Screening a new European hake (Merluccius merluccius) chromosome-level genome assembly suggests an XX/XY sex determining system driven by the SRY-box transcription factor 3 (sox3)

Paulino Martínez Portela[1], Laura Casas[2], Natalia Petit-Marty[2], Andres Blanco[3], Maialen Carballeda[1], Nair Vilas[2], Jèssica Gómez Garrido[4], Fernando Cruz[5], Julio Valeiras[6], Tyler Alioto[4], and Fran Saborido-Rey[2]

[1]University of Santiago de Compostela
[2]Instituto de Investigaciones Marinas
[3]Universidade de Santiago de Compostela
[4]CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology
[5]Fundacion Centro Nacional de Analisi Genomico
[6]Instituto Espanol de Oceanografia Centro Oceanografico de Vigo

December 10, 2024

Abstract

Sex determination is exceptionally diverse and shows high evolutionary rate in fish. European hake (Merluccius merluccius) is a species of great commercial value distributed throughout European coasts, which displays a significant sexual growth dimorphism. We present a chromosome-level genome of M. merluccius, assembled into 215 contigs using long- and short-read sequencing, further scaffolded into the species' 21 chromosomes using Omni-C (715 Mb). RNA-Seq on several tissues from pooled individuals improved annotation (26,625 protein-coding genes and 16,766 ncRNAs). Five males and five females from an Atlantic population were re-sequenced at 30x coverage to look for association with sex across the whole genome. Genetic differentiation between males and females (FST) and intrapopulation fixation index (FIS) pointed to a region on chromosome 9 spanning ˜10 Mb which included several genes related to gonad differentiation and showed strong linkage disequilibrium associated to a putative inversion. Near sox3 (˜25 kb), SNPs were mostly heterozygous in males and homozygous in females, consistent with an XX/XY SD system. These SNP markers were validated in a larger sample of 56 males and 65 females from the same population using MassARRAY. Other genomic regions that were differentiated between males and females and suggestive of sexual conflict were also explored across the genome. Results support a candidate master sex determining (MSD) gene in M. merluccius and indicate some differentiated regions potentially under sexual conflict. This information will be useful for the fisheries management of M. merluccius in the context of climate change, where non-invasive sex identification tools are essential.

Introduction

Sex determination (SD) refers to the mechanism controlling the fate of the undifferentiated gonadal primordium during early-life stages, ultimately responsible for the sex of a mature individual. SD was initially associated with chromosomal heteromorphisms due to the highly conserved SD systems studied in *Drosophila* , mammals, and birds. However, increasing information from ectothermic vertebrates has revealed a much

more diverse scenario (Martínez et al., 2014; Guiguen et al., 2019). Chromosomal heteromorphisms associated with sex are rare in fish, even within the highly diverse Neotropical ichthyofauna (Cioffi et al., 2017), and opposite SD systems are not infrequent between congeneric species (Martínez et al., 2014). According to a recent review by Kitano et al. (2024), up to 21 different master sex-determining (MSD) genes have been consistently identified to date across 114 fish species spanning 18 orders. While some genes such as *amh* and its receptors, as well as *dmrt* , have been repeatedly and independently recruited throughout fish evolution, several other MSD genes are species- or taxa-specific, such as *fshr* in *Solea senegalensis* , *hsd17b1* in the genus *Seriola,* or *sdY* in salmonids. Although most MSD genes are associated with transcription factors at the top of the SD cascade or with genes involved in the TGF-β signaling pathway, recent data suggest that steroidogenesis, a crucial pathway in gonad and sex differentiation, has also been recurrently co-opted for MSD recruitment (Kitano et al., 2024). The growing body of knowledge on SD in fish has been facilitated by the development of highly contiguous and reliable chromosome-level genome assemblies used as reference for high throughput SNP genotyping screening facilitated by advances in sequencing technologies and new scaffolding and bioinformatic approaches (Ramos & Antunes, 2022).

To understand the evolutionary dynamics of SD in fish, intraspecific variation should be explored to identify transitions between SD systems (Bachtrog et al., 2014; Martínez et al., 2014). For this approach, the genetic structure and degree of isolation of populations are essential reference points. According to theory, sexual conflict at specific genomic regions could trigger a shift to a new SD system to restore population fitness (Van Doorn & Kirkpatrick, 2007). Ongoing sexual conflicts associated with transitions between SD systems have been shown in cichlids (Roberts et al., 2009), and intrapopulation variation in SD has been reported in fish species such as Northern pike (*Essox lucius* ; Pan et al., 2021) and gray mullet (*Mugil cephalus* ; Ferrareso et al., 2020). Different mechanisms have been shown to underly the emergence of new SD genes in fish, including gene duplication or the presence of alternative alleles at the MSD gene (Kitano et al., 2024). To fully elucidate SD mechanisms and regions under sexual conflict, comprehensive chromosome-level assemblies and genome-wide genotyping data are required (de la Herrán et al., 2023).

The European hake (*Merluccius merluccius*) is widely distributed in the Mediterranean Sea and the Northeast Atlantic Ocean, ranging from the coasts of Norway and Iceland to the Guinea Gulf (Murua, 2010). Its range also extends eastward into the North Sea, Skagerrak and Kattegat, and occasionally into the Black Sea (Casey & Pereiro, 1995; Kutsyn et al., 2024). It is considered one of the most important commercial species in both the Northeast Atlantic and the Mediterranean, where populations have been extensively and intensively harvested, leading to its classification as overexploited (FAO, 2020; Izquierdo et al., 2021; Morales-Nin et al., 2022). The European hake is a gonochoristic species that exhibits significant sexual size dimorphism, with females reaching greater sizes and weights than males (Piñeiro and Sainza, 2003; Domínguez-Petit and Saborido-Rey, 2010; Cerviño, 2014; El Bouzidi et al., 2022; Apostologamvrou et al., 2023). This is the consequence of the differential maturation pattern between sexes influencing growth and natural mortality, which in turn affects population dynamics and stock productivity (Cerviño, 2014). Moreover, due to the intense size-selective fishing pattern, hake stocks experience size truncation, mostly in females, altering growth patterns and maturation schedules, and more importantly inducing adaptive changes beyond mere phenotypic plasticity (Hidalgo et al., 2014) that may differ among stocks. Beyond its importance to fisheries, the European hake has also been identified as a potential target for aquaculture diversification (Iglesias et al., 2010).

In this study, we generated a high-quality and contiguous chromosome-level genome assembly of *M . merluccius* using short-read and long-read sequencing, further scaffolded with Omni-C, to investigate the SD system of the species. Several males and females from an Atlantic population were screened across the whole genome using millions of SNPs to identify regions associated with sex. Detailed analysis of the most significant regions was conducted to identify candidate MSD genes, and the findings were validated in a larger sample of sexed specimens. Furthermore, other genomic regions differentiated between males and females were inspected and validated as potential sexual conflict regions. Results suggest an SD XX/XY system in European hake driven by the SRY-box transcription factor 3 (*sox3*).

Materials and Methods

Sampling

A single *M. merluccius* female from the Spanish Atlantic Ocean (Galician shelf) was used for genome assembly. Additionally, pooled samples from five different tissues (brain, muscle, gonad, spleen and liver) from three individuals of the same population were used for RNA-seq to improve genome annotation. Five males and five females from the same Atlantic population were collected to explore sex association using whole-genome sequencing (WGS). To validate the candidate SD region, 56 males and 65 females from the same Atlantic population were analyzed. Biological and genomic metadata for the *M. merluccius* samples analysed in this study are provided in Supplemental Table Metadata.

DNA and RNA sequencing

Long-read Whole Genome Sequencing

High molecular weight DNA was purified from the spleen of a single female of *M . merluccius* using the Nanobind CBB Big DNA Kit (Circulomics), following the manufacturer´s instructions, and eluted in EB buffer (Qiagen). Sequencing libraries were prepared with the Ligation sequencing kit SQK-LSK109 from Oxford Nanopore Technologies (ONT). Briefly, 4.0 µg of the DNA was repaired and end-repaired using the NEBNext FFPE DNA Repair Mix (NEB) and the NEBNext UltraII End Repair/dA-Tailing Module NEB. Then, sequencing adaptor ligation, purification by 0.4X AMPure XP Beads and elution in Elution Buffer (SQK-LSK109) were performed. Sequencing runs were carried out on a GridION Mk1 (ONT) using a Flowcell R9.4.1 FLO-MIN106D (ONT), with sequencing data collected over 110 hours. The quality parameters of the sequencing runs were monitored in real time by the MinKNOW platform version 4.1.2, and base-calling was performed with Guppy version 4.2.3.

Short-read whole genome sequencing

The short-insert paired-end libraries for whole genome sequencing (WGS) were prepared from the DNA of the same female used for long-read sequencing, following a PCR free protocol with some modifications using the KAPA HyperPrep kit (Roche). In short, 1.0 µg of genomic DNA was sheared on a Covaris LE220-Plus (Covaris) and size-selected for fragment sizes of 220-550 bp with AMPure XP beads (Agencourt, Beckman Coulter). The genomic DNA fragments were then end-repaired and adenylated. Next, compatible adaptors for Illumina platforms with unique dual indexes and 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 using the Kapa Library Quantification Kit for Illumina platforms (Roche).

RNA-Seq

Total RNA extraction was performed with the RNeasy mini kit (Qiagen) with DNase treatment. RNA quantity and quality were evaluated with the Qubit® RNA BR Assay kit (Thermo Fisher Scientific), and RNA integrity was estimated using the RNA 6000 Nano Bioanalyser 2100 Assay (Agilent). Next, equimolar RNA pools from three individuals (brain, liver, muscle, spleen and gonad) were used for library construction for each tissue following quality evaluation of individual RNA extractions.

The RNA-Seq libraries were prepared with the KAPA Stranded mRNA-Seq Illumina® Platforms Kit (Roche) following the manufacturer´s recommendations. Briefly, 500 ng of total RNA was used for poly-A fraction enrichment with oligo-dT magnetic beads, following the mRNA fragmentation. Strand specificity was achieved during second strand synthesis, which was performed in the presence of dUTP instead of dTTP. The blunt-ended double stranded cDNA was 3´adenylated before ligating Illumina platform-compatible adaptors with unique dual indexes and unique molecular identifiers (Integrated DNA Technologies). The ligation product was enriched using 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 a NovaSeq 6000 (Illumina) in paired-end mode, with a read length of 151 bp for WGS and 100 bp for 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), followed by the generation of FASTQ sequence files.

Hi-C sequencing

Hi-C libraries were prepared using the Omni-C kit (Dovetail Genomics) following the manufacturer's protocol. Briefly, three *M .merluccius* (two females and one immature specimen) were dissected and five different organs (brain, muscle, liver, spleen and gonads) were stored in RNAlater and maintained at -80^{o}C until use. The different samples from each organ were pooled and pulverized using a mortar and pestle immersed in a liquid nitrogen bath. Chromatin was crosslinked with formaldehyde (Sigma Aldrich), digested with DNase I and the DNA was extracted. The DNA ends were repaired, and a biotinylated bridge adapter was ligated, followed by proximity ligation of adapter-containing ends. After reverse crosslinking, the DNA was purified and used to prepare Illumina-compatible paired-end sequencing libraries (omitting the fragmentation step). Biotinylated chimeric molecules were isolated using streptavidin beads prior to PCR enrichment of the library. The library was sequenced on a NovaSeq 6000 (Illumina, 2 × 151 bp) following the manufacturer's protocol for dual indexing.

Genome assembly

High molecular weight DNA was purified from the spleen of a single*M . merluccius* female using the Nanobind CBB Big DNA Kit (Circulomics), following the manufacturer´s instructions, and eluted in EB buffer (Qiagen). The genome was assembled using long reads from Oxford Nanopore Technologies (ONT), Illumina paired-end reads to improve base accuracy, and Omni-C contact data to map sequences to chromosomes using the CLAWS1 pipeline (Figure S1).

Prior to assembly, adaptors present in the Illumina data were trimmed with TrimGalore (https://github.com/FelixKrueger/TrimGalore). A k-mer database (k=20) was subsequently built with Meryl (https://github.com/marbl/meryl) using the trimmed short-read data. The k-mer histogram generated by Meryl was used as input to GenomeScope2.0 to visualize the k-mer distribution and to estimate the haploid genome size, heterozygosity and repeat content. The ONT data were filtered with Filtlong3 (–minlen 1000 –min_mean_q 80) prior to assembly to remove short and low-quality reads.

The filtered ONT data were assembled with Nextdenovo v2.5.04 using the 'nano-raw' mode and a minimum overlap of 1000 bp. To improve base accuracy, the assembly was polished with HyPo5, using both Illumina and ONT data. The polished assembly was then purged with purge_dups6 to remove alternate haplotypes and other artificially duplicated repetitive regions.

The resulting assembly was scaffolded using the Omni-C data with YAHS7. Manual curation of the assembly was performed with PretextView (https://github.com/wtsi-hpag/PretextView).

Finally, the Blobtoolkit8 pipeline was applied to the scaffolded assembly (Fig. 1), using the NCBI nt database (updated on November 2022) and several BUSCO odb10 databases, including actinopterygii, vertebrata, metazoa, eukaryota, fungi and bacteria.

Genome annotation

Repeats in the genome assembly were annotated with RepeatMasker v4-1-2 (http://www.repeatmasker.org) using the custom repeat library available for *Danio rerio* . Additionally, a specific repeat library for our assembly was generated with RepeatModeler v1.0.11. Repeats associated to repetitive protein families were excluded by performing a BLAST1 search against the Uniprot database. RepeatMasker was then rerun with this new library to annotate the specific repeats.

Gene annotation of the *M. merluccius* genome assembly was achieved by integrating transcript alignments, protein alignments, and ab initio gene predictions (Figure S2). A flowchart of the annotation process is shown

in Figure ANN1. RNAseq data from five organs (brain, liver, muscle, spleen and gonad) were used for annotation. Reads were aligned to the genome using STAR2 v-2.7.2a, and transcript models were subsequently generated using Stringtie3 v2.2.1 and merged using TACO4 v0.7.3. High quality junctions for annotation were obtained by running ESPRESSO5 v1.3.0 after mapping with STAR. Finally, PASA assemblies were produced with PASA6 v2.5.2, and TransDecoder program, which is part of the PASA package, was used to identify coding regions in the transcripts. Next, the complete proteomes of *Danio rerio* ,*Chanos chanos* and *Carassius auratus* were downloaded from Uniprot (May 2023) and aligned to the genome using Miniprot7 v0.6. Ab initio gene predictions were performed on the repeat-masked assembly with three programs: GeneID8 v1.4, Augustus9 v3.5.0 and Genemark-ET10 v4.71, with and without RNAseq data. Gene predictors were run with human-trained parameters, except for Genemark, which runs in a self-trained mode. Finally, all data were combined into consensus CDS models using EvidenceModeler-1.1.1 (EVM, Haas et al., 2008). Additionally, two rounds of PASA annotation updates were used to annotate UTRs and alternative splicing forms via. Functional annotation of the annotated proteins was performed with Pannzer's11 online server.

The annotation of non-coding RNAs (ncRNAs) was obtained by running the following steps on the repeat masked version of the genome assembly. First, the program cmsearch12 v1.1 that is part of the Infernal13 package was run against the RFAM database of RNA families v12.0. Additionally, tRNAscan-SE14 v2.08 was used to detect transfer RNA genes present in the genome assembly. Long non-coding RNAs (lncRNAs) were identified by first filtering the set of PASA-assemblies that were not included in the protein-coding gene annotation, retaining only those transcripts longer than 200bp and not covered by more than 80% by a small ncRNA. The resulting transcripts were clustered into genes using shared splice sites or significant sequence overlap as criteria for designation as the same gene.

Comparison of M. merluccius genome assemblies

M . merluccius assembly from this study (female; origin: Galicia, NW Spain), was aligned using minimap2 (Li et al., 2018) against the chromosome-level assembly recently released (PRJEB77069; September, 2024; male; origin: North Sea, Oslofjord, Norway). The result PAF format files (Pairwise mApping Format) were plotting by ggplot2 (Wickham, 2016) and pafr (Winter, 2020) R packages.

MSD gene candidates

DNA extracted from fin clip tissue of five adult males and five adult females from an Atlantic population was sequenced using 150 bp PE reads on an Illumina NovaSeq 6000 System to 30x coverage in the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain) Platform, following the outlined short-read WGS protocol. The high throughput SNP screening of the genome was aimed at identifying sex associated regions to be further validated using 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. Each sample was then aligned independently against the newly assembled *M. merluccius* reference genome using Burrows-Wheeler Aligner v.0.7.17 (Li and Durbin, 2009), with default parameters.

A large SNP dataset was identified and genotyped for five males and five 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 intrapopulation (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 20 consecutive SNPs and explored using sliding windows across each chromosome to look for deviation from the null hypothesis (F_{ST} and $F_{IS} = 0$). Expected F_{ST} and F_{IS} values for regions consistent with an XX/XY or a ZZ/ZW SD system would be 0.5 and – 1, respectively, when partitioning total genetic diversity within and between sexes (Wright, 1951). Given that we averaged these values over 20 SNPs for the sliding windows, we established more relaxed thresholds to identify candidate regions (F_{ST} [?] 0.4 and with F_{IS} [?] - 0.5), considering that not all SNPs within a window would show the same genetic diversity, and, particularly, the expected sex-association pattern. Our strategy, successfully implemented in other species (Martinez et al., 2021; de la Herran et al., 2023), involves a double check to identify the SD region, based not only on genetic differentiation between male and female populations (as in

the pooled DNA strategy), but also on the expected heterozygote excess in the region associated with the heterogametic sex.

Validation of the SD region

To validate the most promising candidate sex-associated genomic regions, we expanded our sample to 60 males and 60 females from the same Atlantic population. DNA was extracted from fin clips stored in pure ethanol using the DNeasy Blood & Tissue Kit (Qiagen) following manufacturer's instructions. A MassARRAY assay was used for SNP genotyping, utilizing a set of markers located in the candidate regions (selecting at least two markers per region showing the closest genotyping pattern consistent with the disclosed XX/XY SD system; see Results). SNP selection for each region was based on technical feasibility, specifically the absence of other polymorphisms within +-100 bp of the SNP, as indicated by the WGS data from the initial five males and five females. The MassARRAY assay consists of two consecutive reactions; the first step involves the PCR amplification of the region containing the target SNP and the second step is a mini-sequencing reaction using an internal 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, taking the flanking sequences from our assembled genome. Primer design and genotyping was done at the UCIM-University of Valencia Genomics Platform, using MALDI-TOF mass spectrometry analysis on an Autoflex spectrometer.

Results

Genome assembly

Table 1: Statistics of the genome assembly of *M* .

merluccius

Assembly	**Nextdenovo**	**Nextdenovo + hypo**	**Nextdenovo + hypo + purged**	**fMerMel2.1**
Contig N50	10,991,627 bp	10,936,441 bp	10,936,441 bp	10,936,441 bp
Scaffold N50	10,991,627 bp	10,936,441 bp	10,936,441 bp	32,021,586 bp
Scaffold L50	21	21	21	10
Total sequences	245	245	215	62
Total length	723,776,638 bp	720,771,172 bp	715,974,550 bp	715,430,492 bp
BUSCO[c]	95.2%	96.1%	96.0%	96.0%
BUSCO[d]	1.2%	1.3%	1.1%	1%
QV	35	41	42	42
Kmer completeness	89.38	90.33	90.27	90.25

BUSCO v5 actinopterygii_odb10 database: [c] complete;[d] duplicated

The initial assembly obtained with Nextdenovo spanned 724 Mb (245 scaffolds). This assembly was scaffolded using the Omni-C data (Fig. S3), and after manual curation, a total of 62 scaffolds were obtained that were assembled in the chromosome-level assembly (716 Mb; 21 chromosomes) (Table 1; Table S1). The high resolution of Onmi-C scaffolding can be observed in the accurate picture of the two arms of several metacentric /submetacentric chromosomes (C) 1, C2, and C14 corresponding to the C1, C2, and C15 of the *M* . *merluccius* karyotype (García-Souto et al., 2015). The contig and scaffold N50 of the final assembly were 10.9 Mb and 32 Mb, respectively, and 50% of the sequence (L50) comprehended the 10 biggest superscaffolds (Fig. 1; Table 1). The consensus quality (QV) of the final assembly was estimated by Merqury as 42 and the gene completeness reported by BUSCO v5 was 96% using the Actinopterygii_odb10 database.

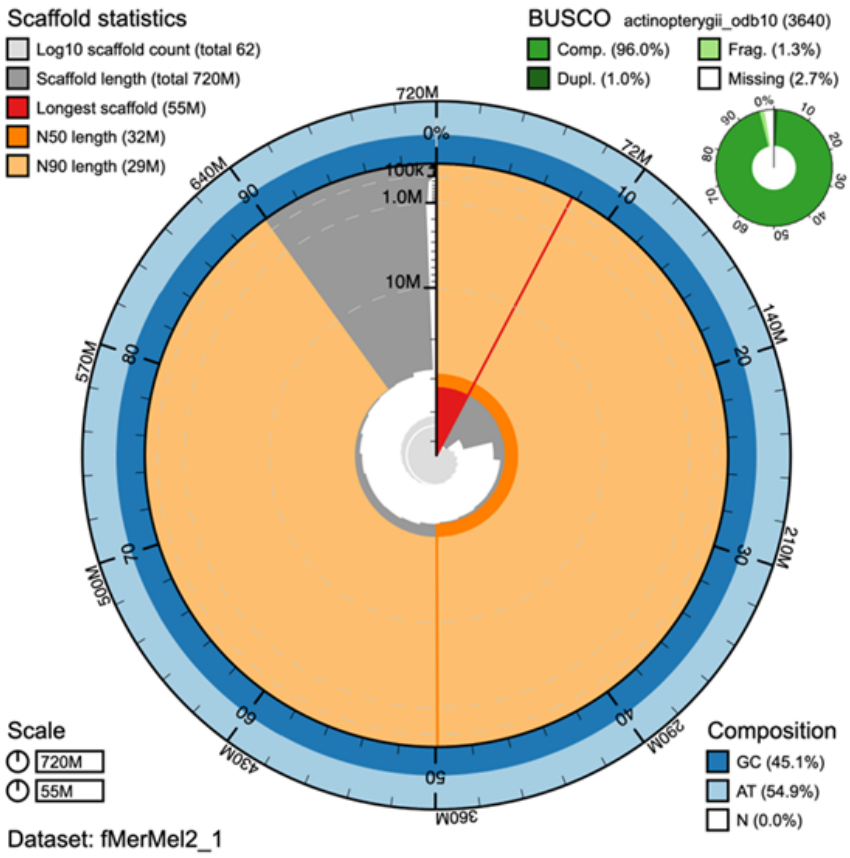

Figure 1. Snail plot showing the main features of the *M .merluccius* genome assembly

Genome annotation

Repeated elements represented 43.3% of the *M. merluccius* genome, being DNA elements, LINEs, LTRs and simple repeats the most frequent. Also 9.1% of repeated elements were unclassified (Table 2).

Table 2: Statistics of repetitive elements of the *M .merluccius* genome

Type	Length (bp)	%
LINEs	28,949,807	4.05
SINEs	3,319,899	0.46
LTR elements	29,597,425	4.14
DNA elements	134,914,544	18.86
Satellites	2,913,070	0.41
Simple repeats	30,402,074	4.25
Low complexity	4,429,408	0.62
Unclassified	64,963,785	9.08
Other	10,287,643	1.44
Total	309,777,655	43.29836253

In total, we annotated 26,625 protein-coding genes that produced 41,543 transcripts (1.56 transcripts per gene) and encoded for 37,855 unique protein products (Table 3; Table S2A and S2B). We were able to assign

functional labels to 78% of the annotated proteins (Table S3). The annotated transcripts contain 11.84 exons on average, with 97% of them being multi-exonic. Additionally, 11,083 noncoding transcripts were annotated, of which 5,683 as lncRNA.

Table 3: Statistics of *M . merluccius* genome annotation

Number of protein-coding genes 26,625

Median gene length (bp) 8,357

Number of transcripts 41,543

Number of exons 285,286

Number of coding exons 266,171

Median UTR length (bp) 1,254

Median intron length (bp) 501

Exons/transcript 11.84

Transcripts/gene 1.56

Multi-exonic transcripts 0.97

Gene density (gene/Mb) 37.21

Comparison of M. merluccius genome assemblies

The chromosome-level genome assembly of *M . merluccius* from this study (female; origin: Galicia, NW Spain; hereafter referred to as MmG), was aligned against the chromosome-level assembly recently released (male; origin: North Sea, Oslofjord, Norway; hereafter referred to as MmN) (Fig. S4). The higher genome size of MmG compared to MmN (715.4 Mb vs 647.0 Mb; 10.6 % difference) was observed in all chromosomes, mainly at their ends, suggesting a more complete and higher-quality assembly. Chromosome numbers were not fully concordant between assemblies reflecting unequal sequence completeness across chromosomes, ordered by size in each assembly. Interestingly, several inversions were suggested when comparing both assemblies, both at the ends (i.e. C2, C4, C7 and C14), but also at the interstitial (C1, C4, C9, C16) regions of chromosomes (Fig. S4; reference MmG). The most notable finding was a large inversion spanning around 20 Mb at C9. While these inversions could result from assembly errors, they might also represent genuine chromosomal polymorphisms in *M .merluccius* .

Identification of the SD region

We screened the whole genome of five males and five females of *M .merluccius* using 1,330,537 SNPs (2.2 SNPs/kb; range: 1.9 (C9) - 2.5 (C17)) filtered by MAF [?] 0.2, assuming an underlying XX/XY or ZZ/ZW SD system (expected frequencies of 0.75 and 0.25 for the X/Z and Y/W alleles, respectively; Table S4). The join inspection of average F_{ST} and F_{IS} using genomic windows of 10 SNPs across each chromosome enabled the identification of candidate regions associated with sex, where F_{ST} and F_{IS} were close to the expected values of 0.5 and -1, respectively, under a XX/XY or a ZW/ZZ SD system (Fig. S5). Interestingly, a side effect of this strategy was the potential recognition of centromere locations of several chromosomes, characterized by very negative and consistent F_{IS} across a broad region, probably related to the presence of tandem repetitive elements. This pattern is particularly evident in three metacentric/submetacentric chromosomes (C1, C2, C16), but also at any of both ends of many acrocentric chromosomes.

Several regions with F_{ST} and F_{IS} above or below the cut-off values (0.4 and -0.5, respectively; see Materials and Methods) were identified at C11, C13, C15 and C19 (Table S5; Fig. S5), but particularly at C9 (Fig. 2A). On this chromosome, multiple peaks or subregions with F_{ST} and F_{IS}values close to those expected under an XX/XY SD chromosome system were detected within a broad region between 8.0 and 18.6 Mb (Fig. 2B). Specifically, a peak of 3.5 kb between 14,327,382-14,330,907 bp displayed a high proportion of

SNPs that were heterozygous in males and homozygous in females (F_{ST} = 0.427; F_{IS}= -0.818) (Fig. 2C). Similar patterns were observed in other conspicuous subregions on this chromosome with high differentiation between sexes and an excess of heterozygotes in males and homozygotes in females (Table S6): i) 8,443,212-8,452,667 bp (F_{ST} = 0.428, F_{IS} = -0.748); ii) 18,505,237-18,510,665 bp (F_{ST} = 0.411, F_{IS} = -0.854); and iii) the broad region of more than 100 kb between 13,144,747-13,249,800 bp (F_{ST} = 0.322, F_{IS} = -0.574)). his consistent pattern suggests the presence of linkage disequilibrium across a substantial region of C9.

Several genes associated with SD or gonadal differentiation were identified within or very close to those regions at C9: (i) androgen receptor alpha (*ar*) and progesterone receptor-like (*pgr*); (ii) glucocorticoid receptor (*nr3c1*); and (iii) motile sperm domain containing 1 (*mospd*), forkhead box O4 (*fox4*), testis expressed 11 (*tex11*) (Table 4). However, the most compelling candidate gene associated with the highest differentiated region was the SRY-box containing gene 3 (*sox3* : 14303025-14303945 bp). This gene mapped at less than 25 kb of the most associated region at C9 consistent with an XX/XY system (14,327,382-14,330,907 bp) (Table S7). To explore a possible gene duplication of *sox3, we compared* the number of reads of this gene in the five males and the five females. No significant differences were detected, ruling out a male-specific duplication of *sox3* .

Potential sexual conflict regions

We also explored other regions showing high genetic differentiation between males and females (F_{ST} > 0.5) across the *M . merluccius* genome, irrespective of their F_{IS} values, using the same sample of males and females. Our intention was to identify and validate genomic regions with allelic variants potentially advantageous to either males or females, reflecting a putative sexual conflict. We initially preselected a list of 14 such regions, which included highly divergent SNPs between the sexes, and mined the genome ±100 kb kb around them looking for clusters of genes with functions potentially specialized by sex.

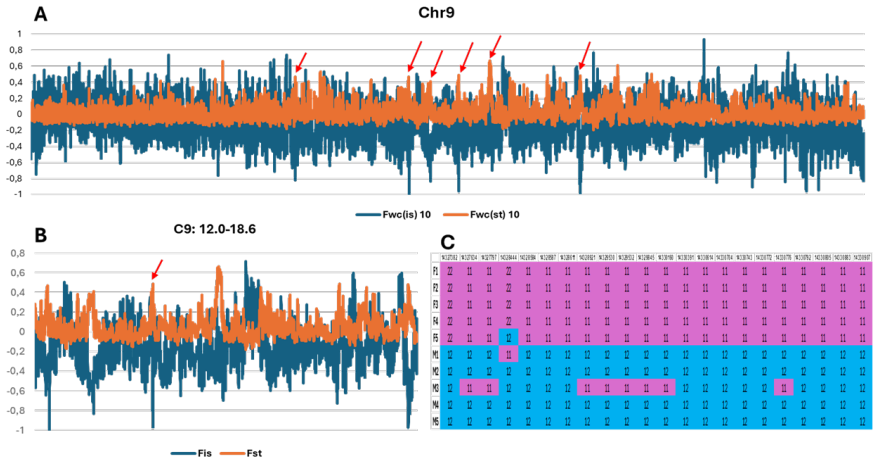

Figure 2: Screening of genetic differentiation (F_{ST}) and intrapopulation fixation index (F_{IS}) using sliding windows across chromosome 9 in the Atlantic population of *M . merluccius* subdivided by sex; A) whole chromosome 9 (positions in kb); B) Zoom between 12 and 18.6 Mb; C) Female and male genotypes in the most associated region of Chr9 (red arrows point to those regions with highest F_{ST} and lowest F_{IS} values; homozygotes in pink and heterozygotes in blue).

	13213	**13224**	**13521**	**13727**	**13750**	**13793**	**13859**
12700	*0,028*	*0,024*	*Motile sperm domain containing 1*	*0,031*	0,056	*Forkhead box O4*	*Forkhead box protein*
13213		**0,000**		**0,001**	*0,038*		
13224			*Mospd2*	0,002	*0,038*	*Fox4*	*FoxN3*

	13213	13224	13521	13727	13750	13793	13859
13727					**0,000**		
13750							
14327							
14328							
14330							
15119							
16318							
19844							
19853							
19854							
21014							

Table 4: Linkage disequilibrium p-values (bold: after Bonferroni correction; underlined: < 0.01; italics < 0.05) between pairs of loci across the most consistent region associated with sex at chromosome 9 of *M . merluccius* . Inserted between SNP markers genotyped with the MassARRAY technology, the annotated genes, mostly associated with testis differentiation.

We finally retained one region at C5 including a large cluster of genes related to olfaction and chemoreception, and another at C12 mostly related to immunity, to be validated in a broader sample (Tables S5 and S7; Fig S4). Among the genes in the C5 olfactory cluster (23.522 and 23.815 Mb), we identified several isoforms of Phosphoinositide phospholipase C, Olfactory receptor C family, v3, Olfactory receptor C family, r1, Olfactory receptor family C subfamily 11 member 4, Olfactory receptor C family, w1 and one Pheromone receptor. Among the genes in the C12 cluster, we identified E3 ubiquitin-protein ligase, HECT domain containing 1, HEAT repeat containing 5a, Heat shock protein nuclear import factor hikeshi, Dehydrogenase/reductase SDR family member 13-like and several isoforms of 4F2 cell-surface antigen heavy chain-like.

Validation of the SD region and potentially sexual conflict regions

The most compelling genomic regions, detected in the sample of five males and five females through WGS, were analyzed in a large sample to confirm their association with sex. 25 SNPs from these regions were selected for genotyping in 121 individuals (56 males and 65 females) coming from the same population using the MassARRAY technology (Table S8). SNPs were selected based on their differentiation pattern between males and females (F_{ST} and F_{IS}) and the absence of additional SNPs within $\pm$100 bp in their flanking regions to avoid interference with primer annealing and genotyping. Allelic variants at all the SNPs conformed to the *in silico* data from WGS, and genotyping was highly consistent, with no missing data across the more than 3000 genotypes obtained (Table S9). Association with sex was highly significant for most SNPs at C9 according to a XX/XY system, consistent with the WGS data, , although the association was not perfect. SNPs 9_13213378 and 9_13224545 showed highly significant differentiation between males and females (F_{ST} = 0.223; P = 0) and the highest heterozygote excess in males (F_{IS}= -0.330) (Table S10). We also explored LD across 12.0-21.0 Mb of C9 using this set of SNPs and as previously suggested, highly significant LD was detected at very distant positions ($>$ 6 Mb) within the region, particularly between 13.3 and 14.3 Mb, where notable genes related to gonad differentiation are located (Table 4).Regarding the other two regions analyzed at C5 and C12, high genetic differentiation was still maintained in this large sample for two SNPs at C5 5_22355441 and 5_23175612, F_{ST} = 0.214; P = 0), and for the three SNPs examined at C12 (12_13762434, 12_18517951 and 12_18533985, F_{ST} $>$ 0.170; P = 0). Furthermore, the putative inversion observed when aligning MmG (female) and MmN (male) genomes (between 4 Mb and 24 Mb; Fig. 3) could be related to the LD detected across the large region at C9 outlined before and could be potentially linked to recombination suppression between X and Y chromosomes.

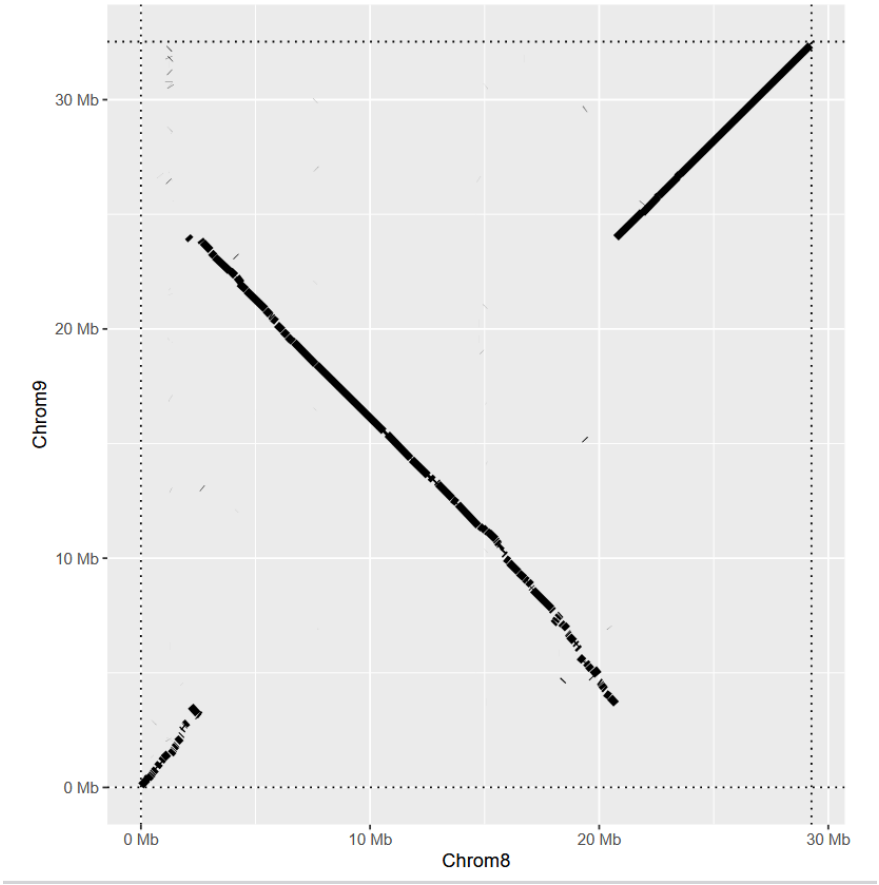

Figure 3 : Alignment of chromosome 9 from the *M .merluccius* MmG genome with the homologous chromosome 8 from the MmN genome (abscises).

Discussion

The chromosome-level genome assembly of *M . merluccius* presented in this study (MmG) represents a significant improvement over the previous version (PRJEB12469; March 2018; scaffold N50: 5.1 kb; total lenght: 401 Mb; no annotation). The MmG genome demonstrate greater contiguity, completeness and annotation quality (N50: 32.0 Mb; 715.4 Mb; 26,625 protein coding genes; 11,083 noncoding transcripts including 5,683 as lncRNA) compared to the recently released MmN genome (PRJEB77069; September, 2024: N50: 28.9 Mb;: 647 Mb; 22,124 protein coding genes). The larger genome size of the MmG is primarily attributed to its higher contiguity at the chromosome ends, while the enhanced annotation likely reflects the integration of RNA-seq data from five tissues using pooled samples. Alignment of the MmG and MmN genomes suggests the presence of multiple inversions, both at chromosome ends-which could reflect technical assembly artifacts, but also at interstitial regions of several chromosomes.

Validation of these inversions at the population level will require whole-genome resequencing or PCR assays targeting the inversion breakpoints (Hanlon et al., 2022). Overall, this newly assembled genome provides an essential resource for population genomics studies of European hake and for studying the genetic basis of adaptive traits to support sustainable fisheries management.

We followed a cost-effective strategy to identify the SD region in*M . merluccius* , which made it possible to suggest a SD system and a putative candidate gene, as successfully demonstrated in other fish species (Martínez et al., 2021; de la Herrán et al., 2023). A consistent contiguous and well-annotated genome,

such as the one here presented, is an essential reference for this strategy ensuring completeness across all chromosomal regions, including centromeres and telomeres. Resequencing a small number of males and females at appropriate coverage (30x in this study) would be enough to identify specific genomic regions, not only differentiated between males and females (F_{ST} ˜ 0.5), but also involving a significant heterozygote excess (F_{IS} ˜ -1) either in males (XX/XY) or females (ZW/ZZ), depending on the SD system. Unlike the pooling strategy frequently used for this purpose (Imarazene et al., 2021; Nakamoto et al., 2021; Kulh et al., 2022), which inherently renders false positives, our approach refines candidate regions by simultaneously analyzing both parameters as averages across sliding windows of adjustable size. . This strategy necessarily assumes a strong genetic component underlying SD, but if this is true, the SD region should be identifiable with a significant reduction of false positives. In fact, we could detect the only genomic region fully compatible with a XX/XY SD system at 14.3 Mb in C9. Furthermore, several other genomic regions showing high differentiation between males and females did not adjust to a typical chromosome SD pattern, but instead represent potential sexual conflict regions (see below).

Our findings suggest that the main SD region in *M .merluccius* is located on C9, consistent with an XX/XY system, and that *sox3 is* the most likely candidate MSD gene. Several lines of evidence support these observations: i) F_{ST} and F_{IS} values above and below the established thresholds, close to the expected 0.5 and -1, respectively; maximum differentiation was observed near *sox3* , where most SNPs were homozygous in males and heterozygous in females; ii) highly significant LD was detected between SNP markers across an extensive region of C9, including the putative MSD gene, in a large sample of males and females; this LD pattern is compatible with the presence of a putative ˜20 Mb inversion encompassing the SD region, potentially suppressing crossing-over between the X and Y chromosomes; iii) several other genes involved in gonad differentiation mapped within the inversion, ensuring their co-transmission as a block linked to the X and Y chromosomes. A duplication of *sox3* has been established as the MSD gene in mammals (Katsura et al., 2018), and it has also been reported as the MSD gene inseveral fish species within the genus *Oryzias* (Takehana et al., 2014; Myosho et al., 2015), in this case not related to gene duplication. Takehana et al. (2014) demonstrated that a cis-regulatory element located downstream of *sox3* drives increasing expression, triggering testis development via *gsdf* , a member of the transforming growth factor beta family associated with SD in several teleost (Kitano et al., 2024). In our study, we did not identify specific allelic variants associated with the X and Y alleles of*sox3* , nor evidence of a gene duplication linked to the Y chromosome. However, the main difference between the X and Y chromosomes was located 25 kb downstream *sox3* , suggesting a potential regulatory element. Further investigation on gene expression of*sox3* in males and females during early stages of gonad differentiation, will be required to confirm this hypothesis.

Additional factors are required to fully elucidate the mechanism underlying SD in *M . merluccius* . Validation of SNPs located in the region most strongly associated with sex, using a large sample of 121 males and females, revealed highly significant associations consistent with an XX/XY system, but a perfect genotype-sex correlation was not observed. Disentangling SD in wild populations is complex, due to potential interactions between genetic factors and environment cues across the species ' large distribution area, which can manifest as intraspecific variation in the SD system. Significant variation in the SD systems has been documented across the geographic ranges of several fish species, often linked to specific environmental factors, such as the Northern pike (*Essox Lucius* ; Johnson et al., 2024), flathead great mullet (*Mugil cephalus* ; Ferrareso et al., 2021), Atlantic silverside (*Menidia menidia* ; Lagomarsino & Conover, 1993) and Argentinian pejerrey (*Odontestes bonariensis* ; Del Fresno et al., 2022), among others. Notably, many species where SD systems have been characterized correspond to aquaculture or to model species, usually reared in stable environments. Changes of the environmental context can significantly influence SD systems, as demonstrated in zebrafish, where the original ZW/ZZ system was lost under domesticated conditions (Kossack & Draper, 2019). For *M . merluccius* , the influence of environmental factors on a potentially non-fully penetrant genetic system cannot be excluded. Preliminary data suggest differences in the SD system between populations at the extremes of the species' distribution range, namely Norway and Galicia (Martínez et al. 2024). These populations exhibit low but significant genetic divergence based on neutral markers (F_{ST} = 1.6%; Westgaard et al., 2017). Intriguingly, the Galician European hake population displays different reproduction seasons in spring

in summer, suggesting potential population subdivision that may be critical to fully understand the SD system and its relationship to the broader population structure of European hake in the Atlantic Ocean.

Sexual conflict has been proposed as a key factor driving the evolution of SD systems (Van Doorn & Kirkpatrick, 2007). The rapid SD turnover reported in fish represents an excellent model to test this hypothesis. For instance, studies on cichlids from lake Malawi have identified a novel MSD gene associated with the correct transmission of allelic variants linked to sexually dimorphic coloration patterns favorable to males or females (Roberts et al., 2009; Parnell & Streelman, 2013). However, beyond color patterns, numerous other traits related to sexual selection or sex-specific reproductive roles could differ between males and females, and manifest as significant genetic differentiation (F_{ST}) at specific genomic regions. Consequently, associations between genetic markers and sex observed in whole genome screenings might sometimes reflect sexual conflict rather than direct involvement in sex determination, suggesting caution when searching for SD genes, particularly through pooling strategies. In this study, whole-genome screening revealed several regions in different chromosomes, some of them associated with clusters of genes potentially involved in sexual conflict. Given the limited sample size (5 males and 5 females), the risk of false positives was mitigated by validating the observed divergence between sexes in a larger sample set. Among the most divergent regions outside C9, regions at C5 and C12 were prioritized for further validation due to their association with gene clusters related to suggestive traits, such as reproduction and immunity. Several SNPs in these regions maintained the highly significant association observed with sex in a larger sample of 56 males and 65 females ($P = 0$). The sex-specific roles in reproduction may explain the associations detected with genes related to olfactory reception and chemosensory communication (Rodríguez-Villamayor et al., 2023; Torres et al., 2024), functions for which sex dimorphism has been extensively documented, not only in humans (Cherry and Baum, 2019), but also in fish species (Kawaguchi et al., 2014; Lei et al., 2019; Shinohara & Kobayashi, 2020; Wang et al., 2020). Similarly, sex differences in immune responses, widely studied in humans (Klein and Flanagan, 2016; Takahashi et al., 2020), have also been reported in fish (Campbell et al., 2021). While many of these physiological differences are influenced by sex hormones, the role of selection on alternative genetic variants should not be discarded. These observations and especially the proposed interpretations would deserve further work for their confirmation, but they suggest an interesting outcome to be tested when performing WGS at individual level.

Conclusions

Here, we present a chromosome-level genome assembly of *Merluccius merluccius* , which outperforms previous versions regarding contiguity, completeness and annotation. This genome was used to search for the SD region in European hake. A large region on chromosome C9 showing strong LD was identified, containing several genes involved in gonad differentiation. Among these, the most promising candidate was*sox3* , which was located near the region of highest divergence between males and females consistent with an XX/XY SD system. Alignment of chromosome-level genomes from one male and one female, revealed a potential inversion encompassing a substantial segment of C9, which might hinder crossing-over between the X and Y chromosomes. The association was validated in a large sample of males and females; however the data suggest that environmental factors or intraspecific variation in the SD system may contribute to the observed attenuation of the association. Additionally, we explored other genomic regions associated with sex across other chromosomes, identifying potential candidates for sexual conflict. Significant genetic divergence between males and females was confirmed in regions at C5, associated with olfactory-chemosensory communication, and C12, linked to immunity-related genes. The genomic resources and molecular tools developed in this study provide valuable insights into the biology of*M . merluccius* and will contribute to the sustainable management of *its* fisheries.

Ethical Statement

All samples used in this study were obtained from commercial fishing activities or scientific research surveys conducted in accordance with national and international fisheries regulations. Ethical approval was not required as the samples were collected as part of routine fisheries operations and no experimental interventions were performed on live organisms.

Acknowledgements

This study was supported by the Regional Government Xunta de Galicia (Spain) (Grant number ED431C 2022/33, Consellería de Cultura, Educación y Universidad, Vicepresidencia Segunda, Consellería de Economía, Empresa e Innovación). This research has been carried out under the framework of the Spain's Recovery and Resilience Plan, and more specifically under the investment line no.1 of its component number 17, where the complementary RTDI plan with the autonomous regions of Spain are foreseen, with one of those being the Complementary RTDI Plan for Marine Science which includes the Marine Science Program for Galicia. The DESCARSEL0922 scientific campaign by EU-Spain has been co-funded by the European Union through the European Maritime and Fisheries Fund (EMFF) within the National Program for the collection, management, and use of data in the fisheries sector, as well as support for scientific advice related to the Common Fisheries Policy and the DESCARSEL project. Natalia Petit-Marty was supported by MSCA-PFFishAdapt 101066785 and Maialen Carballeda by a contract within the framework of the State Plan for Scientific, Technical and Innovation Research 2021-2023 (PID2022-137821OB-C31).

Availability of data and materials

To be uploaded to public repositories upon acceptance

References

Apostologamvrou, C., Vlachou, M., Theocharis, A., Ntavaros, C., & Klaoudatos, D. (2023). Reproductive aspects of European hake, (*Merluccius merluccius* , Linnaeus, 1758) based on histological depiction of both sexes in the Eastern Mediterranean (Aegean Sea).*Regional Studies in Marine Science* , 68, 103281. doi: 10.1016/j.rsma.2023.103281.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C., & Tree of Sex Consortium. (2024). Sex determination: why so many ways of doing it? *PLoS Biology* , 12, e1001899. doi: 10.1371/journal.pbio.1001899

Campbell, J. H., Dixon, B., & Whitehouse, L. M. (2021). The intersection of stress, sex and immunity in fishes.*Immunogenetics* , 73, 111-129. doi: 10.1007/s00251-020-01194-2.

Casey, J., & Pereiro, J. (1995). European hake (*M. merluccius*) in the North-east Atlantic. In: Alheit, J., Pitcher, T.J. (eds) Hake.*Chapman & Hall Fish and Fisheries Series* , vol 15. Springer, Dordrecht. doi: 10.1007/978-94-011-1300-7_5

Cerviño, S. (2014). Estimating growth from sex ratio-at-length data in species with sexual size dimorphism. *Fisheries Research* , 160, 112-119. doi: 10.1016/j.fishres.2013.11.010

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , 34, i884–i890. doi: 10.1093/bioin forma tics/bty560

Cherry, J. A., & Baum, M. J. (2020). Sex differences in main olfactory system pathways involved in psychosexual function. *Genes, Brain and Behavior* , 19, e12618. doi: 10.1111/gbb.12618.

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. doi: 10.3390/genes8100258

de la Herrán, R., Hermida, M., Rubiolo, J.A., Gómez-Garrido, J., Cruz, F., Robles, F., Navajas-Pérez, R., Blanco, A., Villamayor, P. R., Torres, D., Sánchez-Quinteiro, P., Ramirez, D., Rodríguez, M.E., Arias-Pérez, A., Cross, I., Duncan, N., Martínez-Peña, T., Riaza, A., Millán, A., De Rosa, M.C., Pirolli, D., Gut, M., Bouza, C., Robledo, D., Rebordinos, L., Alioto, T., Ruíz-Rejón, C., & Martínez, P. (2023). 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* . *Molecular Ecology Resources* , 23, 886-904. doi: 10.1111/1755-0998.13750.

Del Fresno, P. S., Garcia de Souza, J. R., Colautti, D. C., Yamamoto, Y., Yokota, M., Strüssmann, C. A., & Miranda, L. A. (2023). Sex reversal of pejerrey (*Odontesthes bonariensis*), a species with temperature-dependent sex determination, in a seminatural environment.*Journal of Fish Biology* , 102, 75-82. doi: 10.1111/jfb.15241.

Domínguez-Petit, R., & Saborido-Rey, F. (2010). New bioenergetic perspective of European hake (*Merluccius merluccius* L.) reproductive ecology. *Fisheries Research,* 104*,* 83–88. doi: 10.1016/j.fishres.2009.09.002

El Bouzidi, C., Abid, N., Awadh H., Bakkli, M., & Zerrouk, M.H. (2022). Growth and mortality of the European hake *Merluccius merluccius* (Linnaeus, 1758) from the North of Moroccan Atlantic coasts.*Egyptian Journal of Aquatic Research* , 48, 233-239. doi: 10.1016/j.ejar.2021.12.003

FAO. (2020). The state of the Mediterranean and Black Sea fisheries 2020, General Fisheries Commission for the Mediterranean. doi: 10.4060/cb2429en

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. (2020). fshr: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations. *iScience* , 24, 101886. doi: 10.1016/j.isci.2020.101886

García-Fernández, C., Suca, J., Llopiz, J., Álvarez, P., Dominguez-Petit, P., Saborido-Rey, F. 2021. Spatial and temporal variability in the occurrence and abundance of European hake larvae,*Merluccius merluccius* , on the Galician Shelf (NE Atlantic).*Frontiers in Marine Science* , 8, 2021. doi: 10.3389/fmars.2021.696246.

García-Souto, D., Troncoso, T., Pérez, M., & Pasantes, J.J. (2015). Molecular cytogenetic analysis of the European hake *Merluccius merluccius* (Merlucciidae, Gadiformes): U1 and U2 snRNA gene clusters map to the same location. *PLoS ONE* , 10, e0146150. doi: 10.1371/journal.pone.0146150

Guiguen, Y., Fostier, A., & Herpin, A. (2019). Sex determination and differentiation in fish: Genetic, genomic, and endocrine aspects. In H.-P. Wang, F. Piferrer, S.-L. Chen, & Z.-G. Shen (Eds.), *Sex control in aquaculture* (Vol. 1, 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 Biology* , 9, R7. doi: 10.1186/gb-2008-9-1-r7.

Hanlon, V.C.T., Lansdorp, P.M., & Guryev, V. (2022). A survey of current methods to detect and genotype inversions. *Human Mutation* , 43, 1576-1589. doi: 10.1002/humu.24458

Hidalgo, J., Olsen, E. M., Ohlberger, j., Saborido-Rey, F., Murua, H., Piñeiro, C., & Stenseth, N. C. (2014). Contrasting evolutionary demography induced by fishing: the role of adaptive phenotypic plasticity. *Ecological Applications* , 24, 1101-1114. doi: 10.1890/12-1777.1

Iglesias, J., Lago M., Sánchez F., & Cal, R. (2010). Capture, transport and acclimatization to captivity of European hake, *Merluccius merluccius* L: Preliminary data on feeding and growth. *Aquaculture Research* , 41, 607-609. doi: 10.1111/j.1365-2109.2009.02365.x.

Imarazene, B., Du, K., Beille, S., Jouanno, E., Feron, R., Pan, Q., Torres-Paz, J., Lopez-Roques, C., Castinel, A., Gil, L., Kuchly, C., Donnadieu, C., Parrinello, H., Journot, L., Cabau, C., Zahm, M., Klopp, C., Pavlica, T., Al-Rikabi, A., Liehr, T., Simanovsky, S.A., Bohlen, J., Sember, A., Perez, J., Veyrunes, F., Mueller,. T.D, Postlethwait, J.H., Schartl, M., Herpin, A., Rétaux, S., & Guiguen, Y. (2021). A supernumerary "B-sex" chromosome drives male sex determination in the Pachón cavefish, *Astyanax mexicanus. Current Biology,* 31, 4800-4809.e9. doi: 10.1016/j.cub.2021.08.030.

Izquierdo, F., Paradinas, I., Cerviño, S., Conesa, D., Alonso-Fernández, A., Velasco, F., Preciado, I., Punzón, A., Saborido-Rey, F., & Pennino, M.G. (2021). Spatio-temporal assessment of the European hake (*Merluccius merluccius*) recruits in the Northern Iberian Peninsula. *Frontiers in Marine Science, 8,* 614675. doi: 10.3389/fmars.2021.614675

Johnson, H.A., Rondeau, E.B., Sutherland, B.J.G., Minkley, D.R., Leong, J.S., Whitehead, J., Despins, C.A., Gowen, B.E., Collyard, B.J., Whipps, C.M., Farrell, J.M., & Koop, B.F. (2024). Loss of genetic variation and ancestral sex determination system in North American northern pike characterized by whole-genome resequencing. *G3 Genes, Genomes, Genetics,* 14, jkae183. doi: 10.1093/g3journal/jkae183.

Katsura, Y., Kondo, H.X., Ryan, J., Harley, V., & Satta, Y. (2018). The evolutionary process of mammalian sex determination genes focusing on marsupial SRYs. BMC *Evolutionary Biology* , 18, 3. doi: 10.1186/s12862-018-1119-z

Kawaguchi, Y., Nagaoka, A., Kitami, A., Mitsuhashi, T., Hayakawa, Y., & Kobayashi, M. (2014). Gender-typical olfactory regulation of sexual behavior in goldfish. *Frontiers in Neurosciences* , 8, 91. doi: 10.3389/fnins.2014.00091.

Kitano, J., Ansai, S., Takehana, Y., & Yamamoto, Y. (2024). Diversity and convergence of sex-determination mechanisms in Teleost fish.*Annual Review of Animal Biosciences* , 12, 233-259. doi: 10.1146/annurev-animal-021122-113935

Klein, S., & Flanagan, K. (2016). Sex differences in immune responses.*Nature Reviews Immunology,* 16, 626–638. doi: 10.1038/nri.2016.90

Korta, M., Domínguez-Petit, R., Murua, H., & Saborido-Rey, F. (2010). Regional variability in reproductive traits of European hake*Merluccius merluccius* L. populations. *Fisheries Research* , 104 (1-3): 64-72. doi: 10.1016/j.fishres.2009.03.007.

Kossack, M.E., & Draper, B.W. (2019). Genetic regulation of sex determination and maintenance in zebrafish (*Danio rerio*).*Current Topics in Developmental Biology* , 134, 119-149. doi: 10.1016/bs.ctdb.2019.02.004.

Kuhl, H., Du, K., Schartl, M., Kalous, L., Stöck, M., & Lamatsch, D.K. (2022). Equilibrated evolution of the mixed auto-/allopolyploid haplotype-resolved genome of the invasive hexaploid Prussian carp.*Nature Communications* , 13, 4092. doi: 10.1038/s41467-022-31515-w.

Kutsyn, D. N., Tamoykin, I. Y., Vdodovich, I. V., Klimova, T. N., & Donchik,P. I. (2024). Finding of the European hake *Merluccius merluccius* (Merlucciidae) off the Black Sea shore of Crimea.*Journal of Ichthyology* , 64, 80–89. doi: 10.1134/S003294522401003X

Lagomarsino, I.V., & Conover, D.O. (1993). Variation in environmental and genotypic sex-determining mechanisms across a latitudinal gradient in the fish, *Menidia menidia. Evolution* , 47, 487-494. doi: 10.1111/j.1558-5646.1993.tb02108.x.

Leone, A., Álvarez, P., García, D., Saborido-Rey, F., & Rodriguez-Ezpeleta, N. (2019). Genome-wide SNP based population structure in European hake reveals the need for harmonizing biological and management units. *ICES Journal of Marine Science* , 76(7), 2260-2266. doi: 10.1093/icesjms/fsz161.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* , 25, 1754–1760. doi: 10.1093/bioin forma tics/btp324

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. doi: 10.1093/bioin forma tics/btp352

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences.*Bioinformatics* , 34:3094-3100. Doi: 10.1093/bioinformatics/bty191

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: 10.1016/j.ygeno.2021.04.007

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. Petit, N., Blanco, A., Carballeda, M., Gómez-Garrido, J., Cruz, F., Alioto, T., Saborido, F., & Casas, L. (2024). Sex determination in fish: New insights from the European hake*Merluccius merluccius* chromosome-level genome assembly (Oral communication). In AQUA2024-World Aquaculture Society. 26-30 August 2024. Copenhagen (Dinamarca).

Mehault, S., Dominguez-Petit, R., S. Cerviño, & Saborido-Rey, F. (2010). Variability in Total Egg Production and implications for management of the Southern stock of European hake. *Fisheries Research* , 104 (1-3): 111-122

Morales-Nin, B., Pérez-Mayol, S., MacKenzie, K., Catalán, I. A., Palmer, M., Kersaudy, T., & Mahé, K. (2022). European hake (*Merluccius merluccius*) stock structure in the Mediterranean as assessed by otolith shape and microchemistry. *Fisheries Research* , 254, 106419. doi: 10.1016/j.fishres.2022.106419

Murua, H. (2010). The biology and fisheries of European hake,*Merluccius merluccius* , in the North-East Atlantic. *Advances in Marine Biology* , 58, 97-154. doi: 10.1016/B978-0-12-381015-1.00002-2.

Myosho, T., Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2015). Turnover of sex chromosomes in celebensis group medaka fishes. *G3: Genes, Genomes, Genetics* , 5, 2685–2691. doi: 10.1534/g3.115.021543

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. doi: 10.1371/journal.pgen.1009705.

Pan, O., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., Rondeau, E., Goetz, F. W., Larson, W. A., Bernatchez, L., Tringali, M., Curran, S. S., Saillant, E., Denys, G. P. J., von Hippel, F. A., Chen, S., López, J. A., Verreycken, H., Ocalewicz, K., Guyomard, R., Eche, C., Lluch, J., Roques, C., Hu, H., Tabor, R., De Haan, P., Nichols, K. M., Journot, L., Parrinello, H., Klopp, C., Interesova, E. A., Trifonov, W., Schartl, M., Postlethwait, J., and Guiguen, Y. (2021). The rise and fall of the ancient northern pike master sex-determining gene. *eLife* , 10, e62858. doi: 10.7554/eLife.62858

Parnell, N., & Streelman, J. (2013). Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity,* 110, 239–246. doi: 10.1038/hdy.2012.73

Piñeiro, C., & Saínza, M. (2003). Age estimation, growth and maturity of the European hake (*Merluccius merluccius* (Linnaeus, 1758)) from Iberian Atlantic waters. *ICES Journal of Marine Science* , 60, 1086–1102. doi: 10.1016/S1054-3139(03)00086-9

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. doi: 10.1016/j.ygeno.2022.110277

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

Roberts, R. B., Ser, J. R., & Kocher, T. D. (2009). Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* , 26, 998-1001. doi: 10.1126/science.1174705

Serrat, A., Saborido-Rey, F., Garcia-Fernandez, C., Muñoz, M., Lloret, J., Thorsen, A., & Kjesbu, O.S. (2019). New insights in oocyte dynamics shed light on the complexities associated with fish reproductive strategies. *Scientific Reports* 9, 18411. doi: 10.1038/s41598-019-54672-3.

Shinohara, Y., & Kobayashi, M. (2020). Sexual bipotentiality of the olfactory pathway for sexual behavior in goldfish. *Fisheries Science* , 86, 819–827. doi: 10.1007/s12562-020-01454-w

Takahashi, T., Ellingson, M.K., Wong, P. et al. (2020). Sex differences in immune responses that underlie COVID-19 disease outcomes.*Nature* , 588, 315–320. doi: 10.1038/s41586-020-2700-3

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

Torres, D., Villamayor, P. R., Salisbury, S. J., Ruiz Daniels, R., Blanco, A., Riaza, A., Blanchet, I., Rasines, I., Martín, I., Martínez, P., & Robledo, D. (2023). Characterisation of the olfactory organ of Senegalese sole (*Solea senegalensis*) using single-nuclei RNA-Seq (Oral communication). In Aquaculture Europe 23 (AE2023). 18-21 September 2023. Viena (Austria).

Torres, D., Villamayor, P. R., Román, A., García, P., Martínez, P., & Sanchez-Quinteiro, P. (2024). In-depth histological, lectin-histochemical, immunohistochemical and ultrastructural description of the olfactory rosettes and olfactory bulbs of turbot (*Scophthalmus maximus*). *Cell and Tissue Research* , 397, 215-239. doi: 10.1007/s00441-024-03906-6

Van Doorn, G.S., & Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature* , 449, 909-912. doi: 10.1038/nature06178.

Wang, Y., Jiang, H., & Yang, L. (2020). Transcriptome analysis of zebrafish olfactory epithelium reveal sexual differences in odorant detection. *Genes* (Basel), 11, 592. doi: 10.3390/genes11060592.

Westgaard, J.I., Staby, A., Godiksen, J.A., Geffen, A.J., Svensson, A., Charrier, G., Svedäng, H., & André, C. (2017). Large and fine scale population structure in European hake (*Merluccius merluccius*) in the Northeast Atlantic. *ICES Journal of Marine Science* , 74, 1300–1310. doi: 10.1093/icesjms/fsw249

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. doi: 10.1007/978-3-319-24277-4

Winter D (2020). pafr: Read, Manipulate and Visualize 'Pairwise mApping Format' Data_. R package version 0.0.2. https://CRAN.R-project.org/package=pafr

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* , 15, 323-354. doi: 10.1111/j.1469-1809.1949.tb02451.x