obtained using the molecular tool. Only those individuals with consistent genotyping at ≥ 33% diagnostic SNPs of *fshr* gene were considered for further analysis. Then, the number of reads of the X-linked allele (0 when females were 0/0 or 1, when females were 1/1) and of the Y-linked allele (the opposite: 1 when females were 0/0, or 0 when females were 1/1) were counted in each male for all the stages evaluated. Reads of the X-linked and of the Y-linked alleles were pooled across all exons for each individual and normalized per million reads to be compared at each stage and across all stages. Non-parametric paired samples Wilcoxon tests were used to compare X-linked and Y-linked normalized counts across all stages and at each stage.

Protein structure modelling on X- and Y-linked allelic variants of fshr: comparison with other Pleuronectiformes

We evaluated 3D protein structure models to infer potential functional differences between X and Y linked allelic variants of FSHR. To find potential template structures for modeling, a specific PSI-BLAST sequence search in the PDB was performed

(https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990). The identified template structure showed unresolved regions which encompass non-synonymous mutations analysed in the present study. Two different strategies for modeling were undertaken: I-TASSER (Zhang 2008) and RoseTTAfold (Baek et al., 2021). I-TASSER is a metaserver that automatically employs ten threading algorithms in combination with *ab initio* modelling to build the tertiary structure of a protein as well as replica-exchange Monte Carlo dynamics simulations for the atomic-level refinement. For comparison, a deep learning-based modelling method, RoseTTAFold (https://robetta.bakerlab.org), was also applied. The putative presence of intrinsically disordered regions in the proteins was investigated by the following predictors: PONDR (Romero et al., 2001), DISOPRED (Jones & Cozzetto, 2015), IUPRED3 (Erdös et al., 2021) and PrDOS (Ishida & Kinoshita, 2007).

To compare the evolutionary rate of the X-linked and Y-linked alleles of the follicule stimulating hormone receptor (*fshr*) gene (SD candidate, see Results) from the putative undifferentiated ancestor, the sequences of their encoded protein variants were compared to those from other flatfish with confident annotation and chromosome-level assemblies available in public databases. We took information of the orthogroup corresponding to *fshr* from the analysis performed by de la Herrán et al. (2022) on *S. senegalensis*, *Hippoglossus hippoglossus*, *H. stenolepis*, *Paralychthys olivaceous*, *S. maximus* and *C. semilaevis*, using *D. rerio* as outgroup. The protein sequence corresponding to *S. senegalensis* in that analysis was replaced by the protein sequences encoded by the X-linked and Y-linked alleles and then, a phylogenetic tree was constructed after whole sequence alignment using Clustal W2 (Larkin et al., 2007). Phylogenetic reconstructions were performed using the function "build" of ETE3 3.1.2 (Huerta-Cepas et al., 2016) and a ML tree was inferred using PhyML v20160115, where branch supports were computed out of 100 bootstrapped trees (Guindon et al., 2010).

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Results

Genome assembly and annotation

The initial assembly comprised 82 contigs with an N50 of 23.4 Mb (sizes ranging from 0.3 to 30.1 Mb) for a total assembly size of 614 Mb (Table S1). A high-density genetic map was used to place 51 contigs, representing 98.9% of the whole assembly (607.9 Mb), into the 21 chromosomes of the *S. senegalensis* haploid karyotype (n = 21; Vega et al., 2002) (see below). This assembly is highly contiguous (contig N50: 23.4 Mb, scaffold N50: 29.0 Mb) and displays high consensus quality (QV = 43.17, which corresponds with a sequence accuracy of 99.995%), and gene (98.4% Complete BUSCO genes) and k-mer (98.18%) completeness (Table S1). Consistent with the low percentage of duplicated BUSCOs (1.0%), the k-mer spectra (Fig. S1) did not reveal any evidence of artificial duplications. The repetitive peak observed at 360x very likely corresponds to true long repetitive regions captured by the nanopore reads (Fig. S1).

Repetitive sequences made up to 8.2% of the *S. senegalensis* genome (Table S2). These were constituted of three main categories: simple repeats (2.8%), low-complexity motifs (0.3%), and transposable elements (TEs) (4.7%). The TE-derived fraction was very similar to that found in other high-quality flatfish genome assemblies, 5.8% in *C. semilaevis* (Chen et al., 2014) and 5.0% *in S. maximus* (Figueras et al., 2016), respectively. The *S. senegalensis* genome displayed a higher TE proportion than *T. nigroviridis* and *Fugu rubripes* (< 3%), but much lower than that observed in other fish such as *D. rerio* (> 40%) (Gao et al., 2016).

In total, 24,264 protein-coding genes producing 40,511 transcripts (1.67 transcripts per gene) and encoding for 37,259 unique protein products were annotated (Table S3). We were able to assign functional labels to 85% of the annotated proteins. The annotated transcripts contained 12.79 exons on average, with 95.5% of them being multi-exonic. The median length of the

protein-coding genes present in this annotation was 7,566 bp, value that is consistent with the annotation of the *S. senegalensis* genome assembly published by Guerrero-Cózar et al (7,368 bp). In addition, 52,888 candidate non-coding RNAs were annotated (6,871 long-non-coding and 46,017 small non-coding RNAs).

Genetic map construction and genome scaffolding

2b-RAD-seq was used for genotyping three full-sib families, consisting of 81 (Fam1), 77 (Fam2) and 71 offspring (Fam3) and the six correspondent parents (Table S4). The *gstacks* module rendered a total of 156,981 loci, 35,441 of them containing at least one single nucleotide polymorphism (SNP), and 16,890, 15,457 and 16,715 SNPs were retained for map construction in Fam1, Fam2 and Fam3, respectively. These represent 29,126 unique SNPs (47.5 SNPs per Mb), with 4889 SNPs shared among the three families.

Separate male and female genetic maps were built in each family. A LOD score of 9.0 was applied to match with the n = 21 chromosomes of the species. The average number of markers per LG across all maps was 434 (range: 93 to 694) (Tables S5 and S6). Female maps were slightly longer than male maps (average F *vs* M recombination ratio of 1.1:1). Shared markers across families were used to build the female, male, and species consensus maps. The final species consensus map included 28,838 markers across 21 LGs spanning 40,704.47 cM (Fig. S2). a length exceeding that expected based on genome size. This artifact elongation is related with the limitations of the software to accommodate such number of markers in a consensus framework from several individual maps, as previously reported (Maroso et al., 2018).

Using Chromonomer, we anchored and oriented 51 of the initial 82 genome contigs to the 21 LGs of the consensus map (Table S7). Those 51 contigs assembled represented 98.9% of the 614 Mb of the genome placed into the 21 chromosomes of the *S. senegalensis* karyotype. Only one contig was split into two fragments assigned to LG6 and LG9 (Fig. 1). After this anchoring

516 step supported by the genetic map, the new genome significantly improved upon the 90.0%
517 reported by Guerrero-Cózar et al. (2021) and it was similar to other flatfish genomes recently
518 assembled (Table S8). The new assembly showed a one-to-one correspondence at the
519 chromosome level with the previous version of Guerrero-Cózar et al. (2021) (Fig. S3; Table
520 S5). However, the substantial fragmentation of the previous version (1937 scaffolds *vs* 82
521 contigs) gave rise to discrepancies related to wrong orientation of many minor contigs across
522 most chromosomes.

523 *Cytogenetic map and mapping integration*

524 A total of 141 BACs were used to anchor the LGs / scaffolds to the chromosomes of the *S.*
525 *senegalensis* karyotype using previous BAC-FISH information (Table S9; Fig. S4). On average
526 6.6 BACs per scaffold were used to establish the correspondence between the genetic, physical
527 and cytogenetic maps (range: 4 to 14 BACs), thus providing a robust mapping integration (Fig.
528 1). Ten out of 141 BACs were localized in more than one chromosome or in different locations
529 within the same chromosome suggesting paralogous regions (Table S9). The minor 5S rDNA
530 was located on C6 and C11, while signals of the major rDNA (18S + ITS1 + 5.8) were found
531 at C6 and C20.

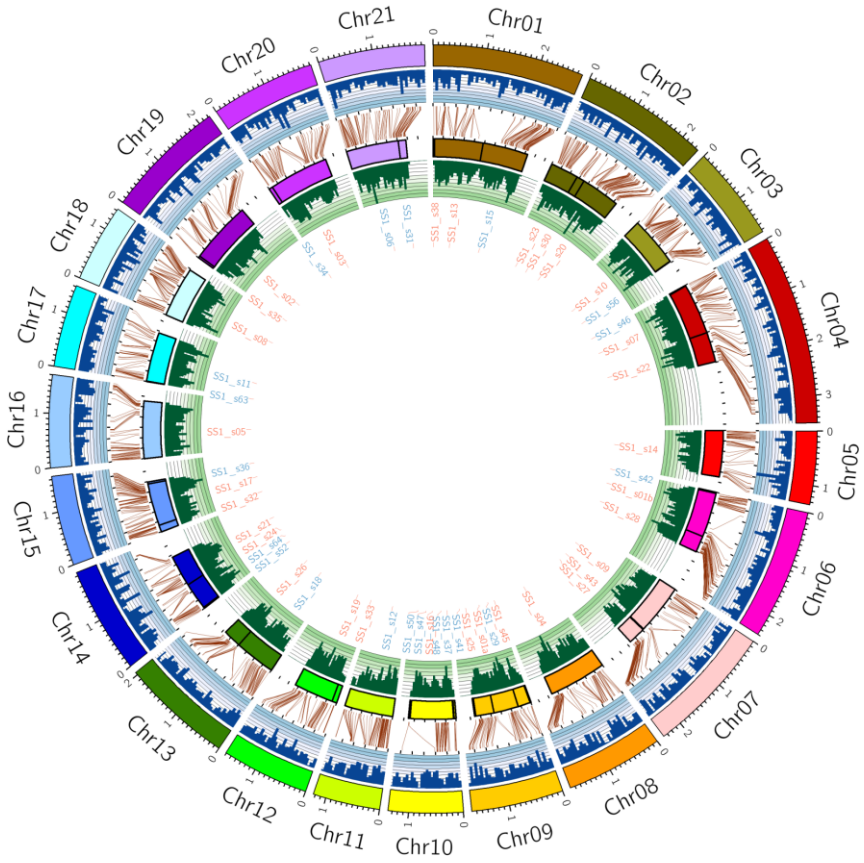

Figure 1: Circos plot of the genome map and anchored scaffolds in *S. senegalensis*. From outer to inner circles are represented: the 21 LGs/chromosomes (tick marks every 100 cM); histograms of the number of markers per 50 cM (in dark blue); brown lines anchoring the genome scaffolds through collinear markers in the genetic map; the 51 anchored contigs (tick marks every 5 Mb); histograms of the number of markers per Mb (in dark green); and the names of the scaffolds (in red those anchored in the reverse strand).

Sex determining (SD) gene candidate

Whole-genome resequencing of six males and females identified a total of 9,078,413 SNPs. A consistent pattern of genetic differentiation between males and females (average F_{ST} = 0.304) and significant heterozygote excess (average F_{IS} = - 0.519) was detected at chromosome (C) 12, between 10,024,822 and 10,054,590 bp (Table S10; Fig. 2).

This region included the 14 exons of the *fshr* gene and a small fragment of the 5' end (~8012 bp) of neurexin, a long gene (~227,983 bp) involved in neuron synaptic connection (Fig. 2). Out of 284 SNPs within the gene, 168 were heterozygous in males and homozygous in females (sex diagnostic markers) consistent with the XX / XY system reported for this species (Molina-Luzón et al., 2015) (Table S11). A total of 33 diagnostic variants were located within exons of *fshr* (24 non-synonymous), 16 of them in exon 14 (11 non-synonymous) and 5 in exon 1 (3 non-synonymous). The accumulation of non-synonymous variants might suggest the degeneration of the Y-linked *fshr* allele that could represent a non-functional variant, although, as shown below, this is an active allele involved on the SD mechanism of Senegalese sole.

A molecular tool for sexing

Three diagnostic SNPs surrounded by ± 100 bp conserved regions (no polymorphisms) were selected for assessing genetic sex using a SnaPshot assay. Three sets of three primers (two external and one internal) were designed (Table S12) and tested in five males and five females. One marker differentiated males and females and matched the *in silico* genotype expectations, and was further validated in 48 males and 48 females from the broodstock of a *S. senegalensis* farm (SS-sex marker). In all but three fishes, the genetic sex matched the phenotypic sex. These three fish were phenotypic males sexed as genetic females. Nonetheless, among a total of 134 fish with known genetic and histological sex information in this study (see Material and Methods section), only these three males showed a discordant genotype (2.2%).

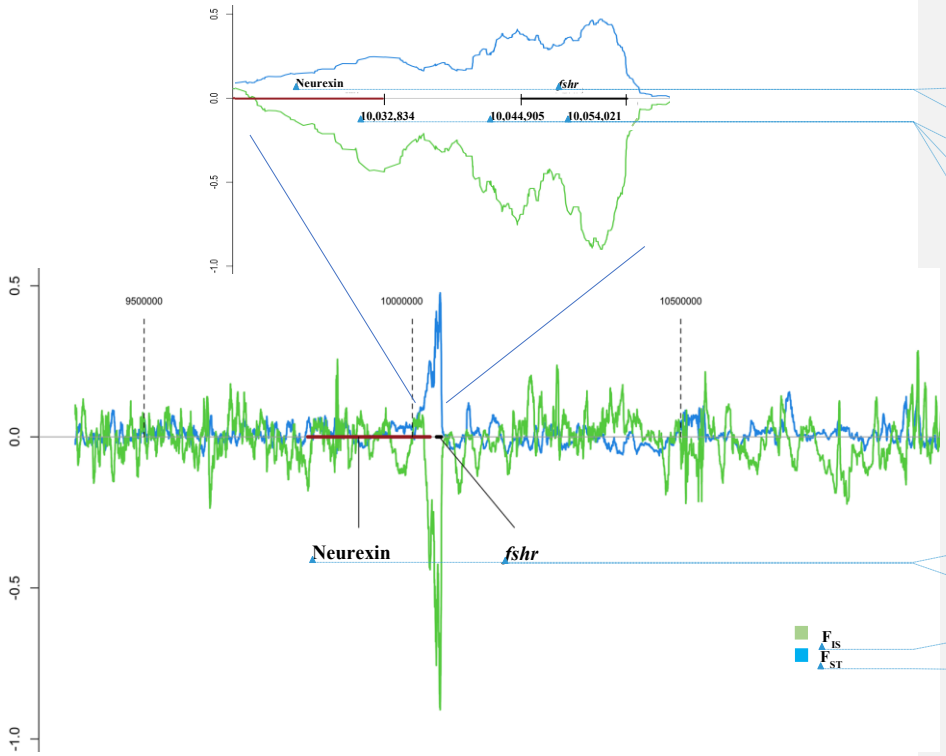

Figure 2: Genetic differentiation (F_{ST}) and intrapopulation fixation index (F_{IS}) between male and female populations at a section of C12 of *S. senegalensis* (between 9,372 and 11,072 kb; coordinates showed with dashed lines). The zoom corresponds to the highest differentiated region (10,024,822- 10,054,590 bp) including the full follicle stimulating hormone receptor gene (*fshr*; coordinates: 10,044-10,054 kb) and a small fragment of the neurexin gene.

Gonad differentiation: gene expression analysis

We analysed gene expression of *fshr* and other relevant genes for gonadal differentiation (*amh*, *cyp19a1*, *gsdf*, *vasa*, and *sox9a*), from the undifferentiated primordium until mature adults, including 84, 98 and 126 dpf, juveniles (315 dpf) and adults (810 dpf), using five males and

Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman
Formatted: Font: (Default) Times New Roman

five females at each stage, sexed either macroscopically or with the SS-sex marker (undifferentiated stages). Gonads were macroscopically identified at all stages (Fig. S5) and histological observations were in accordance with previous information by Viñas et al. (2013) (Figs. S6, S7, and S8).

Interestingly, the *fshr* gene was significantly overexpressed in males at all stages, despite the putative degeneration suggested for the Y-linked allele, especially in juveniles, but even at the undifferentiated 84 and 98 dpf stages (Fig. 3), an observation corroborated by RNA-Seq data (Fig. S9). RNA-Seq was also used to check the expression of the X-linked and the Y-linked *fshr* alleles, taking advantage of the presence of diagnostic SNPs associated to each variant. Only one male from 84 dpf, did not pass the filtering criteria (≥ 8 reads per SNP and ≥ 33% genotyped exons) and was excluded for the analysis (Table S13). Interestingly, the Y-linked showed higher expression than the X-linked allele across all stages (paired samples Wilcoxon test; $P = 0$) and at each stage (paired samples Wilcoxon tests; $P < 0.05$), even at 84 dpf (Y-linked: 19.853 vs X-linked: 13.944, total normalized read count), although not significant at this stage ($P = 0.144$) (Fig. 4; Table S13). In adults of both sexes *fshr* expression was nearly undetectable. Normalized read counts across the 14 exons of *fshr* showed a very similar profile both in males and in females, and no signs of alternative splicing were detected (Fig. S10).

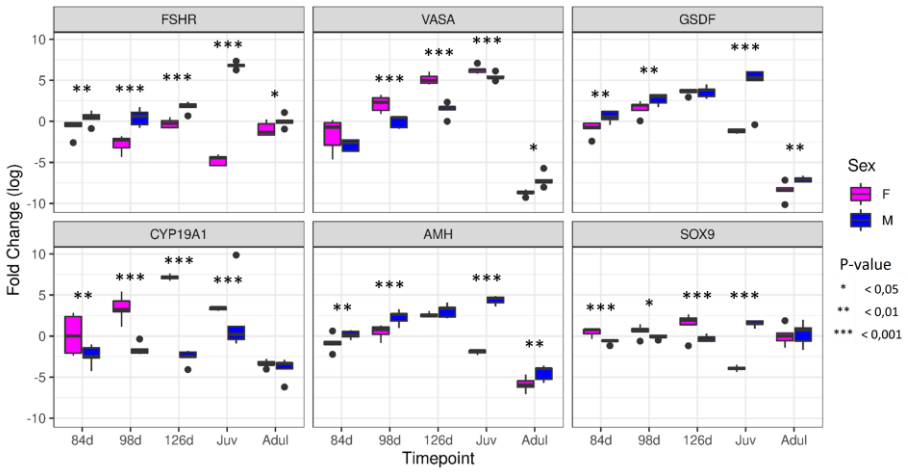

Figure 3: Box plots of the qPCR for the follicule stimulating hormone receptor SD gene of *S. senegalensis* and other key marker genes across gonad development. T-tests were performed to check for statistical differences between sex at each stage.

Additionally, marker genes of male (*sox9a* and *amh*) and female (*cyp19a1a*) gonadal differentiation and of germinal cell proliferation (*gsdf* and *vasa*) were evaluated by qPCR (Fig. 3). While adult male and females hardly showed differential expression for the genes evaluated, juveniles showed the greatest differentiation for most of them. A progressive expression increase was observed for sex marker genes, *cyp19a1a* in females and *amh* in males, concomitant with the two germinal cell proliferation markers, *vasa* in females and *gsdf* in males. However, *sox9a*, a marker of testis differentiation, showed higher expression in females, although not as markedly as for *cyp19a*, until the juvenile stage, when the pattern was reversed.

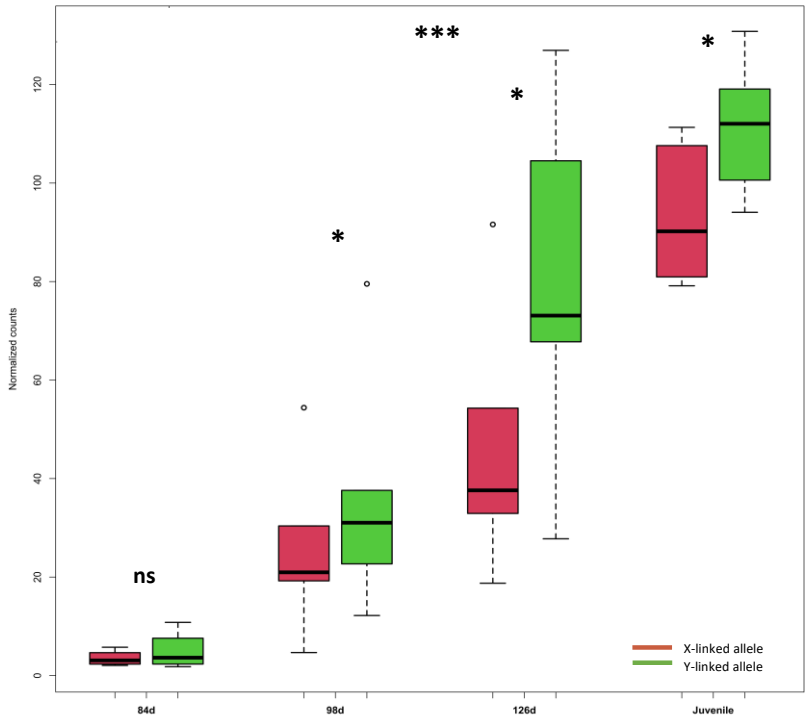

Figure 4: Boxplots of normalized counts (per million reads) for the X-linked and Y-linked allelic variants of the *fshr* gene across different gonad developmental stages in males of *S. senegalensis*. Five individuals were evaluated at each stage, but only those passing the filtering genotyping criteria were considered. ns: not significant; * P < 0.05; *** P < 0.001 (top, across all stages).

3D structure and evolutionary rate of the X-linked and Y-linked encoding follicule stimulating hormone receptor

The model structures obtained by RoseTTAfold were of higher quality than those generated by I-TASSER (Table S14). The models of the follicule stimulating hormone (FSH) receptor

showed differences at key sites of the protein involved in the reception of the hormone and the G-coupled signal transduction domains (Fig. 5). In particular, differences between the X-linked and Y-linked protein variants were found in the three main external domains related to the hormone binding, hairpin loop and hinge region domains (Ulloa-Aguirre et al., 2018), suggesting that the proper binding of the FSH could be ~~hampered~~affected. Model structures highlighted the differences at the flexible hairpin loop with eight variants falling within the LRR extracellular domain (Fig. 5A) that might alter the recognition properties of the receptor (Fig. 5B). Furthermore, the E326Q variant might affect the backbone conformation, between residues 326-336 located at the C-terminal of the hinge region which behaves as an internal agonist unit for the receptor (Ulloa-Aguirre et al., 2018). In the internal side, related to signal transduction, the R365P replacement at IL1, adjacent to the conserved K367, L368 and F373 residues, might influence the interaction with the adapter protein APPL1; the I439L variant, close to the ERW motif, might affect the activation of the G-protein (Ulloa-Aguirre et al., 2018), and aa replacements at IL3 (N529T, I530V and P540H) might affect the stabilization of the inactive conformation.

Furthermore, the X-linked variant of *S. senegalensis* showed higher homology with the FSHR proteins of other flatfish species (Fig. S11), which indicates the divergence of the Y-linked allele from the putative ancestor, as shown by the multiple aminoacid substitutions detected on the encoded protein. All in all, results suggest that the Y-linked could be a non-functional allele but expressed over the X-linked variant in males, although we cannot exclude that a neofunctionalization process is involved.

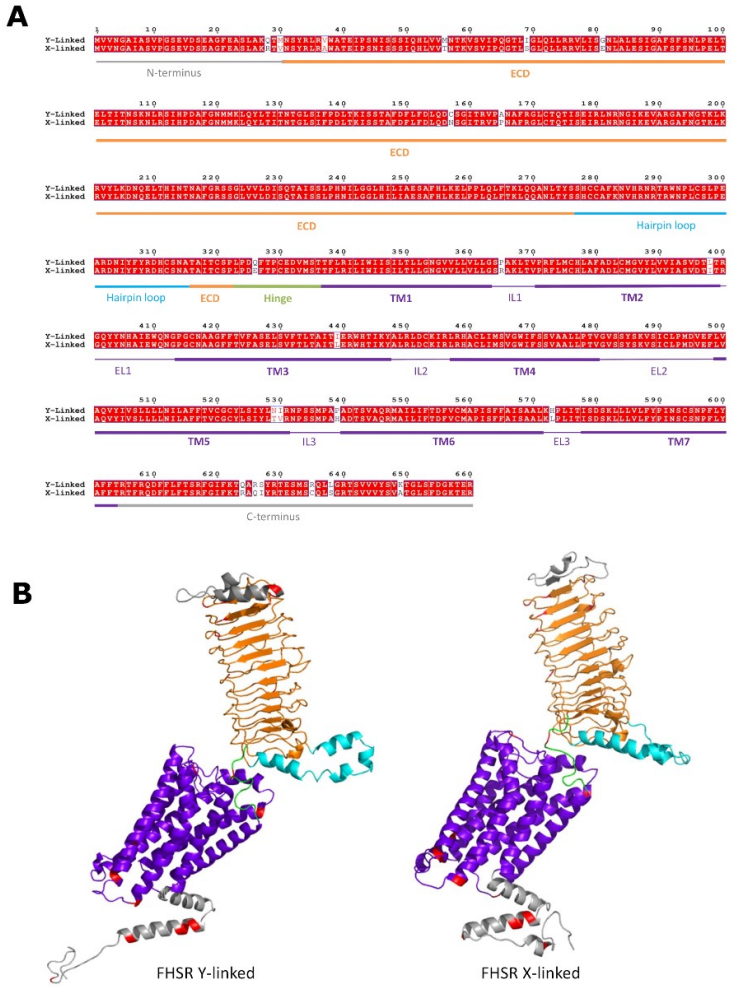

Figure 5: Molecular modeling of the follicule stimulating hormone receptor (FSHR) encoded by X-linked and Y-linked alleles of the sex determining *fshr* gene of *S. senegalensis*. (A) Structure-based sequence alignment of the two variants. Different domains and regions of the receptors are also indicated (Ulloa-Aguirre et al., 2018): ECD, leucine-rich repeat (LRR) extracellular domain; TM, transmembrane helix; EL, extracellular loop; IL, intracellular loop. (B) 3D model structures represented as coloured ribbons. Colour code is the same as for panel A: grey, N-terminus and C-terminus; orange, ECD; light blue, hairpin loop; green, hinge; purple blue, TMD; Non-conserved residues between Y- and X-linked variants are shown in red.

Discussion

A contiguous S. senegalensis genome assembly

The new genome assembly of the *S. senegalensis* is amongst the most contiguous fish genomes assembled to date (Ramos and Antunes 2022) and meets the standards outlined by the Earth Biogenome Project initiative (https://www.earthbiogenome.org/assembly-standards). The contiguity achieved (82 contigs; N50: 23.4 Mb) was facilitated by the small proportion of low complexity sequences; repetitive elements only constitute 8.2% of the *S. senegalensis* genome, a similar proportion to that reported in other flatfish species (Chen et al., 2014; Figueras et al., 2016). The new *S. senegalensis* assembly comprises 21 scaffolds corresponding to the n = 21 chromosomes of its haploid karyotype (Vega et al., 2002) and contains 98.9% of the whole assembly, similar to other flatfish genomes recently reported (Einfeldt et al., 2021; Lü et al., 2021; Martínez et al., 2021; Ferchaud et al., 2022; Jasonowicz et al., 2022) and notably improved the previous version reported by Guerrero-Cózar et al. (2021; 90.0% assembled into chromosomes). These 21 scaffolds were associated to the corresponding chromosomes of this species using previous BAC-FISH information (Ramírez-Torres et al., 2022), thus providing a sound reference for comparative genomics and for studying the genetic architecture of relevant traits, such as sex determination. The new assembly displayed one to one chromosome correspondence with the recently reported genome by Guerrero-Cózar et al. (2021), but ~~notable~~ important discordances were detected across all chromosomes mostly related to orientation of minor scaffolds due to the higher fragmentation of the previous assembly (1937 scaffolds~~; 90.0% of the total length anchored to chromosomes~~). This upgraded *S. senegalensis* genome allowed the identification and annotation of 24,264 protein-coding genes encoding for 37,259 unique protein products (85% with functional annotation), along with 52,888 non-coding RNAs (6871 long non-coding and 46,017 small RNAs).

The master SD gene of S. senegalensis

Guerrero-Cózar et al. (2021) reported a broad genomic region (1.4 Mb) associated with sex at SseLG18 of *S. senegalensis* (Sse C12 in our study), compatible with a nascent XX / XY system. They suggested a SD mechanism with incomplete penetrance and proposed *fshr* as the most promising master SD gene among the tens of genes existing int that region (43 in our genome), considering its role in gonad development and previous information~~data~~ ~~from~~ on SD in the flathead gr~~e~~~~a~~y mullet (Curzon et al., 2021). These authors prompted for more refined association studies and functional information to validate this gene. Here, we showed that *fshr* was the only ~~full~~ complete gene within the region of maximum association and demonstrated the presence of a Y-linked *fshr* allelic variant (*fshry*) in *S. senegalensis*, compatible with the proposed XX / XY SD system and congruent with phenotypic sex in 97.8% of the individuals analysed. Genetic differentiation between males and females (29 kb) included only the full *fshr* gene (~9 kb), although a small fragment of the 5' end of neurexin gene, involved in neuron synapsis, was detected at the least associated region. Our results strongly support a SD system mainly driven by a single gene, unlike the previous report by Guerrero-Cózar et al. (2021), who suggested a genetic SD system with incomplete penetrance involving significant environmental influence. The existence of minor genetic or environmental factors influencing SD has been usually reported, even in species with sound genetic SD systems (Martínez et al., 2014).

The *fshry* coding region showed a high number of sex diagnostic variants (33), heterozygous in males and homozygous in females, twenty-four of them representing non-synonymous substitutions, which suggests a divergent function or a non-functional allele. This hormone receptor is a G protein-coupled with seven transmembrane domains linked to an adenyl cyclase for intracellular transduction of the signal (Levavi-Sivan et al., 2010; Ulloa-Aguirre et al., 2018). Interestingly, many non-synonymous variants were ~~mostly~~ located at exon 14 and exon

1, which are part of the intracellular and extracellular ~~regions~~ domains ~~and are~~ related to signal transduction and hormone reception, respectively. In fact, 3D structure modelling supports significant differences between both FSHR isoforms at several key extracellular (hormone receptor, hairpin loop and hinge) and intracellular (internal loops, C-tail) domains. Also, the phylogenetic analysis performed within Pleuronectiformes using *D. rerio* as outgroup showed a ~~higher~~ divergence of the Y-linked variant from the putative ancestor, associated with the non-synonymous substitutions detected with respect to the X-linked variant, closer to other flatfish. Nonetheless, *fshry* was not only expressed at all stages of gonadal development, but also at a higher level than the X-linked allele (*fshrx*). Moreover, we could not detect evidence of a duplication of the *fshr* gene in the male *S. senegalensis* genome assembled by Guerrero-Cózar et al. (2021), which could suggest an extra copy of *fshr* gene on the Y chromosome taking over the process of SD in the species, as reported in other fish (Matsuda et al., 2002; Hattori et al., 2012; see for review Martínez et al., 2014)~~and~~. ~~therefore~~Consequently, the *fshry* allele appears to be the key factor for the male fate of the undifferentiated primordium. A similar SD mechanism has been reported in flathead gr~~e~~ay mullet, where a *fshry* allele including only two non-synonymous variants was suggested ~~as~~ to be responsible ~~for~~ of the male fate of the undifferentiated primordium (Curzon et al., 2021). Considering the 24 non-synonymous variants detected, it can be speculated that the mechanism driven by the *fshry* allele might be at a more advanced evolutionary stage in *S. senegalensis* than in flathead gr~~e~~ay mullet, and in fact, Ferrareso et al. (2021) reported interpopulation variation of the *fshry* in flathead gr~~e~~ay mullet, suggesting an incomplete penetrance and a recent evolutionary origin of the SD gene.

The expression of *fshr* measured by both qPCR and RNA-SeqRNA-Seq was consistently higher in males than in females starting at 84 dpf (undifferentiated gonads), and this difference increased progressively until the juvenile stage. In females, expression was very low at most

stages, with the exception of 98 dpf and 126 dpf, and in adults, where expression was nearly undetectable both in males and females. The higher expression of *fshr* in males has also been reported in other fish species like clownfish (Kobayashi et al., 2017) and flathead grey mullet (Curzon et al., 2021), and besides, knocking-out the *fshr* gene produced female to male sex reversal in zebrafish and medaka (Murozumi et al., 2014; Zhang et al., 2015). Furthermore, *fshr* has been suggested as the masculinisation transducer of cortisol via suppression of germinal cell proliferation in response to high temperature during the sex determination period in medaka (Hayashi et al., 2010).

The analysis of *fshr* and other key genes at the initial stages of gonad differentiation in *S. senegalensis* revealed a progressive differential upregulation of *fshr*, *amh* and *gsdf* in males, while *cyp19a1* and *vasa* displayed a similar pattern in females. It should be noted that the expression of *amh*, a marker gene of testis development, also documented as master SD gene in several fish species (or its receptor *amhr*), has been reported to be down-regulated by the follicle stimulating hormone (Sambroni et al., 2013), so impairing its function could activate the testis pathway.

All in all, our data strongly support *fshr* as the master SD gene of *S. senegalensis* and suggest that the *fshry* allele could be hampering the action of the FSH, driving the undifferentiated gonad toward testis, potentially by avoiding the suppression of *amh* activity. We hypothesize that the presence of 11 and 3 non-synonymous variants in the intracellular and extracellular domains of the receptor, respectively, should affect the transduction signal of *fshr*, thus ~~affecting~~ blocking the FSH signalling. Also, the FSHRy protein could alter trafficking of the FSH receptors in the endoplasmic reticulum impeding their final integration in the cell membrane, thus hampering FSH function, as reported in humans (Zariñán et al., 2010; Ulloa-Aguirre et al., 2018). Alternatively, neofunctionalization of the *fshry* could be another

784 explanation, but additional data would be necessary to understand how it could be involved in
785 the testis triggering pathway.

786 *Diversification of the SD gene in Pleuronectiformes*

787 Consistent chromosome orthology has been recently reported within flatfishes taking advantage
788 of their chromosome-level genome assemblies facilitated by their compact genomes (Lü et al.,
789 2021; Martínez et al., 2021; Jasonowicz et al., 2022; de la Herrán et al., 2022). According to
790 this information, the SD gene-bearing chromosome of *S. senegalensis*, Sse C12, would not be
791 orthologous to any other consistently proved SD chromosome in flatfish species. The low
792 differentiated Z and W chromosomes of *C. semilaevis* (tongue sole,Chen et al., 2014), as
793 recently reported when compared to *S. maximus* (turbot, Martínez et al., 2021), matched to a
794 single *S. senegalensis* chromosome, Sse C6, which in turn is syntenic to the SD-Hst9 of *H.*
795 *stenolepis* (Pacific halibut, Drinan et al., 2018) (Fig. S12). This reinforces the orthology of sex
796 chromosomes of *C. semilaevis* and *H. stenolepis*, not being conserved in other flatfish (Martínez
797 et al., 2021), although the *dmrt1* SD gene of *C. semilaevis* (Chen et al., 2014) is different from
798 the candidate SD gene recently suggested for *H. stenolepis*, *bmpr1ba* pertaining to the TGF-β
799 family (Jasonowicz et al., 2022). Moreover, conserved synteny with another *S. senegalensis*
800 autosome (Sse C5) was observed for the *H. hippoglossus* (Atlantic halibut) SD chromosome,
801 Hhi12, where *gsdf* has been identified as the most likely master gene for the XX / XY SD
802 system in this species, recently derived from an ancestral ZZ/ZW in the sister *Hippoglossus*
803 species, as recently reported (Edvardsen et al., 2022; Einfeldt et al., 2021). Interestingly, Sse
804 C12 is orthologous to the *S. maximus* Sma C18, where a minor SD-QTL was identified and
805 where sex associated markers were detected in the brill (*S. rhombus*), a congeneric XX / XY
806 species of turbot (Taboada et al., 2014). Moreover, Sse C12 is syntenic to the chromosome 21
807 of *Reinhardtius hippoglossoides* (Greenland halibut), where *sox9a* was suggested as a possible

SD candidate gene of this XX / XY species, although other genes, such as *gdf6* and *sox2* in chromosome 10, were also suggested as potential candidates (Ferchaud et al., 2022). The latter gene, *sox2*, would point to a similar SD system to that reported in *S. maximus* (Martínez et al., 2021). The syntenic relationships between flatfish adds evidence to the huge heterogeneity of SD systems in Pleuronectiformes, even between closely related species (Drinan et al., 2018; Einfeldt et al., 2021; Martínez et al., 2021; Ferchaud et al., 2022), but also highlights the independent recruitment of major SD drivers from common gene families (e.g., *fshr*, *dmrt*, *amh*, *gsdf* or *sox*) across different fish and vertebrate species (Martínez et al., 2014; Guiguen et al., 2019).

Conclusions

The chromosome-level *S. senegalensis* genome assembly reported in this study is among the most contiguous fish assemblies to date. Its integration with previous resources has generated a robust genomic framework for future studies in this important commercial species. This new high-quality assembly enabled the identification of a very consistent SD gene for the species, the follicle stimulating hormone receptor (*fshr*), a new SD gene reported for the first time in Pleuronectiformes. Our hypothesis, supported by functional data, is that the Y-linked variant (*fshry*), through some unknown mechanism, reduces FSH signalling, impeding the down-regulation of *amh* thus driving the undifferentiated gonad toward testis, although we cannot exclude a putative neofunctionalization of this allelic variant. This information made possible to validate a molecular tool for sexing, very useful for production and for management of wild populations of *S. senegalensis*.

Acknowledgements

This study was supported by the Spanish Ministry of Economy and Competitiveness, FEDER Grants (RTI2018-096847-B-C21, RTI2018-096847-B-C22 and RTI2018-097110-B-C21), Junta de Andalucía-FEDER Grant (P20-00938) and the European Union's Horizon 2020 research and innovation programme under grant agreement No 81792 (AQUA-FAANG). We thank Geneaqua SL for their participation and financial support on sequencing. We acknowledge the bioinformatic support of the Centro de Supercomputación de Galicia (CESGA).

Data accessibility

All sequences generated for genome annotation, genetic map construction, and gonadal differentiation have been uploaded to ENA (PRJEB47818) and NCBI (PRJNA820527) Bioprojects detailed by sample in the Supplemental Table Metadata, which additionally includes biological and analysis information for all samples in this study. Genotypes of offspring and parentals of the three families are provided as a Supplemental Table named "Genotypes Mapping families".

Benefit-Sharing section

Benefits from this research accrue from the sharing of our data and results on public databases as described above

Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt, Bold
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: Bold

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403-410. (https://doi.org/10.1016/S0022-2836(05)80360-2)

Arias-Pérez, A., Ramírez, D., Rodríguez, M.E., Portela-Bens, S., García, E., Merlo, M.A., García-Angulo, A., Cross, I., Liehr, T., & Rebordinos L. (2018). In silico detection and

FISH analysis to determine location of miRNAs in *Solea senegalensis* chromosomes using BACs. *OBM Genetics, 2,* 044. (https://doi.org/10.21926/obm.gemmsenet.1804044)

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee G.R., Wang, J., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science, 373,* 871-876. (https://doi.org/10.1126/science.abj8754)

Bao L., Tian C., Liu, S., Zhang, Y., Elaswad, A., Yuan, Z., Khalil, K., Sun, F., Yang, Y., Zhou, T., Li, N., Tan, S., Zeng, Q., Liu, Y., Li, Y., Li, Y., Gao, D., Dunham, R., Davis, K., Waldbieser, G., & Liu, Z. (2019) . The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. *BMC Biology, 17*, 6. (https://doi.org/10.1186/s12915-019-0627-7)

Bao, W., Kojima, K.K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA, 6,* 11. (https://doi.org/10.1186/s13100-015-0041-9)

Buchfink, B., Reuter, K., & Drost, H.G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods, 18*, 366-368. (https://doi.org/10.1038/s41592-021-01101-x)

Cabral, H.N. (2000). Comparative feeding ecology of sympatric *Solea solea* and *Solea senegalensis*, within nursery areas of the Tagus estuary, Portugal. *Journal of Fish Biology, 57*, 1550-1562. (https://doi.org/10.1111/j.1095-8649.2000.tb02231.x)

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. *Mol Ecol* **22**: 3124–3140.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W.A. (2020). Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3- Genes Genom Genet* **10**: 4115–4128.

Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J.N., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics, 46,* 253–260. (https://doi.org/10.1038/ng.2890)

Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*, i884–i890. (https://doi.org/10.1093/bioinformatics/bty560)

Cioffi, M.B., Yano, C.F., Sember, A., & Bertollo, L.A.C. (2017). Chromosomal evolution in lower vertebrates: Sex chromosomes in Neotropical fishes. *Genes, 8,* 258. (https://doi.org/10.3390/genes8100258)

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21,* 3674-3676. (https://doi.org/10.1093/bioinformatics/bti610)

Cross, I., Merlo, A., Manchado, M., Infante, C., Cañavate, J., & Rebordinos, L. (2008). Cytogenetic characterization of the sole *Solea senegalensis* (Teleostei: Pleuronectiformes: Soleidae): Ag-NOR, (GATA)n, (TTAGGG)n and ribosomal genes by

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

one-color and two-color FISH. *Genetica*, *128*, 253-259. (doi: 10.1007/s10709-005-5928-9)

Cui, X., Lu, Z., Wang, S., Wang, J.J.-Y., & Gao, X. (2016). CMsearch: simultaneous exploration of protein sequence space and structure space improves not only protein homology detection but also protein structure prediction. *Bioinformatics, 32,* i332-i340. (https://doi.org/10.1093/bioinformatics/btw271)

Curzon, A.Y., Dor, L., Shirak, A., Meiri-Ashkenazi, I., Rosenfeld, H., Ron, M., & Seroussi, E. (2021). A novel c.1759T>G variant in follicle-stimulating hormone-receptor gene is concordant with male determination in the flathead grey mullet (*Mugil cephalus*). *G3-Genes Genomes Genetics, 11.* jkaa044. (https://doi.org/10.1093/g3journal/jkaa044)

de la Herrán, R., Hermida, M., Rubiolo, J., Gómez-Garrido, J., Cruz, F., Robles, F., Navajas-Pérez, R., Blanco, A., Villamayor, P.R., et al. (2022). A chromosome-level genome assembly enables the identification of the follicle stimulating hormone receptor as the master sex determining gene in *Solea senegalensis*. *bioRvix* (https://doi.org/10.1101/2022.03.02.482245)

Díaz-Ferguson, E., Cross, I., Barrios, M., Pino, A., Castro, J., Bouza, C., Martínez, P., Rebordinos, L. (2012). Genetic characterization, based on microsatellite loci, of *Solea senegalensis* (Soleidae, Pleuronectiformes) in Atlantic coast populations of the SW Iberian Peninsul. *Ciencias Marinas*, *38*, 129–142

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T.R. (2013). STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics, 29,* 15-21. (https://doi.org/10.1093/bioinformatics/bts635)

Drinan, D.P., Loher, T., & Hauser, L. (2018). Identification of genomic regions associated with sex in Pacific halibut. *Journal of Heredity, 109,* 326–332. (https://doi.org/10.1093/jhered/esx102)

Edvardsen, R.B., Wallerman, O., Furmanek, T., Kleppe, L., Jern, P., Wallberg, A., Kjærner-Semb, E., Mæhle, S., Olausson, S.K., Sundström, E., Harboe, T., Mangor-Jensen, R., Møgster, M., Perrichon, P., Norberg, B., Rubin, C.J. (2022). Heterochiasmy and the establishment of gsdf as a novel sex determining gene in Atlantic halibut. *PLoS Genetics*, *8*, 18, e1010011. doi: 10.1371/journal.pgen.1010011.

Einfeldt, A.L., Kess, T., Messmer, A., Duffy, S., Wringe, B.F., Fisher, J., den Heyer, C., Bradbury, I.R., Ruzzante, D.E., & Bentzen, P. (2021). Chromosome level reference of Atlantic halibut *Hippoglossus hippoglossus* provides insight into the evolution of sexual determination systems. *Molecular Ecology Resources, 21,* 1686–1696. (https://doi.org/10.1111/1755-0998.13369)

Erdös, G., Pajkos, M., & Dosztányi, Z. (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Research, 49,* W297-W303. (https://doi.org/10.1093/nar/gkab408)

Ferchaud, A.L., Mérot, C., Normandeau, E., Ragoussis, J., Babin, C., Djambazian, H., Bérubé, P., Audet, C., Treble, M., Walkusz, W., & Bernatchez, L. (2022). Chromosome-level assembly reveals a putative Y-autosomal fusion in the sex determination system of the

Greenland Halibut (*Reinhardtius hippoglossoides*). *G3- Genes Genomes Genetics, 12,* jkab376. (https://doi.org/10.1093/g3journal/jkab376)

Feron, R., Zahm, M., Cabaum C., Klopp, C., Roques, C., Bouchez, O., Eché, C., Valière, S., Donnadieu, C., Haffray, P., et al. (2020). Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Molecular Ecology Resources, 20,* 531–543. (https://doi.org/10.1111/1755-0998.13133)

Ferraresso, S., Bargelloni, L., Babbucci, M., Cannas, R., Follesa, M.C., Carugati, L., Melis, R., Cau, A., Koutrakis, M., Sapounidis, A., Crosetti, D., & Patarnello, T. (2021). *Fshr*: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations. *iScience,* 24, 101886. (https://doi.org/10.1016/j.isci.2020.101886)

Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J.A., Gómez-Garrido, J., Carreté, L., Bello, X., Gut, M., et al. (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Research, 23,* 181–192. (https://doi.org/10.1093/dnares/dsw007)

Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., & Song, C. (2016). The contribution of transposable elements to size variations between four teleost genomes. *Mobile DNA, 7,* 4. (https://doi.org/10.1186/s13100-016-0059-7)

García, E., Cross, I., Portela-Bens, S., Rodríguez, M., García-Angulo, A., Molina, B., Cuadrado, A., Liehr, T., & Rebordinos, L. (2019). Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a Rex transposon located in a proto-sex chromosome. *Scientific Reports*, *9*, 17146. (doi: 10.1038/s41598-019-53673-6)

García-Angulo, A., Merlo, M., Portela-Bens, S., Rodríguez, M., García, E., Al-Rikabi, A., Liehr, T., & Rebordinos, L. (2018). Evidence for a Robertsonian fusion in *Solea senegalensis* (Kaup, 1858) revealed by zoo-FISH and comparative genome analysis. *BMC Genomics 19*, 818. (doi:10.1186/s12864-018-5216-6)

García-Angulo, A., Merlo, M., Rodríguez, M., Portela-Bens, S., Liehr, T., & Rebordinos, L. (2019). Genome and Phylogenetic Analysis of Genes Involved in the Immune System of *Solea senegalensis* – Potential Applications in Aquaculture. *Frontiers in Genetics*, *11*, 529. (doi: 10.3389/fgene.2019.00529)

García-Angulo, A., Merlo, M., Iziga, R., Rodríguez, M., Portela-Bens, S., Al-Rikabi, A., Liehr, T., & Rebordinos, L. (2020). Gene clusters related to metamorphosis in *Solea senegalensis* are highly conserved. 2020. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *35*, 100706. (doi: 10.1016/j.cbd.2020.100706)

García-Cegarra, A., Merlo, M.A., Ponce, M., Portela-Bens, S., Cross, I., Manchado, M., & Rebordinos, L. (2013). A preliminary genetic map in *Solea senegalensis* (pleuronectiformes, soleidae) using BAC-FISH and next-generation sequencing. *Cytogenetic and Genome Research, 14,* 227-240. (https://doi.org/10.1159/000355001)

Guerrero-Cózar, I., Gomez-Garrido, J., Berbel, C., Martinez-Blanch, J.F., Alioto, T., Claros, M.G., Gagnaire, P.A., & Manchado, M. (2021). Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish. *Scientific Reports, 11,* 13460. (https://doi.org/10.1038/s41598-021-92601-5)

Formatted: Spanish (Spain)

Formatted: Font: (Default) Times New Roman

Formatted: Spanish (Spain)

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*. *59*, 307-321. (doi: 10.1093/sysbio/syq010)

Guiguen, Y., Fostier, A., & Herpin, A. (2019). Sex determination and differentiation in fish: genetic, genomic, and endocrine aspects. In *Sex control in aquaculture* (eds. Wang, H.-P., Piferrer, F., Chen, S.-L., Shen, Z.-G.), Vol. 1 of, pp. 35–63, John Wiley & Sons Ltd.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., & Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biolo1gy, 9: R7. (https://doi.org/10.1186/gb-2008-9-1-r7)

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., & Chen, J. (2020). *RIdeogram*: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science, 6,* e251. (https://doi.org/10.7717/peerj-cs.251

Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., & Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences USA, 109,* 2955–2959. (https://doi.org/10.1073/pnas.1018392109)

Hayashi, Y., Kobira, H., Yamaguchi, T., Shiraishi, E., Yazawa, T., Hirai, T., Kamei, Y., & Kitano, T. (2010). High temperature causes masculinization of genetically female medaka by elevation of cortisol. *Molecular Reproduction and Development, 77,* 679–686. (https://doi.org/10.1002/mrd.21203)

Herpin, A., Schartl, M., Depincé, A., Guiguen, Y., Bobe, J., Hua-Van, A., Hayman, E.S., Octavera, A., Yoshizaki, G., Nichols, K.M., Goetz, G.W., & Luckenbach, J.A. (2021). Allelic diversification after transposable element exaptation promoted *gsdf* as the master sex determining gene of sablefish. *Genome Research, 31,* 1366–1381. (https://doi.org/10.1101/gr.274266.120)

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics, 36*, 2253–2255. (https://doi.org/10.1093/bioinformatics/btz891)

Huerta-Cepas, J., Serra, F., Bork, P. (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*. *33*,1635-8. (doi: 10.1093/molbev/msw046)

Ishida, T., & Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research, 35,* W460-464. (https://doi.org/10.1093/nar/gkm363)

Jasonowicz, A.J., Simeon, A.E., Zahm, M., Cabau, C., Klopp, C., Roques, C., Lampietro, C., Lluch, J., Donnadieu, C., Parrinello, H., Drinan, D., Hauser, L., Guiguen, Y., & Planas, J.V. (2022). Generation of a chromosome-level genome assembly for Pacific halibut (*Hippoglossus stenolepis*) and characterization of its sex-determining region. *Molecular Ecology Resources* (https://doi.org/10.1111/1755-0998.13641)

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., López, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30,* 1236-1240. (https://doi.org/10.1093/bioinformatics/btu031)

Jones, D.T., & Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics, 31, 857-863. (https://doi.org/10.1093/bioinformatics/btu744)

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., & Kikuchi, K. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics, 8,* e1002798. (https://doi.org/10.1371/journal.pgen.1002798).

Kobayashi, Y., Nozu, R., & Nakamura, M. (2017). Expression and localization of two gonadotropin receptors in gonads of the yellowtail clownfish, *Amphiprion clarkii*. *Journal of Aquaculture & Marine Biology, 5,* 00120. (https://doi.org/10.15406/jamb.2017.05.00120)

Koyama, T., Nakamoto, M., Morishima, K., Yamashita, R., Yamashita, T., Sasaki, K., Kuruma, Y., Mizuno, N., Suzuki, M., Okada, Y., Ieda, R., Uchino, T., Tasumi, S., Hosoya, S., Uno, S., Koyama, J., Toyoda, A., Kikuchi, K., & Sakamoto, T. (2019). A SNP in a steroidogenic enzyme is associated with phenotypic sex in *Seriola* fishes. *Current Biology, 29,* 1901-1909.e8. (https://doi.org/10.1016/j.cub.2019.04.069)

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., & Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research, 19,* 1639–1645. (https://doi.org/10.1101/gr.092759.109)

Kubista, M., Sindelka, R., Tichopad, A., Bergkvist, A., Lindh, D., Forootan, A. (2007). The prime technique. Real-time PCR data analysis. *GIT Laboratory Journal*, *9-10*, 33–5.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10,* R25. (https://doi.org/10.1186/gb-2009-10-3-r25)

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007). Clustal W and Clustal X version 2.0, *Bioinformatics*, *23*, 2947–2948. (https://doi.org/10.1093/bioinformatics/btm404)

Levavi-Sivan, B., Bogerd, J., Mañanós, E.L., & Gómez A, Lareyre, J.J. (2010). Perspectives on fish gonadotropins and their receptors. *General and Comparative Endocrinology, 165,* 412–437. (https://doi.org/10.1016/j.ygcen.2009.07.019)

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, *25*, 1754-60. (https://doi.org/10.1093/bioinformatics/btp324)

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25,* 2078-2079. (https://doi.org/10.1093/bioinformatics/btp352)

Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*, 2987-93. (https://doi.org/10.1093/bioinformatics/btr509)

Lomsadze, A., Burns, P.D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research, 42*, e119. (https://doi.org/10.1093/nar/gku557)

Lü, Z., Gong, L., Ren, Y., Chen, Y., Wang, Z., Liu, L., Li, H., Chen, X., Li, Z., Luo, H., et al. (2021). Large-scale sequencing of flatfish genomes provides insights into the polyphyletic origin of their specialized body plan. *Nature Genetics, 53,* 742–751. (https://doi.org/10.1038/s41588-021-00836-9)

Luckenbach, J.A., Borski, R.J., Daniels, H.V., & Godwin, J. (2009). Sex determination in flatfishes: Mechanisms and environmental influences. *Seminars in Cell & Developmental Biology, 20,* 256–263. (https://doi.org/10.1016/j.semcdb.2008.12.002)

Maroso, F., Hermida, M., Millán, A., Blanco, A., Saura, M., Fernández, A., Dalla Rovere, G., Bargelloni, L., Cabaleiro, S., Villanueva, B., et al. (2018). 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 Research, 25,* 439-450. (https://doi.org/10.1093/dnares/dsy015)

Martín, I., Carazo, I., Rasines, I., Rodríguez, C., Fernández, R., Martínez, P., Norambuena, F., Chereguini, O., & Duncan, N. (2019). Reproductive performance of captive Senegalese sole, *Solea senegalensis*, according to the origin (wild or cultured) and gender. *Spanish Journal of Agricultural Research, 17,* e0608. (https://doi.org/10.5424/sjar/2019174-14953).

Martínez, P., Viñas, A.M., Sánchez, L., Díaz, N., Ribas, L., & Piferrer, F. (2014)., Robledo D, Taboada X, Blanco A, Moser M, Maroso F, Hermida M, Gómez-Tato A, Álvarez-Blázquez B, Cabaleiro S, et al. 2021. A genome-wide association study, supported by a new chromosome-level genome assembly, suggests sox2 as a main driver of the undifferentiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). *Genomics* **113**: 1705–1718.

Martínez, P., Viñas, A.M., Sánchez, L., Díaz, N., Ribas, L., & Piferrer F. 2014. Genetic architecture of sex determination in fish: Applications to sex ratio control in aquaculture. *Frontiers in Genetics* **5**: 340. (doi:10.3389/fgene.2014.00340).

Martínez P., Robledo D., Taboada X., Blanco A., Moser M., Maroso F., Hermida, M., Gómez-Tato, A., Álvarez-Blázquez, B., Cabaleiro, S., Piferrer, F., Bouza, C., Lien, S., & Viñas, A.M. (2021). A genome-wide association study, supported by a new chromosome-level genome assembly, suggests sox2 as a main driver of the undifferentiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). *Genomics*, *113*, 1705-1718. doi.org/10.1016/j.ygeno.2021.04.007

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., et al. (2002). *DMY* is a Y-specific DM-

domain gene required for male development in the medaka fish. *Nature, 417,* 559–563. (https://doi.org/10.1038/nature751)

Merlo, M., Iziga, R., Portela-Bens, S., Cross, I., Kosyakova, N., Liehr, T., Manchado, M., & Rebordinos, L. (2017). Analysis of the histone cluster in *Senegalese sole* (*Solea senegalensis*): Evidence for a divergent evolution of two canonical histone clusters. *Genome 60*, 441-453. (doi.org/10.1139/gen-2016-0143)

Merlo, M., Portela-Bens, S., Rodríguez, M., García-Angulo, A., Cross, I., Arias-Pérez, A., García, E., & Rebordinos L. (2021). A Comprehensive Integrated Genetic Map of the Complete Karyotype of *Solea senegalensis* (Kaup 1858). *Genes*, *12*, 49. (doi: 10.3390/genes12010049)

Molina-Luzón, M.J., López, J.R., Robles, F., Navajas-Pérez, R., Ruiz-Rejón, C., De la Herrán, R., & Navas, J.I. (2015). Chromosomal manipulation in Senegalese sole (*Solea senegalensis* Kaup, 1858): induction of triploidy and gynogenesis. *Journal of Applied Genetics, 56,* 77-84. (https://doi.org/10.1007/s13353-014-0233-x)

Morais, S., Aragão, C., Cabrita, E., Conceição, L. E., Constenla, M., Costas, B., ... & Dinis, M. T. (2016). New developments and biological insights into the farming of Solea senegalensis reinforcing its aquaculture potential. *Reviews in Aquaculture*, *8*(3), 227-263. (doi: https://doi.org/10.1111/raq.12091)

Murozumi, N., Nakashima, R., Hirai, T., Kamei, Y., Ishikawa-Fujiwara, T., Todo, T., & Kitano, T. (2014). Loss of follicle-stimulating hormone receptor function causes masculinization and suppression of ovarian development in genetically female medaka. *Endocrinology, 155,* 3136–3145. (https://doi.org/10.1210/en.2013-2060)

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics, 191,* 163–170. (https://doi.org/ 10.1534/genetics.111.137497)

Nacif, C. L., Kratochwil, C. F., Kautt, A. F., Nater, A., Machado-Schiaffino, G., Meyer, A., & Henning, F. (2022). Molecular parallelism in the evolution of a master sex-determining role for the anti-Mullerian hormone receptor 2 gene (amhr2) in Midas cichlids. Molecular Ecology.

Nakamoto, M., Uchino, T., Koshimizu, E., Kuchiishi, Y., Sekiguchi, R., Wang, L., Sudo, R., Endo, M., Guiguen, Y., Schartl, M., Postlethwait, J.H., & Sakamoto, T. (2021). A Y-linked anti-Müllerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLoS Genetics, 17,* e1009705. (https://doi.org/10.1371/journal.pgen.1009705)

Nawrocki, E.P., & Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics, 29,* 2933-2935. (https://doi.org/10.1093/bioinformatics/btt509)

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., & Finn, R.D. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research, 43,* D130-137. (https://doi.org/10.1093/nar/gku1063)

Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busne, J.M., Bobe, J., Concordet. J.P., et al. (2019). Identification of the master sex determining gene

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genetics, 15,* e1008013. (https://doi.org/10.1371/journal.pgen.1008013)

Parra, G., Blanco, E., & Guigo, R. (2000). GeneID in *Drosophila*. *Genome Research, 10,* 511-515. (https://doi.org/10.1101/gr.10.4.511.)

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., & Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-Seq reads. *Nature Biotechnology, 33*, 290-295. (10.1038/nbt.3122)

Portela-Bens, S., Merlo, M., Rodríguez, M., Cross, I., Manchado, M., Kosyakova, N., Liehr, T., & Rebordinos, L. (2017). Integrated gene mapping and synteny studies give insights into the evolution of a sex proto-chromosome in *Solea senegalensis*. *Chromosoma*, *126*, 261-277. (doi: 10.1007/s00412-016-0589-2)

Ramírez, D., Rodríguez, M.E., Cross, I., Arias-Pérez, A., Merlo, M.A., Anaya, M., Portela-Bens, S., Martínez, P., Robles, F., Ruiz-Rejón, C., & Rebordinos, L. (2022). Integration of Maps Enables a Cytogenomics Analysis of the Complete Karyotype in *Solea senegalensis*. 2022. *International Journal of Molecular Sciences*; *23*, 5353. (doi: 10.3390/ijms23105353)

Ramos, L., & Antunes, A. (2022). Decoding sex: Elucidating sex determination and how high-quality genome assemblies are untangling the evolutionary dynamics of sex chromosomes. *Genomics, 114,* 110277. (https://doi.org/10.1016/j.ygeno.2022.110277)

Ramos-Júdez, S., González-López, W.Á., Ostos, J.H., Mamani, N.C., Alemán, C.M., Beirão, J., & Duncan, N. (2021). Low sperm to egg ratio required for successful in vitro fertilization in a pair-spawning teleost, Senegalese sole (*Solea senegalensis*). *Royal Society Open Science, 8,* 201718. ((https://doi.org/10.1098/rsos.201718)

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity, 86,* 248–249. (https://doi.org/10.1093/oxfordjournals.jhered.a111573)

Rhie, A., Walenz, B.P., Koren, S., & Phillippy, A.M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology, 21,* 245. (https://doi.org/10.1186/s13059-020-02134-9)

Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir A., & Merisov J.P. (2017). Variant review with the Integrative Genomics Viewer (IGV) *Cancer Research*, *77*, 31-34. doi: 10.1158/0008-5472.CAN-17-0337

Robledo, D., Hernández-Urcera, J., Cal, R.M., Pardo, B.G., Sánchez, L., Martínez, P. & Viñas, A. (2014). Analysis of qPCR reference gene stability determination methods and a practical approach for efficiency calculation on a turbot (*Scophthalmus maximus*) gonad dataset. *BMC Genomics, 15,* 648. (https://doi.org/10.1186/1471-2164-15-648)

Robledo, D., Ribas, L., Cal, R., Sánchez, L., Piferrer, F., Martínez, P., & Viñas, A. (2015). Gene expression analysis at the onset of sex differentiation in turbot (*Scophthalmus maximus*). *BMC Genomics, 16,* 973. (https://doi.org/10.1186/s12864-015-2142-8)

Robledo, D., Fernández, C., Hermida, M., Sciara, A., Álvarez-Dios, J.A, Cabaleiro, S., Caamaño, R., Martínez, P., & Bouza, C. (2016). Integrative transcriptome, genome and quantitative trait loci resources identify single nucleotide polymorphisms in candidate

Formatted: Font: (Default) Times New Roman

genes for growth traits in turbot. *International Journal of Molecular Sciences, 17(2),* 243 (https://doi.org/10.3390/ijms17020243).

Robledo, D., Hermida, M., Rubiolo, J.A., Fernández, C., Blanco, A., Bouza, C., & Martínez, P. (2017). Integrating genomic resources of flatfish (Pleuronectiformes) to boost aquaculture production. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 21*, 41–55. (https://doi.org/10.1016/j.cbd.2016.12.001)

Rodríguez, M., Molina, B., Merlo, M., Arias-Pérez, A., Portela-Bens, S., García-Angulo, A., Cross, I., Liehr, T., & Rebordinos, L. (2019). Evolution of the Proto Sex-Chromosome in *Solea senegalensis*. *International Journal of Molecular Sciences*, *20*, 5111. (doi: 10.3390/ijms20205111)

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., & Dunker, A.K. (2001). Sequence complexity of disordered protein. *Proteins, 42,* 38-48. (https://doi.org/10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3)

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology, 132,* 365-386. (https://doi.org/10.1385/1-59259-192-2:365)

Sambroni, E., Lareyre, J.J., & le Gac, F. (2013). Fsh controls gene expression in fish both independently of and through steroid mediation. *PLoS ONE, 8,* e76684. (https://doi.org/10.1371/journal.pone.0076684)

Shao, C., Bao, B., Xie, Z., Chen, X., Li, B., Jia, X., Yao, Q., Ortí, G., Li, W., Li, X., et al. (2017). The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nature Genetics,* 49: 119–124. (https://doi.org/10.1038/ng.3732)

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., & Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31,* 3210–3212. (https://doi.org/10.1093/bioinformatics/btv351)

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

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *The Plant Journal, 3,* 739–744. (https://doi.org/10.1111/j.1365-313X.1993.00739.x)

Stanke, M., Schoffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics, 7,* 62. (https://doi.org/10.1186/1471-2105-7-62)

Taboada, X., Pansonato-Alves, J.C., Foresti, F., Martínez, P., Viñas, A., Pardo, B.G., & Bouza, C. (2014). Consolidation of the genetic and cytogenetic maps of turbot (*Scophthalmus maximus*) using FISH with BAC clones. *Chromosoma, 123,* 281–291. (https://doi.org/10.1007/s00412-014-0452-2)

Takehana, Y., Matsuda, M., Myosho, T., Suster, M.L., Kawakami, K., Shin-I, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., et al. (2014). Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature Communications, 5,* 4157. (https://doi.org/10.1038/ncomms5157)

Ulloa-Aguirre, A., Zariñán, T., Jardón-Valadez, E., Gutiérrez-Sagal, R., & Dias, J.A. (2018). Structure-function relationships of the follicle-stimulating hormone receptor. *Frontiers in Endocrinology (Lausanne), 9,* 707. (https://doi.org/10.3389/fendo.2018.00707)

Vega, L., Díaz, E., Cross, I., & Rebordinos, L. (2002). Caracterizaciones citogenética e isoenzimática del lenguado *Solea senegalensis* Kaup, 1858. *Boletín. Instituto Español de Oceanografía,18,* 245–250.

Viñas, J., Asensio, E., Cañavate, J.P., & Piferrer, F. (2013). Gonadal sex differentiation in the Senegalese sole (*Solea senegalensis*) and first data on the experimental manipulation of its sex ratios. *Aquaculture, 384–387*, 74–81. (https://doi.org/10.1016/j.aquaculture.2012.12.012)

Voorrips, R.E. (2002). Mapchart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity, 93,* 77–78. (https://doi.org/10.1093/jhered/93.1.77)

Wang, L., Sun, F., Wan, Z.Y., Yang, Z., Tay, Y.X., Lee, M., Ye, B., Wen, Y., Meng, Z., Fan, B., et al. (2022). Transposon-induced epigenetic silencing in the X chromosome as a novel form of *dmrt1* expression regulation during sex determination in the fighting fish. *BMC Biology, 20,* 5. (https://doi.org/10.1186/s12915-021-01205-y)

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

Wen, M., Pan, Q., Jouanno, E., Montfort, J., Zahm, M., Cabau, C., Klopp, C., Iampietro, C., Roques, C., Bouchez, O., et al. (2022). An ancient truncated duplication of the anti-Mullerian hormone receptor type 2 geneis apotential conserved master sex determinant in the Pangasiidae catfish family. *Molecular Ecology Resources, 22,* 2411-2428. (https://doi.org/10.1111/1755-0998.13620)

Wu, Y., Close, T.J., & Lonardi, S. (2011). Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8, 381–394. (https://doi.org/10.1109/TCBB.2010.35)

Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (*sdY*) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications, 6,* 486–496. (https://doi.org/10.1111/eva.12032)

Zariñán, T., Perez-Solís, M.A., Maya-Núñez, G., Casas-González, P., Conn, P.M., Dias, J.A., & Ulloa-Aguirre, A. (2010). Dominant negative effects of human follicle-stimulating hormone receptor expression-deficient mutants on wild-type receptor cell surface expression. Rescue of oligomerization-dependent defective receptor expression by using cognate decoys. *Molecular and Cellular Endocrinology, 321,* 112–122. (https://doi.org/10.1016/j.mce.2010.02.027)

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Informatics, 9,* 40. (https://doi.org/10.1186/1471-2105-9-40)

Zhang, Z., Lau, S.W., Zhang, L., & Ge, W. (2015). Disruption of zebrafish follicle-stimulating hormone receptor (*fshr*) but not luteinizing hormone receptor (*lhcgr*) gene by TALEN leads to failed follicle activation in females followed by sexual reversal to males. *Endocrinology (United States), 156,* 3747–3762. (https://doi.org/10.1210/en.2015-1039)

Formatted: Font: (Default) Times New Roman

Zheng, S., Tao, W., Yang, H., Kocher, T. D., Wang, Z., Peng, Z., Jin, L., Pu, D., Zhang, Y., & Wang, D. (2022). Identification of sex chromosome and sex-determining gene of southern catfish (*Silurus meridionalis*) based on XX, XY and YY genome sequencing. *Proceedings of the Royal Society B*, *289*, 20212645. (https://doi.org/10.1098/rspb.2021.2645)

Formatted: Font: (Default) Times New Roman

Data Accessibility and Benefit-Sharing

Sequence data generated in this study have been uploaded to ENA and NCBI under accession number bioprojects PRJEB47818 and PRJNA820527, respectively (details by sample are shown in Supplemental Table Metadata).

Author's Contributions

RH: identification and characterization of the SD gene, and molecular tool for sexing
MH: responsible of genetic map construction and genome scaffolding
JR: responsible of orthology and phylogenetic on candidate genesin flatfish
JGG: genome annotation and collaboration in genome assembly
FC: genome assembly
FR: qPCR analyses on gonad differentiation
RNP: repetitive DNA analysis
AB: bioinformatics support across multiple tasks
PRV: sampling, anatomy and histology of gonads
DT: sampling, anatomy and histology of gonads
PSQ: supervision of anatomy and histology work
DRA: integration of cytogenetic, genetic and physical maps
MER: integration of cytogenetic, genetic and physical maps
AA-P: integration of cytogenetic, genetic and physical maps
IC: Anchoring BACs to scaffolds
ND: production of full-sib families for genetic map construction
TMP: collaboration in sampling at Stolt Sea Farm SA
AR: supervision of sampling at Stolt Sea Farm SA
MCDR: 3D protein modelling
DP: 3D protein modelling
AM: planning and design of the study
MG: coordination and supervision of genome sequencing
CB: sex determination comparative genomics in flatfish
DR gene expression analyses (qPCR and RNA-SeqRNA-Seq)
LR: coordination of cytogenetic mapping integration
TA: coordination and supervision of genome assembly and annotation
CRR: planing and design of the study
PM: coordination, planning and design of the study; gene expression analysis
Manuscript writing: PM, MH, DR, CB, TA, JGG, FC, PRV
All authors revised the manuscript and contributed to the final version

Legends Tables and Figures

Figure 1: Circos plot of the genome map and anchored scaffolds in *S. senegalensis*. From outer to inner circles are represented: the 21 LGs/chromosomes (tick marks every 100 cM); histograms of the number of markers per 50 cM (in dark blue); brown lines anchoring the genome scaffolds through collinear markers in the genetic map; the 51 anchored contigs (tick marks every 5 Mb); histograms of the number of markers per Mb (in dark green); and the names of the scaffolds (in red those anchored in the reverse strand).

Figure 2: Genetic differentiation (F_{ST}) and intrapopulation fixation index (F_{IS}) between male and female populations at a section of C12 of *S. senegalensis* (between 9,372 and 11,072 kb). The enlarged region corresponds to the follicle stimulating hormone receptor gene (*fshr*).

Figure 3: Box plots of the qPCR for the follicule stimulating hormone receptor SD gene of *Solea sonegalensis* and other key marker genes across gonad development.

Figure 4: Number of counts for the X-linked and Y-linked allelic variants of the *fshr* gene across different gonad developmental stages in *S. senegalensis*. Five replicates were evaluated at each stage, but only those passing the filtering genotyping criteria were further analysed.

Figure 5: Molecular modeling of the follicule stimulating hormone receptor (FSHR) encoded by X-linked and Y-linked alleles of the sex determining *fshr* gene of *S. senegalensis*. (A) Structure-based sequence alignment of the two variants. Different domains and regions of the receptors are also indicated (Ulloa-Aguirre et al., 2018): ECD, leucine-rich repeat (LRR) extracellular domain; TM, transmembrane helix; EL, extracellular loop; IL, intracellular loop. (B) 3D model structures represented as coloured ribbons. Colour code is the same as for panel

A: grey, N-terminus and C-terminus; orange, ECD; light blue, hairpin loop; green, hinge; purple blue, TMD; Non-conserved residues between Y- and X-linked variants are shown in red.

Supplementary information

Supplementary Tables

Supplemental Table Metadata: Biological and genomic information of all fishes (larvae, postlarvae, fry, juvenile and adults) studied

Table S1: Genome assembly statistics of *S. senegalensis*; fSolSen1_LG: set of contigs anchored to linkage groups (LG) using the genetic map.

Table S2. Statistics of TE-derived sequence and other simple repeats in the genome of *S. senegalensis*.

Table S3: Genome annotation statistics of protein coding genes in the genome of *S. senegalensis*

Table S4: Number of reads and averages obtained by the 2b-RAD-seq method for SNP genotyping in three families (Fam1, Fam2, Fam3) of *S. senegalensis* and their offspring across the different filtering steps of the established pipeline from the initial raw reads to the final valid alignment against the genome for constructing a highly dense genetic map.

Table S5: Statistics of the genetic maps constructed in *S. senegalensis* using three full-sib families. Maps were constructed via male and via female in each family and the consensus per sex and for the whole species were obtained. The correspondence between linkage groups in the consensus map and that reported by Guerrero-Cózar et al. (2021) is also provided. LG codes were arranged from the longest to the shortest within each map, but their correspondence was established according to the codes of the consensus map, which in turn followed the

chromosome number of the karyotype after mapping integration. LG: linkage group; families: F1, F2 and F3; male: M; female: F.

Table S6: Marker positions for all genetic maps constructed in *S. senegalensis* and their integration by sex and species

Table S7: List of anchored scaffolds of the *S. senegalensis* genome on the genetic map indicating the orientation and size.

Table S8. Comparative statistics of the *S. senegalensis* genome with other pleuronectiform chromosome-level assembled genomes.

Table S9: One hundred and forty-one BAC clones used for integrating cytogenetic, genetic and physical maps of *S. senegalensis*. Notice that some clones matched to several regions, either in different or the same chromosome, among them, the 5S rDNA gene clusters.

Table S10: Wright F-statistics for male and female populations per SNP and using 50 SNP-sliding windows averaged over FIS and FST across the contig 19 of the *S. senegalensis* genome pertaining to C12, where the *fshr* gene is located.

Table S11: SNPs localized in the follicule stimulating hormone receptor (*fshr*) gene in six males and six females re-sequenced at 20x coverage using the *S. senegalensis* assembled genome; diagnostic SNPs are homozygous in females and heterozygous in males consistent with a XX / XY sex determining system; in the last row it is indicated the non-synonymous or synonymous condition of aminoacid substitutions; the boundaries of exons are highlighted in bold type.

Table S12: Sets of primers designed with Primer 3 to develop a molecular tool for sexing using diagnostic markers between males (heterozygous) and females (homozygous) of *S. senegalensis* located at exon 13 of the *fshr* gene.

Table S13: Genotypes and allelic counts of diagnostic SNPs located at the *fshr* gene (exons, 5' and 3' UTR, introns; detailed information in Table S11) from gonad RNA-SeqRNA-Seq data of samples of five males (M) and five females (F) collected across gonad development of *S. senegalensis*, from the initial undifferentiated or low differentiated stages (84D, 98D and 126D post fertilization) until juveniles and adults; SNP ID makes reference to the contig and position where they are located in the contig of the genome assembly; REF and ALT alleles refers to the allele in the genome (0) and the alternative allele (1) detected after resequencing six females and six males; Genotypes: homozygous in females (0/0 or 1/1) and heterozygous in males (0/1); ./. missing genotypes because allelic counts did not reach the minimum threshold (8 reads); allelic counts: the first number refers to the "0" allele and the second, separated by semicolon, to "1" allele; Colors: pink (missing genotypes), green (valid genotypes for counting), purple (females); blue (males); red (individuals not considered because they did not reach a minimum genotyping data).

Table S14. Quality assessment of modeled Y-linked and X-linked allelic variants of the *fshr* SD gene of *S. senegalensis*

Supplementary Figures

Figure S1: K-mer distribution on: A) initial Illumina reads, B) the final assembly of *S. senegalensis*.

Figure S2: Mapping marker density across linkage groups in the *Senegalese sole* consensus map

Figure S3: LASTZ plots between scaffolds / chromosomes (this study; ordinates) and pseudo-chromosomes (Guerrero-Cózar et al., 2021; abscises) of *S. senegalensis* genome assemblies.

Figure S4: Idiogram of the *S. senegalensis* chromosomes where position of BACs used for mapping integration is shown.

Figure S5. Macroscopic anatomy and topography of sole gonads of individuals of 126, 98 and 84 dpf. (A, B) 126 dpf female sole; (C) 98 dpf individual - unidentified sex; (D, E) 126 dpf presumably sole male identified by 'lacking a female gonad'. (F) 84 dpf individual with unidentified sex. Insets: site of gonad. Scale bars: 600 µm (A); 400 µm (B-F)

Figure S6. Histological sections of adult and juvenile sole gonads. (A-C) Adult female. (D-F) Juvenile female. (G-I) Adult male. (J-L) Juvenile male. (A-F) (*) Atresia stages; (^) Post-ovulatory follicles; Arrow: Nuclei; Arrowhead: Nucleolus; (1) Oogonia; (2) Early oocyte; (3) Late oocyte. (G-L) White arrows: Show the radial disposition of seminiferous lobules (*) from the central medulla (m) to the cortex (c) and tunica albuginea (ta); (1) and black arrowhead: Spermatids; 2 and black arrow: Spermatozoa; Black square: Interstitial tissue. Stain: Hematoxilin-Eosin (HE). Scale bars: 250 µm (A, D, G); 100 µm (B, C, E, H, J, K); 50 µm (F, I, L).

Figure S7. Histological sections of female gonad of individuals of 126, 98 and 84 dpf. (A) In 126 dpf female sole, previtellogenic oocytes can be visualized; (B, C) Gonad of 98 and 84 dpf females, respectively, both undifferentiated, although at 98 dpf there are more potential-oocyte-cells. (D-F) Higher magnification of A-C, respectively. Gonads of 98 and 84 dpf individuals correspond to females identified with the SS-sex. Stain: HE. Scale bars: 250 µm (A); 100 µm (B); 50 µm (C, D, E, F).

Figure S8: Histological sections of male gonad of individuals of 126, 98 and 84 dpf. (A) 126 dpf; (B) 98 dpf; (C) 84 dpf; (D, E) Higher magnification of A, B, respectively. The three stages are undifferentiated, with the oldest one (126 dpf) showing more potentially-spermatogonia

cells. All samples were genotyped with the SS-sex marker. * Gonad; K: kidney; c: Cartilage. Stain: HE. Scale bars: 100 µm (A, B, C); 50 µm (D, E).

Figure S9: RNA-SeqRNA-Seq data of the *fshr* gene in *S. senegalensis* males and females across different gonad developmental stages (abscissas) using a count log scale (ordinates).

Figure S10: Read distribution across the 14 exons of the *fshr* gene in males and in females of *S. senegalensis* at some stages with a number of counts high enough to assess differences.

Figure S11: Phylogeny of the *fshr* gene in Pleuronectiformes including the Y- and X-linked allelic variants of *S. senegalensis* using *D. rerio* as outgroup. Confidence bootstrapping values of each grouping are shown in brackets at the nodes and genetic distances to the nodes above each branch or in parentheses in the short terminal branches.

Figure S12: Diversification of the sex determinant (SD) systems across Pleuronectiformes.A) Concatenated synteny of SD chromosomes and genes (or markers / MAS) in the six flatfish species studied ([a] Phylogenomic divergence (million years, My) among flatfish families; Lü et al. 2011); B) Comparative mapping between flatfish SD-gene bearing chromosomes.