









## RESOURCE ARTICLE

# Generation of a chromosome-level genome assembly for Pacific halibut (*Hippoglossus stenolepis*) and characterization of its sex-determining genomic region

Andrew J. Jasonowicz<sup>1</sup> | Anna Simeon<sup>1,2</sup> | Margot Zahm<sup>3</sup>  | Cédric Cabau<sup>4</sup>  |  
Christophe Klopp<sup>3</sup>  | Céline Roques<sup>5</sup> | Carole Iampietro<sup>5</sup>  | Jérôme Lluch<sup>5</sup> |  
Cécile Donnadieu<sup>5</sup>  | Hugues Parrinello<sup>6</sup>  | Daniel P. Drinan<sup>2</sup> | Lorenz Hauser<sup>2</sup> |  
Yann Guiguen<sup>7</sup>  | Josep V. Planas<sup>1</sup> 

<sup>1</sup>International Pacific Halibut Commission, Seattle, Washington, USA

<sup>2</sup>School of Aquatic and Fishery Science, University of Washington, Seattle, Washington, USA

<sup>3</sup>SIGENAE, Bioinfo Genotoul, UMIAT, INRAE, Castanet-Tolosan, France

<sup>4</sup>SIGENAE, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France

<sup>5</sup>INRAE, GeT-PlaGe, Castanet-Tolosan, France

<sup>6</sup>MGX-Montpellier GenomiX, University of Montpellier, CNRS, INSERM, Montpellier, France

<sup>7</sup>INRAE, LPGP, Rennes, France

## Correspondence

Josep V. Planas, International Pacific Halibut Commission, Seattle, Washington, WA 98199, USA.

Email: [josep.planas@iphc.int](mailto:josep.planas@iphc.int)

## Funding information

Agence Nationale pour la Recherche, Grant/Award Number: ANR-10-INBS-09; International Pacific Halibut Commission; North Pacific Research Board, Grant/Award Number: 2110

**Handling Editor:** Shotaro Hirase

## Abstract

The Pacific halibut (*Hippoglossus stenolepis*) is a key species in the North Pacific Ocean and Bering Sea ecosystems, where it also supports important fisheries. However, the lack of genomic resources limits our understanding of evolutionary, environmental and anthropogenic forces affecting key life history characteristics of Pacific halibut and prevents the application of genomic tools in fisheries management and conservation efforts. In the present study, we report on the first generation of a high-quality chromosome-level assembly of the Pacific halibut genome, with an estimated size of 602 Mb, 24 chromosome-length scaffolds that contain 99.8% of the assembly and a N<sub>50</sub> scaffold length of 27.3 Mb. In the first application of this important resource, we conducted genome-wide analyses of sex-specific genetic variation by pool sequencing and characterized a potential sex-determining region in chromosome 9 with a high density of female-specific SNPs. Within this region, we identified the *bmpr1ba* gene as a potential candidate for master sex-determining (MSD) gene. *bmpr1ba* is a member of the TGF- $\beta$  family that in teleosts has provided the largest number of MSD genes, including a paralogue of this gene in Atlantic herring. The genome assembly constitutes an essential resource for future studies on Pacific halibut population structure and dynamics, evolutionary history and responses to environmental and anthropogenic influences. Furthermore, the genomic location of the sex-determining region in Pacific halibut has been identified and a putative candidate MSD gene has been proposed, providing further support for the rapid evolution of sex-determining mechanisms in teleost fish.

## KEYWORDS

candidate master sex-determining gene, genome assembly, Pacific halibut, pool-sequencing, sex-associated genomic region, whole genome sequencing

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 International Pacific Halibut Commission. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

While the phenotypic outcome of sex determination is highly conserved, a diverse array of mechanisms underlie sex determination throughout the tree of life with a large portion of this diversity being represented in vertebrate species (Bachtrog et al., 2014; Pennell et al., 2018). In some vertebrate clades, such as mammals and birds, sex determination systems are highly conserved (Pennell et al., 2018). Both rely on genetic sex determination mechanisms (GSD) where, in mammals, males are the heterogametic sex (XX/XY system) and in birds, females are the heterogametic sex (ZZ/ZW system) (Bachtrog et al., 2014; Capel, 2017). Teleost fishes on the other hand, exhibit a remarkable array of sex determination mechanisms. This includes, but is not limited to, temperature-dependent sex determination (TSD), behavioural sex determination (BSD), and GSD mechanisms (Bachtrog et al., 2014; Godwin & Roberts, 2018). In some cases, taxonomically closely related species display alternate mechanisms of sex determination. Both XX/XY and ZZ/ZW systems have been documented in medaka (*Oryzias spp.*) (Takehana et al., 2008), mosquitofish (*Gambusia spp.*) (Kottler et al., 2020), Esociformes (Pan, Feron, et al., 2021), and the Pleuronectiformes or flatfishes.

The diversity of sex determination mechanisms exhibited among flatfishes makes them a particularly interesting group to study from an evolutionary perspective. Not only are both XX/XY and ZZ/ZW systems represented but there is some evidence that environmental factors may play a role in sex determination for some flatfish species (Luckenbach et al., 2009). It is estimated that flatfishes diverged from a common ancestor with threadfins (Polynemidae) between 57.3–72.6 million years ago (Ma) (Harrington et al., 2016). Among the Pleuronectidae family, the genus *Hippoglossus* consists of two species, the Pacific (*Hippoglossus stenolepis*) and Atlantic halibut (*H. hippoglossus*). Levels of genetic divergence suggest that a relatively recent common ancestor was shared between the sister species as it is estimated that these species diverged from one another 0.9–4.5 Ma and that a common ancestor dispersed from the Pacific to the Atlantic ocean sometime after the Bering Strait opened or that some degree of interocean migration was maintained until relatively recently in an evolutionary time scale (Grant et al., 1984). Interestingly, alternate genetic sex determination systems have been documented for these sister species. Females are the heterogametic sex in Pacific halibut (ZZ/ZW) (Drinan et al., 2018; Galindo et al., 2011) while males are the heterogametic sex (XX/XY) in Atlantic halibut (Einfeldt et al., 2021; Palaiokostas et al., 2013) and in Greenland halibut (Ferchaud et al., 2022). Moreover, both XX/XY and ZZ/ZW sex determination systems are documented in more distantly related flatfish species. The turbot (*Scophthalmus maximus*) (Martinez et al., 2021) and half-smooth tongue sole (*Cynoglossus semilaevis*) (Chen et al., 2014), utilize a ZZ/ZW system while an XX/XY system has been proposed for the Senegalese sole (*Solea senegalensis*) (Guerrero-Czar et al., 2021; Portela-Bens et al., 2017). While sex determination systems in other vertebrate lineages are much older (therian mammals ~166 Ma (Wallis et al., 2008), avian ~102 Ma

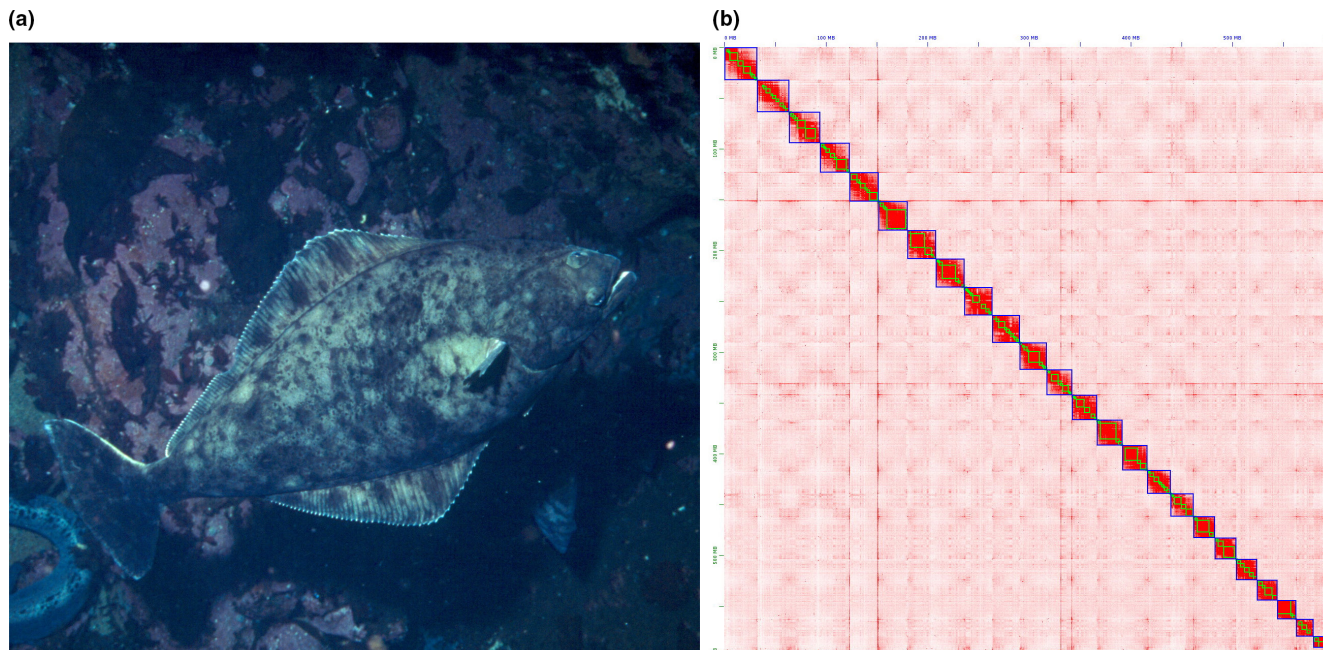
(Zhou et al., 2014)), flatfish offer a unique opportunity to study the biological mechanisms behind sex determination.

The Pacific halibut is a migratory demersal flatfish species of the order Pleuronectiformes that occupies a top trophic role in the North Pacific Ocean and Bering Sea ecosystems (Best, 1981). It is one of the largest flatfish species worldwide and exhibits adult sexual dimorphic growth, females growing larger than males to sizes reaching 230 kg in weight and 2.4 m in length (Figure 1a). This species also supports important commercial, recreational, and subsistence fisheries. Recent research to identify genetic markers diagnostic of sex in Pacific halibut for their application in fisheries management, as the commercial fishery is currently operating under minimum size limits (Stewart et al., 2021), resulted in the identification of a number of sex-linked loci that, in the absence of a Pacific halibut genome, mapped to the Z chromosome of the half-smooth tongue sole and largely to single linkage groups in Atlantic halibut and turbot, suggesting that genetic sex in Pacific halibut is likely under the control of loci present on a single chromosome (Drinan et al., 2018). Through the use of high-resolution genetic marker assays derived from the identified sex-linked loci (Drinan et al., 2018) it is now known that the Pacific halibut commercial catches in waters off Canada and the United States (including Alaska) are comprised of over 80% females (Stewart & Webster, 2021), providing direct evidence for sex-specific fishing pressure. Therefore, while promoting improved understanding of the mechanisms responsible for sex determination is a relevant aspect to the basic biology of the species, studies on sex determination in Pacific halibut have profound implications for fisheries management and conservation. In this study, we present the first chromosome-level reference genome assembly for Pacific halibut and utilize this genomic resource to identify genomic regions associated with sex in Pacific halibut using a pool sequencing approach. We identified a potential sex-determining region in chromosome 9 of Pacific halibut and, within this region, a potential candidate master sex-determining (MSD) gene. The genomic resource developed here will be essential to future genomic studies aimed at resolving population structure and dynamics and at understanding the evolutionary history of Pacific halibut populations.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling and genomic DNA extraction

Freshly sampled blood (500 µl) from an adult Pacific halibut female captured by angling from the mouth of Resurrection Bay near Seward, Alaska (59.80°N, 149.23°W) was mixed with 25 ml of TNE-S-Urea lysis buffer (4 M urea; 10 mM Tris-HCl, pH 7.5; 125 mM NaCl; 10 mM EDTA; 1% SDS; (Asahida et al., 1996)). From this blood sample, high-molecular weight genomic DNA (gDNA) was purified by phenol-chloroform extraction. A separate blood sample from the same Pacific halibut female (1.5 ml) was gently mixed with 270 µl of dimethyl sulfoxide (DMSO) in a 2 ml cryotube (15% DMSO) and



**FIGURE 1** (a) Photograph of a Pacific halibut (*H. stenolepis*). Photo credit: Guy Becken. (b) Heatmap of chromosome interaction intensity in the Pacific halibut Hi-C assembly. The x- and y-axis represent the length of the chromosomes

slowly cryopreserved in an isopropanol-containing Nalgene Mr Frosty Freezing container (Sigma) at  $-80^{\circ}\text{C}$  for generating a chromosome contact map (Hi-C) (see Section 2.2.3).

Fin clip samples from 30 adult female and 30 adult male Pacific halibut captured by a chartered commercial longline vessel in the Portlock Bank region of the Gulf of Alaska ( $56^{\circ}59'\text{N}$ - $58^{\circ}55'\text{N}$ ,  $148^{\circ}41'\text{W}$ - $152^{\circ}44'\text{W}$ ) were collected for whole-genome sequencing of pools of individuals (Pool-seq). Genomic DNA from fin clips was extracted using the DNeasy Blood & Tissue Kit (Qiagen) and the concentration was standardized across all samples and pooled in equimolar ratios by individual and sex, resulting in two gDNA pools, one for females and another for males. Ovarian and testicular samples were collected from the same fish, preserved in RNAlater (Invitrogen) and stored at  $-80^{\circ}\text{C}$  until processed for RNA extraction. Ovarian and testicular samples were also collected, fixed in 10% buffered formalin and processed for histological analyses to determine the reproductive stage, as described previously (Fish et al., 2020). Collected females and males for Pool-sequencing were at the early vitellogenic and early spermatogenic stages, respectively (data not shown).

## 2.2 | DNA library construction and sequencing

### 2.2.1 | Illumina sequencing

DNA library construction was performed according to the manufacturer's instruction using the Truseq DNA nano library prep kit (Illumina). Briefly, gDNA was quantified using the HS dsDNA Assay kit on the Qubit (Invitrogen). 200ng of gDNA were sonicated on a Bioruptor (Diagenode). Sonicated gDNA was end repaired and size selected on magnetic beads aiming for fragments of an average size

of 550bp. Selected fragments were adenylated on their 3' ends before ligation of Illumina's indexed adapters. The library was amplified using an eight cycle PCR and verified on a fragment analyser using the HS NGS fragment kit (Agilent). The library was quantified by qPCR using the KAPA Library quantification kit (Roche, ref. KK4824) and sequenced on half a lane of HiSeq2500 in paired-end  $2 \times 250\text{nt}$  using the clustering and SBS rapid kit following the manufacturer's instructions. This sequencing generated 57,277,046 raw clusters, leaving 56,057,418 clusters after Illumina's purity filtering.

### 2.2.2 | Nanopore sequencing

The quality and purity of gDNA were assessed using spectrophotometry, fluorometry and capillary electrophoresis. Additional purification steps were performed using AMPure XP beads (Beckman Coulter). All library preparations and sequencing were performed using Oxford Nanopore Ligation Sequencing Kit SQK-LSK109 according to the manufacturer's instructions "1D gDNA selecting for long reads." Genomic DNA (10  $\mu\text{g}$ ) was purified and then sheared to 20kb using the megaruptor system (Diagenode). A size selection step using a 10 kb cutoff was performed on the BluePippin Size Selection system (Sage Science) with the 0.75% agarose cassettes, Marker S1 high Pass 6–10 kb. For each library, a one-step-DNA-damage repair + ENDrepair + dA-tail-of-double-stranded-DNA-fragments procedure was performed on 2  $\mu\text{g}$  of DNA. Adapters were then ligated to DNAs in the library. Libraries were loaded onto R9.4.1 flowcells and sequenced on either a GridION or PromethION instrument at a concentration of 0.019 pmol for 48h and 0.013 pmol for 64h, respectively. The GridION flowcell produced 8.6 Gb of data and the PromethION flowcell produced 46.3 Gb of data.

### 2.2.3 | Hi-C sequencing

In situ Hi-C sequencing was performed as previously described (Foissac et al., 2019). Cryopreserved blood cells were thawed, washed twice with PBS and counted. Five million cells were then cross-linked with 1% formaldehyde in PBS, quenched with 0.125 mM glycine and washed twice with PBS. Membranes were disrupted using a Dounce pestle, and nuclei were permeabilized using 0.5% SDS and then digested with HindIII endonuclease. Five-prime overhangs at HindIII-cut restriction sites were filled-in, in the presence of biotin-dCTP with the Klenow large fragment, and then religated at an NheI restriction site. Nuclei were lysed and DNA was precipitated and then purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Qubit fluorometric quantification system (Thermo). T4 DNA polymerase was used to remove unligated biotinylated ends. Then, the Hi-C library was prepared according to Illumina's protocols using the Illumina TruSeq Nano DNA HT Library Prep Kit with a few modifications: DNA (1.4 µg) was fragmented to 550bp by sonication. Sheared DNA was then sized (200–600bp) using Agencourt AMPure XP beads, and biotinylated ligation junctions were captured using M280 Streptavidin Dynabeads (Thermo) and then purified using reagents from the Nextera Mate Pair sample preparation kit (Illumina). Using the TruSeq nano DNA kit (Illumina), the 3' ends of blunt fragments were adenylated. Next, adaptors and indexes were ligated and the library was amplified for 10 cycles. Library quality was assessed by quantifying the proportion of DNA cut by endonuclease NheI using a fragment analyser (Advanced Analytical Technologies). Finally, the library was quantified by qPCR using the Kapa Library Quantification Kit (Roche). Sequencing was performed on an Illumina HiSeq3000 apparatus (Illumina) using paired-end 2 × 150nt reads. This produced 51 million read pairs (1.5 Gb of raw nucleotides).

### 2.2.4 | Pool genome sequencing (Pool-seq) and analysis

Male and female pooled DNA libraries were constructed using the TruSeq Nano DNA HT Library Prep Kit (Illumina) following the manufacturer's protocol. Briefly, DNA from each male and female pool (200ng/pool) was sonicated using a M220 focused-ultrasonicator (COVARIS) and size-selected using SPB beads (kit beads) to retain fragments of ~550bp. Blunt DNA fragments were monoadenylated, ligated to indexes and Illumina's adapter and amplified using Illumina-specific primers. Libraries were checked for quality using a fragment analyser (Advanced Analytical Technologies Inc.) and quantified by qPCR using the Kapa Library Quantification Kit (Roche Diagnostics Corp.). Sequencing of male and female Pool-seq libraries was performed on a NovaSeq S4 lane (Illumina) using a paired-end read length of 2 × 150nt mode Illumina NovaSeq reagent kits following the manufacturer's instructions. The run produced 105 million read pairs for the male pool library and 96 million read pairs for the female pool library. Reads from the male and female pools were analysed as

previously described (Adolfi et al., 2021; Imarazene et al., 2021; Pan, Feron, et al., 2021), using the pooled sequencing analyses for sex signal (PSASS; <https://github.com/SexGenomicsToolkit/PSASS>), a software that compares pooled sequencing data sets from two sexes, and our Pacific halibut assembly as a reference assembly. PSASS was run with default parameters to compute three metrics in 50kb non-overlapping windows over the Pacific halibut genome: (1)  $F_{ST}$  between males and females over the window, (2) density of sex-specific SNPs, defined as SNPs heterozygous in one sex and homozygous in the other sex, and (3) read depth for each sex.

### 2.2.5 | Estimation of linkage disequilibrium on chromosome 9

In order to investigate patterns of linkage disequilibrium (LD) throughout the Pacific halibut genome, the restriction-site associated DNA (RAD) sequence data generated by Drinan et al. (2018) was reanalysed using our Pacific halibut assembly as a reference. Complete details regarding RAD sequencing and raw data processing can be found in Drinan et al. (2018). Trimmed sequence reads were aligned to the Pacific halibut reference using minimap2 (v2.23) (Li, 2018) and invoking the short read preset option (`-x sr`). Samtools (v1.14) (Li et al., 2009) was used to sort and filter unmapped reads and alignments with mapping qualities (MAPQ) less than 20. The reference aligned data were processed using stacks (v2.60). First, the gstacks module was used to create RAD loci, identify SNPs, and genotype individuals while removing PCR duplicates. Then, the populations module was used to filter loci and export a vcf file of variants. Loci were retained for further analysis if they had a minor allele frequency (MAF) of at least 0.01 and genotypes for at least 20% of the individuals, these filters were applied using the `--filter-haplotype-wise` option. Patterns of linkage disequilibrium were analysed separately for males and females. First, vcftools (v0.1.16) (Danecek et al., 2011) was used to apply more stringent SNP filters to each data set separately, requiring a MAF of at least 0.05 and genotypes for at least 80% of the individuals within each data set. Plink (v1.9) (Chang et al., 2015; [www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/)) was used to calculate pairwise  $r^2$  for all SNPs located on Chr 09. Pairwise  $r^2$  values were summarized in nonoverlapping 100kb bins by taking the median within each bin and plotting.

## 2.3 | Genome assembly and analysis

GridION and PromethION data were trimmed using porechop version 0.2.1 (Wick, 2017/2019) and assembled using Flye version 2.9 (Kolmogorov et al., 2019) with default parameters. The assembly base pair quality was improved by two polishing steps, including one round of long read alignment to the draft genome with minimap2 version 2.7 (Li, 2018) followed by samtools sort and index version 1.10 then by racon version 1.3.1 (Vaser et al., 2017), as well as one round of short read alignments using bwa mem version 0.7.17



followed by samtools sort and index and Hapo-G version 1.2 (Aury & Istace, 2021). Default parameters were used for all these software packages. The polished genome assembly was then scaffolded using Hi-C as a source of linking information. Reads obtained by in situ Hi-C sequencing (see Section 2.2.3) were aligned to the draft genome using juicer (Durand et al., 2016) with default parameters. A candidate assembly was then generated with the 3D de novo assembly (3D-DNA) pipeline (Dudchenko et al., 2017) with the -r 0 parameter. Finally, the candidate assembly was manually reviewed using juicebox assembly tools (Durand et al., 2016). Genome completeness was estimated using Benchmarking universal single-copy orthologues (BUSCO) version 4.0 (Simão et al., 2015) based on 3640 BUSCO orthologues derived from the Actinopterygii lineage.

## 2.4 | Genome annotation

The first annotation step was to identify repetitive content using RepeatMasker v4.0.7 (Tarailo-Gravoac & Chen, 2009), DustMasker of NCBI toolkit v18 (described in Morgulis et al., 2006), and TandemRepeatFinder version 4.09 (Benson, 1999). A species-specific de novo repeat library was built with RepeatModeler version 1.0.11 (<http://www.repeatmasker.org/RepeatModeler/>) and repeated regions were located using RepeatMasker with the de novo and *Danio rerio* libraries. Bedtools version 2.26.0 (Quinlan & Hall, 2010) was used to aggregate repeated regions identified with the three tools and to soft mask the genome. 23.3% of the assembly was soft masked.

The second annotation step aimed to generate a transcriptome annotation using Maker3 genome annotation pipeline version 3.01.02-beta (Holt & Yandell, 2011). It combines annotations and evidence from three approaches: (1) similarity with fish proteins, assembled transcripts (see below), and de novo gene predictions. Protein sequences from 11 fish species (*Astyanax mexicanus*, *D. rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Xiphophorus maculatus*) found in Ensembl were aligned to the masked genome using exonerate version 2.2.0 (Slater & Birney, 2005) with the alignment model protein2genome that allows translated alignments with modelling of introns. (2) RNA-seq reads from 12 tissues (white muscle, liver, ovary, testis, red muscle, heart, pituitary, brain, spleen, gill, retinas and head kidney) (Jasonowicz et al., 2022) were used for genome annotation and aligned to the chromosomal assembly using HISAT2 version 2.2.1 (Kim et al., 2019). Bam files were merged using samtools merge and signal wiggle files were generated using STAR version 2.5.1b (Dobin et al., 2013) in run mode inputAlignmentsFromBAM with outWigType and outWigStrand options. Cufflinks version 2.2.1 (Trapnell et al., 2010) was used to assemble the transcripts, which were used as RNA-seq evidence. (3) Braker version 2.0.4 (Hoff et al., 2016) provided de novo gene models with wiggle files provided by STAR as hint files for GeneMark and Augustus training. The best supported transcript for each gene

was chosen using the quality metric called annotation edit distance (AED) (Eilbeck et al., 2009). Genome annotation gene completeness was assessed by BUSCO using the Actinopterygii group. From the 52 scaffolds of the genome assembly, 40 were annotated. These annotated scaffolds were kept for further analyses.

## 2.5 | Real-time quantitative PCR

The mRNA expression levels of *bmpr1ba* in testis and ovary from adult Pacific halibut were measured by real-time quantitative PCR (rt qPCR) and normalized to the mRNA expression levels of two housekeeping genes: elongation factor 1-alpha 1 (*eef1a1*) and ubiquitin (*ubq*). The housekeeping genes *eef1a1* and *ubq* were identified by geNorm (Vandesompele et al., 2002) and NormFinder (Andersen et al., 2004) as the most suitable in both sets of samples. Primers to *bmpr1ba*, *eef1a1* and *ubq* were designed in Primer3 using Geneious (<https://www.geneious.com>) (Table S1). RNA was extracted from testis ( $N = 7$ ) and ovary ( $N = 10$ ) samples from fish used in Pool-Seq using the PureLink RNA Mini Kit (Invitrogen) with  $\beta$ -mercaptoethanol (ThermoFisher Scientific) added to the lysis buffer at a final concentration of 4%. One microgram of total RNA from each sample was treated with DNase I amplification grade (Life Technologies) to remove contaminating gDNA. Resulting RNA concentrations were measured on a NanoDrop 8000 spectrophotometer (Thermo Scientific) and 300 ng of each RNA sample were reverse transcribed using the SuperScript IV VILO Master Mix cDNA kit with ezDNase (Invitrogen) using the protocol specified by the manufacturer. All PCR reactions were run in a QuantStudio 6 real-time PCR system (Applied Biosystems) and followed the requirements of the MIQE guidelines. Each sample was run in triplicate 10  $\mu$ l reactions with 2  $\mu$ l of cDNA (diluted 1:20), 400 nM of each primer, and PowerUp SYBR Green Mastermix (Applied Biosystems) using the manufacturer's standard PCR and melt curve protocols. The specificity of the reaction and absence of primer dimers were confirmed by the melt curve profile, and products of new primer pairs were run on a 1.5% agarose gel to confirm expected amplicon size. Primer efficiency was calculated by analysing serial dilutions of cDNA samples. The expression level of *bmpr1ba* was normalized to the geometric mean of the two reference genes ( $M < 0.5$ ) and calculated using the  $2^{-\Delta\Delta C_T}$  method (Livak & Schmittgen, 2001).

## 2.6 | Chromosome synteny among flatfish genomes

D-genies (Cabanettes & Klopp, 2018) was used to compare the Pacific halibut genome assembly (RefSeq: GCF\_022539355.2) to representative genome assemblies that are publicly available for other flatfish species: Atlantic halibut (*H. hippoglossus*) (RefSeq: GCF\_009819705.1), Greenland halibut (*Reinhardtius hippoglossoides*) (GenBank: GCA\_006182925.3), half-smooth tongue sole (*C. semilaevis*) (RefSeq: GCF\_000523025.1), Senegalese sole (*S. senegalensis*)

(RefSeq: GCF\_019176455.1), Japanese flounder (*Paralichthys olivaceus*) (RefSeq: GCF\_001970005.1), and turbot (*S. maximus*) (RefSeq: GCF\_013347765.1). Alignments were performed in D-genies using the minimap2 aligner treating the Pacific halibut genome as the query reference and the other flatfish genomes as the target reference. Only primary alignments and alignments between assembled chromosomes were retained for further analysis. Coverage of the other flatfish genomes by our Pacific halibut assembly was estimated by first calculating the length of each alignment on the target genome determined by the start and end positions of the target alignments, then summing all alignment lengths, and dividing by the size of the target genome. Syntenic relationships among flatfish species for which a RefSeq annotation is available were also explored by MCscan implemented in the python package jcv (Tang et al., 2008). This approach differs from that implemented in D-genies by detecting groups of orthologous genes and comparing their arrangement to identify colinear segments in the genomes being compared. MCscan was used to explore microsyntenic relationships, focusing on the local arrangement of genes located near the bone morphogenetic protein receptor type-1B (*bmpr1ba*) gene in Pacific halibut and other flatfish species. Spliced coding sequences were generated for each transcript using gffread (v0.12.6) (Pertea & Pertea, 2020). Orthologous coding sequences were identified using jcv's compara module requiring a C-score of 0.99 and syntenic gene blocks were identified by setting the minspan parameter to 30 genes. The microsyntenic arrangement of genes in the region near the *bmpr1ba* gene were also investigated by examining the order and orthologous relationships of genes within a five gene window upstream and downstream of *bmpr1ba* in Pacific halibut and other flatfish species.

## 2.7 | Bone morphogenetic protein receptor type-1B sequence comparison

Multiple sequence alignments of the *Bmpr1ba* protein sequences were done using the MUSCLE aligner (v3.8.1551) (Edgar, 2004). Sequences from the flatfish *Bmpr1ba* proteins were also compared to sequences from zebrafish (*D. rerio*), and medaka (*O. latipes*). A phylogenetic tree was generated using the web version of PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>) (Guindon et al., 2010). The substitution model was selected using the Smart Model Selection (SMS) (Lefort et al., 2017) implementation in PhyML using Akaike information criterion (AIC) as a model selection criterion. Branch support values were estimated using approximate likelihood-ratio test for branches (aLRT) (Anisimova & Gascuel, 2006).

## 2.8 | Regulatory elements of the bone morphogenetic protein receptor type-1ba gene

Putative transcription factor (TF) binding sites were identified by scanning the genomic region containing the *bmpr1ba* gene and 5 kb of upstream sequences using the R package TFBSTools (v1.26) (Tan

& Lenhard, 2016). Transcription factor binding profiles were obtained as position weight matrices from the JASPAR 2020 database (Fornes et al., 2019). Two scans were carried out. The JASPAR POLII collection was used to identify binding sites linked to RNA polymerase II core promoters. Additionally, a scan using vertebrate binding profiles from the JASPAR CORE collection was carried out to identify binding sites for other TFs that may play a role in the expression of the *bmpr1ba* gene. To evaluate the potential for functional significance of the sex-specific SNPs, their location was compared to the structure of the *bmpr1ba* gene (intron, exon, etc.) and the potential transcription factor binding sites identified. First, the *bmpr1ba* gene annotation was manually curated as (1) 5'UTR, defined as the transcript upstream of start codon in the annotation, (2) 3'UTR, defined as the transcript downstream of the stop codon in the annotation, (3) Intron-exon boundaries, defined using the annotation, and (4) 5' flanking region, defined as the 5 kb region upstream of 5'UTR (as indicated above).

Sex-specific SNPs within the *bmpr1ba* gene and 5 kb of upstream sequences were compared to the manual curation of the *bmpr1ba* gene to identify the gene regions containing SNPs and then this list was joined to the list of putative TF binding sites using bedtools (v2.29.2) (Quinlan & Hall, 2010). With this, the sex-specific SNP positions were compared to putative TF binding sites within the *bmpr1ba* gene to assess potential changes in the transcriptional regulation of this gene due to the presence of SNPs in regulatory regions.

## 2.9 | Functional annotation of Pacific halibut genes and gene ontology analysis

To infer the potential function of genes identified in the Pacific halibut genome annotation, InterProScan (v5.52-86.0) (Jones et al., 2014) was used to screen all protein products identified in the genome annotation against the InterPro consortium member databases (Blum et al., 2021). Gene ontology (GO) terms for the InterPro scan matches were assigned to genes in Pacific halibut genome annotation. The R (v4.0.5) (R Core Team, 2019) packages ViSEAGO (v1.4.0) (Brionne et al., 2019) and topGO (v2.44.0) (Alexa & Rahnenfuhrer, 2020) were used to identify enriched GO terms associated with genes in the putative sex determining region of chromosome 9 using GO terms associated with the entire genome as a background. GO terms were scored using the classic algorithm and fisher test implemented in topGO for all three GO domains (cellular component, biological process, and molecular function).

## 2.10 | Mapping of RAD-tags associated with sex to the Pacific halibut genome assembly

Drinan et al. (2018) used restriction-site associated DNA (RAD) sequencing to identify 56 RAD-tags associated with sex in Pacific halibut. The consensus sequences for these RAD-tags were aligned to Pacific halibut genome assembly using minimap2 (v2.17) using the short read preset (-x sr).

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Generation of a chromosome-level assembly of the Pacific halibut genome

We have generated the first high-continuity, chromosome-length de novo assembly of the Pacific halibut genome ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_022539355.2](https://www.ncbi.nlm.nih.gov/assembly/GCF_022539355.2)) by combining Oxford Nanopore Technologies (ONT) long reads, Illumina short linked reads, and a chromosome contact map (Hi-C). The complete genome assembly after the Hi-C integration step yielded 52 scaffolds with a total genome size of 602.15 Mb, an  $N_{50}$  scaffold length of 27.29 Mb and an  $L_{50}$  scaffold number of 11 (Table 1). The complete genome assembly includes 24 chromosome-length scaffolds that represent 99.78% of the complete genome sequence. Based on the discrete blocks identified in the Hi-C contact map (Figure 1b) and the reported 24 chromosomes in Atlantic halibut, as determined by karyotype (Brown et al., 1997) and linkage analyses (Reid et al., 2007) and genome sequencing (Edvardsen et al., 2022; Einfeldt et al., 2021), we propose that the 24 chromosome-length scaffolds identified here correspond to the 24 chromosomes in Pacific halibut. Pacific halibut chromosome sizes range from 11.27 to 32.85 Mb (Table S2). Genome assembly completeness, as estimated by BUSCO in genome mode with the Actinopterygii lineage as reference, is demonstrated by the high percentage (98.4%) of complete BUSCOs. By using protein, transcript and de novo gene prediction analyses we annotated 26,772 protein-coding genes in the Pacific halibut genome, whereas the annotation conducted by the NCBI Eukaryotic Genome Annotations Pipeline (NCBI Hi Annotation Release 100 [[https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Hippoglossus\\_stenolepis/101/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Hippoglossus_stenolepis/101/)];) yielded 27,944 annotated genes.

TABLE 1 Pacific halibut (*H. stenolepis*) genome assembly statistics and assembly completeness

		Complete assembly	Chromosomes only
Assembly metrics	Number of scaffolds	52	24
	Total size of scaffolds	602,147,347	600,901,890
	Longest scaffold	32,845,235	32,845,235
	Shortest scaffold	582	11,267,222
	Mean scaffold size	11,579,757	25,037,579
	Median scaffold size	184,878	25,468,064
	$N_{50}$ scaffold length	27,288,303	27,288,303
	$L_{50}$ scaffold count	11	11
	% of assembly in chromosomes	–	99.78%
	% of assembly in unanchored scaffolds	–	0.22%
Assembly completeness	Complete BUSCOs (C)	3580 (98.4%)	
	C and single-copy BUSCOs	3540 (97.3%)	
	C and duplicated BUSCOs	40 (1.1%)	
	Fragmented BUSCOs	17 (0.5%)	
	Missing BUSCOs	43 (1.1%)	

Compared to other related flatfish species with sequenced genomes, the size of the Pacific halibut genome assembly is most similar to that of Atlantic halibut (596.8 Mb) (Einfeldt et al., 2021) and Greenland halibut (598.5 Mb) (Ferchaud et al., 2022), followed by turbot (568 Mb) (Figueras et al., 2016), Japanese flounder (546 Mb) (Shao et al., 2017) and half-smooth tongue sole (477 Mb) (Chen et al., 2014). The quality of the Pacific halibut genome assembly was also comparable to that of some closely related species that have been sequenced recently, like Atlantic and Greenland halibut (Table 2). The genomic coverage of the Atlantic and Greenland halibut genomes by the Pacific halibut assembly was very high (99.31 and 99.60%, respectively) (Figure S1). Furthermore, a high degree of nucleotide level synteny was observed between Pacific halibut and the other flatfish genomes, with the highest levels of synteny observed between its closest relatives in the subfamily Hippoglossinae (*H. hippoglossus* and *R. hippoglossoides*) (Atta et al., 2022) (Figure S2). Inspection of the dotplots identified orthologous chromosomes for all Pacific halibut chromosomes in both the Atlantic halibut and Greenland halibut genome assemblies (Figure S2) that is likely a reflection of their recent common evolutionary history (Grant, 1987; Grant et al., 1984).

#### 3.2 | Identification of the sex-determining region in the Pacific halibut genome

In order to identify the sex-determining (SD) region and putative MSD gene in Pacific halibut we set out to detect whole genome sex differences by Pool-seq analysis (Schlotterer et al., 2014). We mapped the Pool-seq reads from male and female pools to the Pacific halibut genome assembly to characterize genomic regions containing sex-specific single nucleotide polymorphism (SNP) differences. By conducting whole genome analysis of SNP distribution, we identified a

clear sex-linked region in Pacific halibut females on chromosome 9 (Chr9) alone (Figure 2). This large region of approximately 12 Mb in size (Chr9: 14,000,000–26,000,000bp) shows differentiation between the two sexes, contains a high-density of female-specific SNPs that peak in regions of high differentiation (Figure 3) and is very likely the SD region in Pacific halibut. From a total of 20,048 sex-specific SNPs identified in this region, 18,452 SNPs were female-specific (strictly heterozygotes in females and homozygotes in males) and 1596 SNPs were male-specific. In addition, increased LD was observed in this region (Figure 4). Therefore, we propose that the female-specific region in Chr9 likely corresponds to the nonrecombining region on the sex chromosome in Pacific halibut. Read coverage between males and females was similar throughout Chr9 including the SD region, suggesting low levels of sex chromosome degeneration despite the accumulation of female-specific SNPs. However, close examination of the SD region revealed a small 18 kb region at Chr9:17,625,400–17,643,850 with an approximately four-fold excess of female over male reads (Figure 3d). This region is probably a nonresolved W-specific copy number variation (CNV) region in the assembly that contains one complete gene annotated as kinetochore protein *ndc80* (*ndc80*) and a 3' end truncated gene annotated as nipped-B-like protein A (*nipbla*) (Figure S3). Mapping of previously identified Pacific halibut RAD-tags associated with sex (Drinan et al., 2018) to the Pacific halibut genome assembly resulted in the alignment of 55 of the 56 RAD-tags, all of which mapped to the putative SD region (14,601,685–25,693,202bp) on Chr9 (NC\_061491.1) (Figure 3a; Figure S4). The large number of sex-associated RAD-tags mapping to this region of Chr9 provides additional support that this region of the genome may play a key role in the sexual determination system of Pacific halibut.

Whole-genome macrosyntentic relationships between Pacific halibut and other flatfish species show that Chr9 of Pacific halibut is orthologous to Chr9 of Atlantic halibut, Greenland halibut and turbot, linkage group (LG) 5 of Senegalese sole, and to both Z and W chromosomes of half-smooth tongue sole (Figure 5). In contrast to the regions of Pacific halibut Chr9 that map to the half-smooth tongue sole sex chromosomes, highlighting the importance of this genomic region in the sex determination system of Pacific halibut, the reported SD regions in the genomes of Atlantic halibut, Greenland halibut, turbot and Senegalese sole are present in chromosomes or LGs that are not orthologous to Chr9 of Pacific halibut: Chr12 in Atlantic halibut (Einfeldt et al., 2021), Chr10 and 21 in Greenland halibut (Ferchaud et al., 2022), Chr12 in turbot (Martinez et al., 2021) and LG18 in Senegalese sole (Guerrero-Czar et al., 2021). These differences in SD chromosomes among phylogenetically related flatfish species are consistent with the known rapid turnover of sex chromosomes in teleosts, attributed to allelic diversification or gene duplication and translocation events associated with the emergence of new MSD genes (Kikuchi & Hamaguchi, 2013; Pan et al., 2018). Closer inspection of the alignment of Chr9 in Pacific halibut to orthologous chromosomes in the other Hippoglossinae species revealed the presence of structural differences between these species in this region of their genome (Figure 6). Both species contain inversions on Chr9 but only the inversion in Atlantic halibut overlaps with the SD region of Pacific halibut (Figure 6). It has been discussed

that inversions offer a potential mechanism for the suppression of recombination which is often implicated in sex chromosome evolution (Furman et al., 2020). This is supported by our analysis of LD, which identified a large block of SNPs showing higher levels of LD on Chr9 in females than in males (Figure 4), suggesting a reduced recombination rate for females in this region of the Pacific halibut genome. Given the relatively recent divergence between Pacific and Atlantic halibut (Grant, 1987; Grant et al., 1984), we hypothesize that the increased levels of LD on Chr9 in Pacific halibut, coupled with the structural differences that exist between the SD region of Pacific halibut and the orthologous region of Chr9 of Atlantic halibut, could be associated with the early stages of sex chromosome degradation; however, further studies are needed to shed light onto this issue.

### 3.3 | Search for a putative sex-determining gene in Pacific halibut

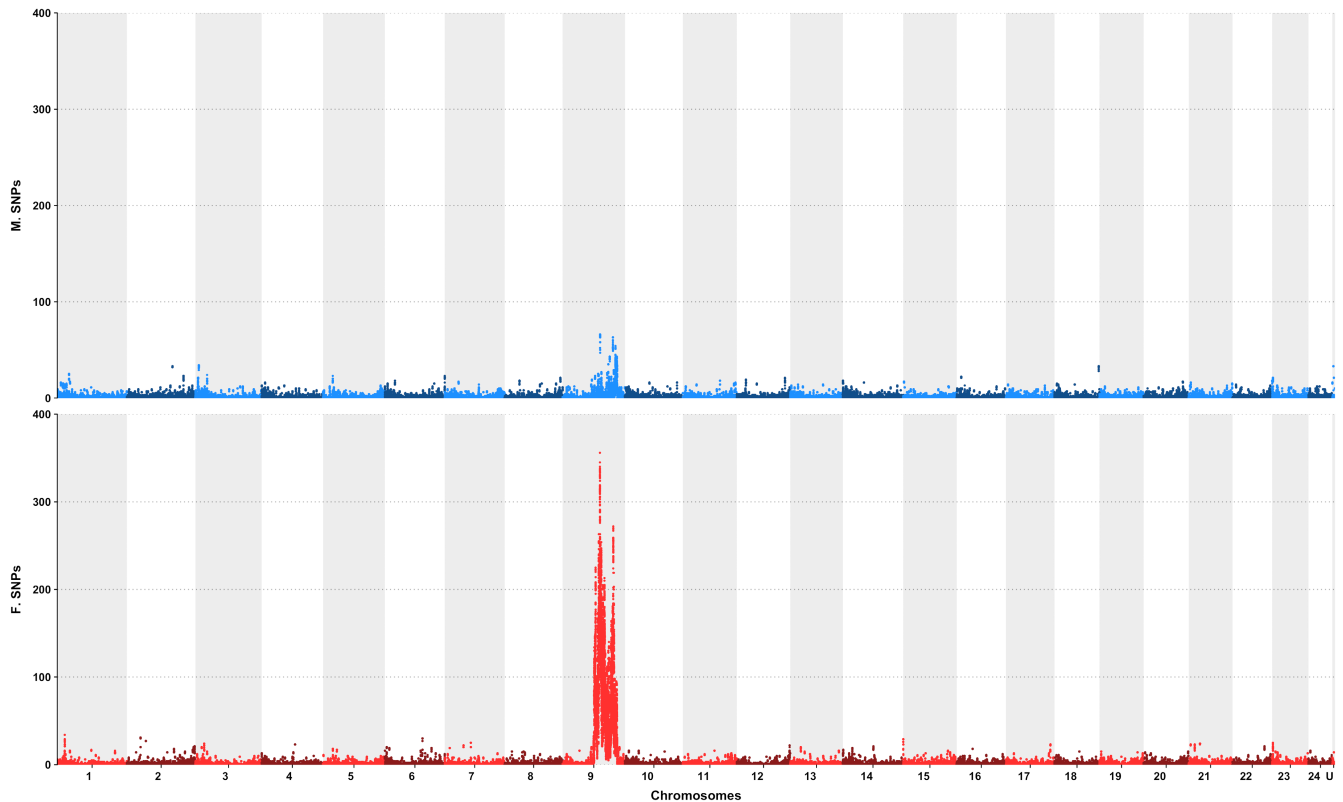
We next searched the putative SD region in Chr9 for potential candidate MSD genes in Pacific halibut. We compared the genome annotation to the locations of sex-specific SNPs in the putative SD region and identified 482 annotated genes that contained a total of 13,949 SNPs, with 12,915 female-specific and 1034 male-specific SNPs (Table S3). Enrichment analysis of GO terms associated with genes in the SD region of Chr9 yielded several significantly enriched GO terms (biological process [BP]: 26; cellular component: 4; molecular function [MF]: 19), including “regulation of translational fidelity” as the top BP term and “hormone receptor activity” and “growth factor receptor binding” among the top MF terms (Table S4). Although several annotated genes containing a large number of sex-specific SNPs were identified in the SD region of Chr9, most notably netrin receptor *UNC5C* (*unc5ca*), glutamate receptor ionotropic, delta-2 (*grid2*) and transmembrane protein 132C, with 702, 665 and 493 SNPs, respectively, none of these genes corresponded to known MSD genes in vertebrates.

However, by specifically searching among all the annotated genes present in this region for members of the few gene families known to have produced MSD genes in teleost species (e.g., SOX, DMRT, TGF- $\beta$ ) (Pan et al., 2018), we identified the bone morphogenetic protein receptor-type IB (*bmpr1ba*) gene in the SD region at position 17,407,676–17,433,892bp on Chr9, located ~2.2 kb downstream of the largest  $F_{ST}$  peak (Figure 3). The *bmpr1ba* gene is a member of the TGF- $\beta$  superfamily of polypeptide growth factors and, therefore, is part of a complex TGF- $\beta$  signalling pathway whose components have been recruited to play important roles in sex determination during teleost evolution (Pan, Kay, et al., 2021). The *bmpr1ba* gene codes for a type I transmembrane serine/threonine kinase receptor that upon ligand binding is activated by type II receptors (e.g., Amhr2) triggering a downstream signal, as part of the characteristic hexameric signalling complex through which TGF- $\beta$  family ligands, such as anti-Müllerian hormone (Amh) or bone morphogenetic protein (Bmp), exert their function (Miyazono et al., 2005). Examples of teleost MSD genes belonging to the TGF- $\beta$  family include genes coding for ligands and receptors that are derived either by gene duplication or



**TABLE 2** Comparative analysis of assembly metrics between Pacific halibut (*H. stenolepis*) and the two other Pleuronectoidei species with recent chromosome-level genome assemblies: Atlantic halibut (*H. hippoglossus*) and Greenland halibut (*R. hippoglossoides*)

Assembly metrics	Pacific halibut	Atlantic halibut	Greenland halibut
Genome size (Mb)	602.1	596.8	598.5
Chromosome-level scaffolds mapped to total assembly (%)	99.8	99.8	96.3
N <sub>50</sub> (Mb)	27.3	26.3	25.0
Complete BUSCO (%)	98.5	96.9	96.5

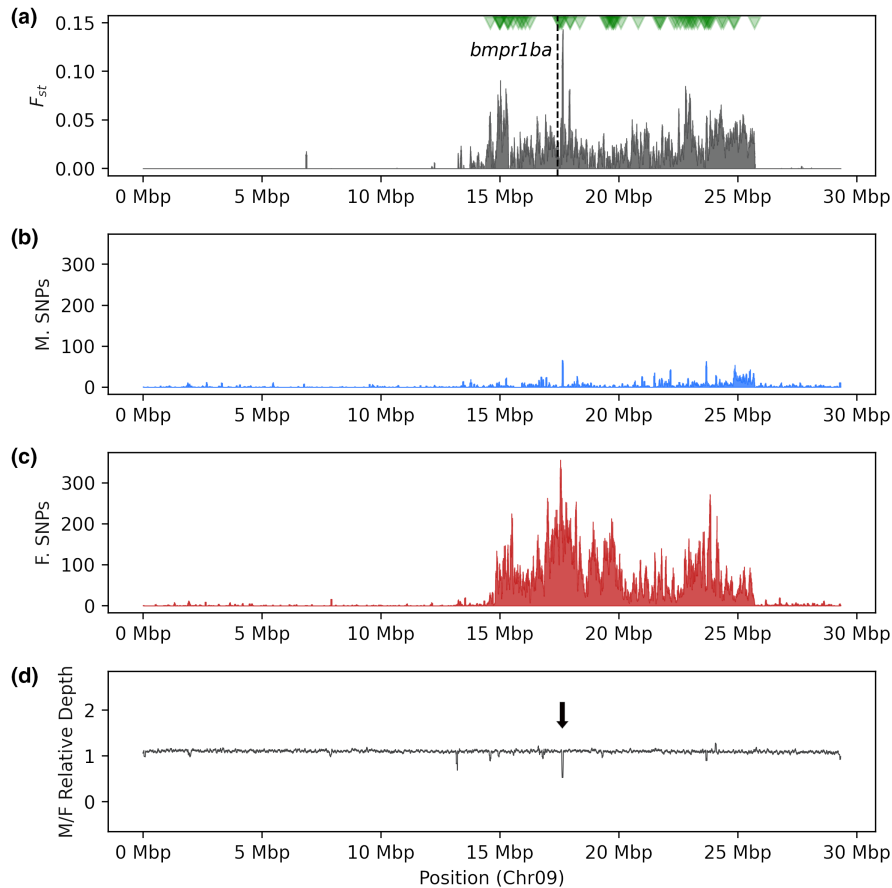


**FIGURE 2** Genome-wide Manhattan plot visualization of male-specific (in blue, top) and female-specific (in red, bottom) single-nucleotide polymorphisms (SNPs) along the 24 chromosomes of the Pacific halibut genome assembly. Sex-specific SNPs are represented as dots (total per 50 kb window size) of alternating colours to distinguish their location on adjacent chromosomes. U indicates unplaced chromosomes

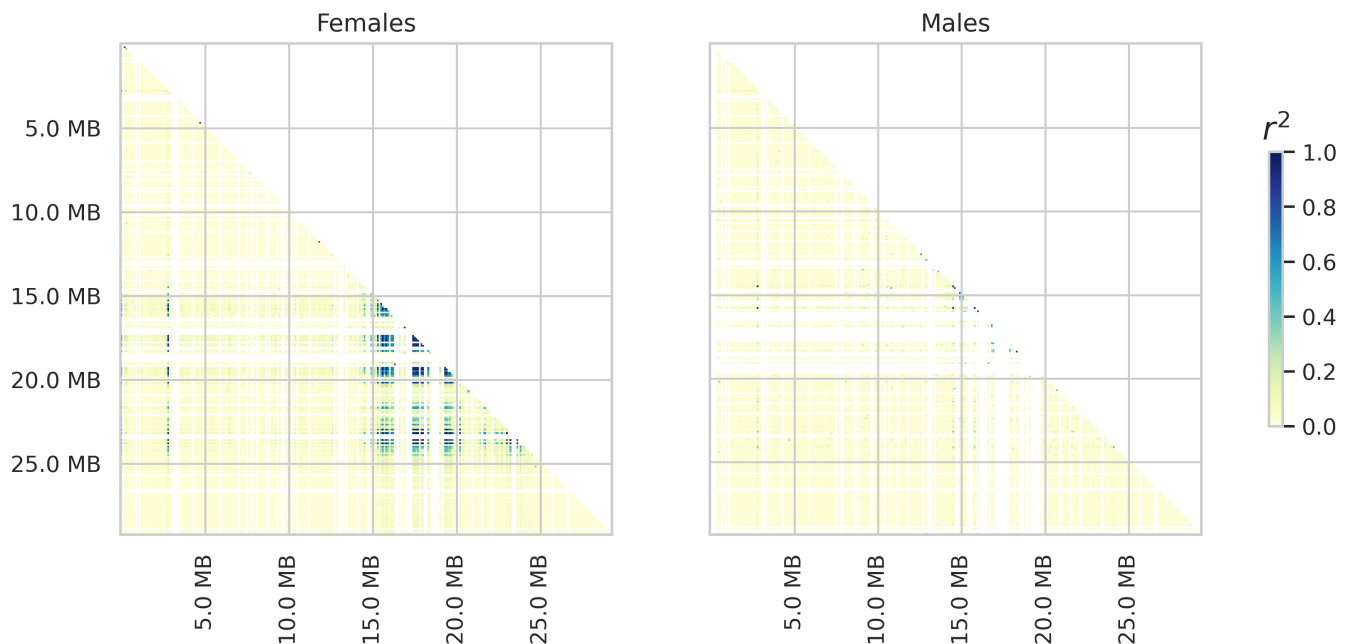
allelic diversification from *amh* (*amhY* and *amhBY*), from *amh* receptor II or *amhr2* (*amhr2Y* and *amhr2BY*), from gonadal soma-derived growth factor or *gsdf* (*gsdfY*), from growth differentiation factor 6 or *gdf6* (*gdf6Y* and *gdf6BY*) and from *bmp* receptor 1bb or *bmpr1bb* (*bmpr1bbY*) (reviewed in Pan, Kay, et al., 2021). Interestingly, *bmpr1ba* is a paralogue of *bmpr1bbY*, a gene encoding a truncated form of a BMP type I receptor with constitutive activity that originated by gene duplication and that is the candidate MSD gene in Atlantic herring (*Clupea harengus*) (Rafati et al., 2020). Therefore, we propose that *bmpr1ba* is a potential candidate MSD gene in Pacific halibut that, if confirmed, would represent the first potential MSD gene coding for a receptor to TGF- $\beta$  family ligands identified in a flatfish species. Other candidate MSD genes identified in flatfish species include *dmrt1* in half-smooth tongue sole (Chen et al., 2014), *gsdf* in Atlantic halibut (Einfeldt et al., 2021) and *sox2* in turbot (Martinez et al., 2021) as well as in Greenland halibut (Ferchaud et al., 2022). Like species of the genus *Oryzias* and *Gasterosteus*, Pacific halibut and Atlantic

halibut, with female and male heterogametic systems, respectively, provide another example of closely related teleost species that have recruited different sex-linked chromosomes and different MSDs, further establishing the lack of conservation of sex determining mechanisms in teleost fish (Pan et al., 2016). It is conceivable that the relatively recent evolutionary divergence of Atlantic halibut from Pacific halibut was accompanied by a shift from female to male heterogamety driven possibly by the emergence of a new MSD. A shift between male and female heterogamety driven by the rise of a new dominant female determinant was recently proposed for certain species of the Esocidae family (Pan, Feron, et al., 2021).

We found a total of 115 female-specific SNPs in the *bmpr1ba* gene (Table S5), with only four SNPs located within exons that are synonymous; therefore, no changes in the sequence of the translated protein nor its function are expected between males and females. In absence of obvious sequence characteristics of the product of the *bmpr1ba* gene that could support a potential role for this gene



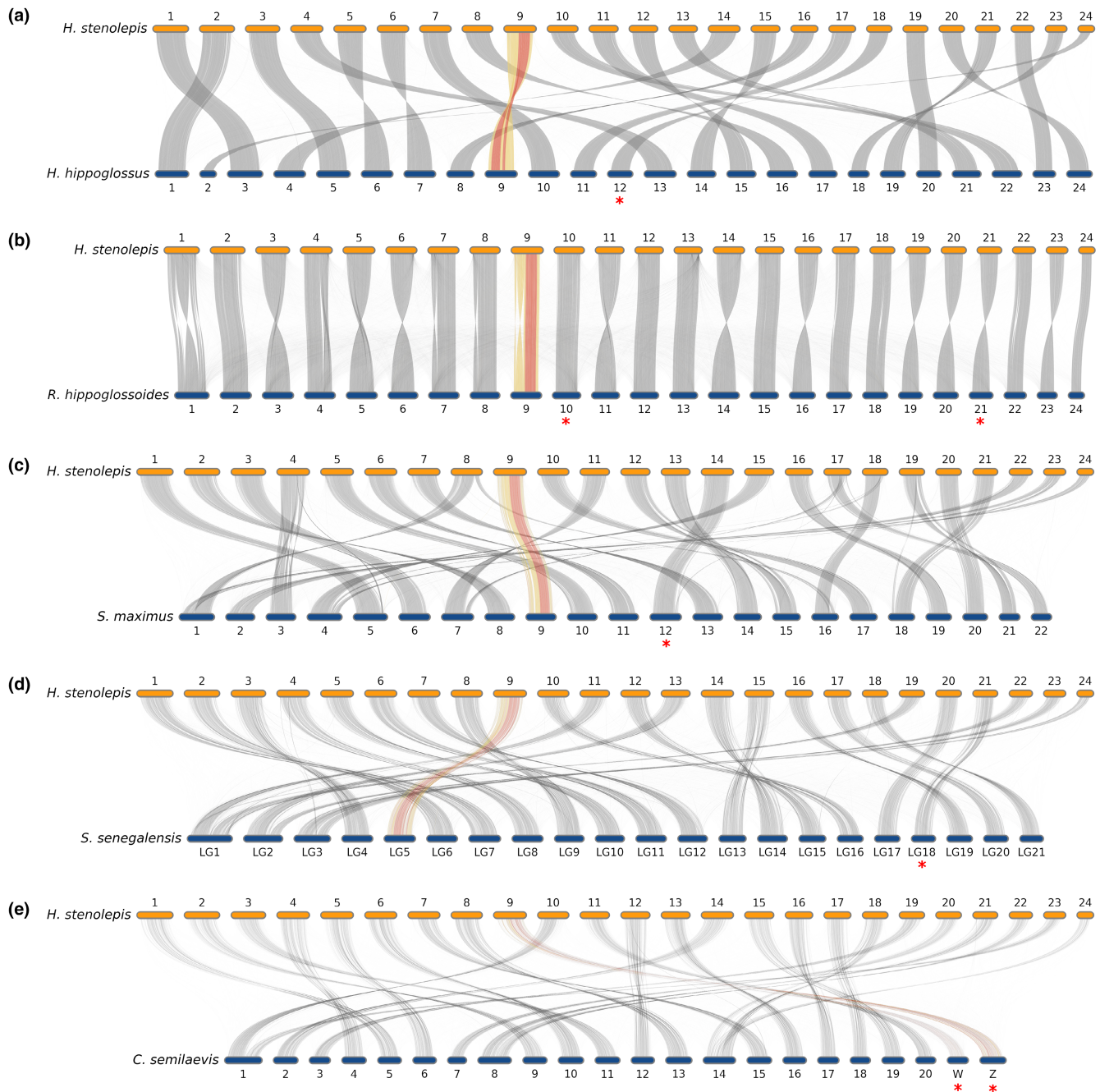
**FIGURE 3** Distribution of female-specific SNPs on chromosome 9 (Chr9).  $F_{ST}$  values are shown in (a). SNPs were counted using 50 kb sliding window with an output point every 1000 bp and male (M)- and female (F)-specific SNPs were indicated by blue and red colours, respectively (b, c). A large sex-determination region was identified on Chr9. Read depth ratio between males and females is shown in (d). The position of the *bmpr1ba* gene is indicated in (a) by a vertical dashed line. The location of RAD-tags associated with sex identified by Drinan et al. (2018) are indicated by inverted green triangles in (a). The location of the small 18 kb region within the sex-determination region that contains an excess of female over male reads is indicated in (d) by an arrow



**FIGURE 4** Linkage disequilibrium ( $r^2$ ) for RAD-seq derived SNPs on chromosome 9. Information for females and males is shown in the left and right panels, respectively. Colour represents the median  $r^2$  for all comparisons in nonoverlapping 100 kb bins

in driving sex determination, we next investigated whether *bmpr1ba* gene expression differed between males and females by measuring the mRNA expression levels of *bmpr1ba* in male and female gonads from adult Pacific halibut by qPCR. Our results show similar levels of

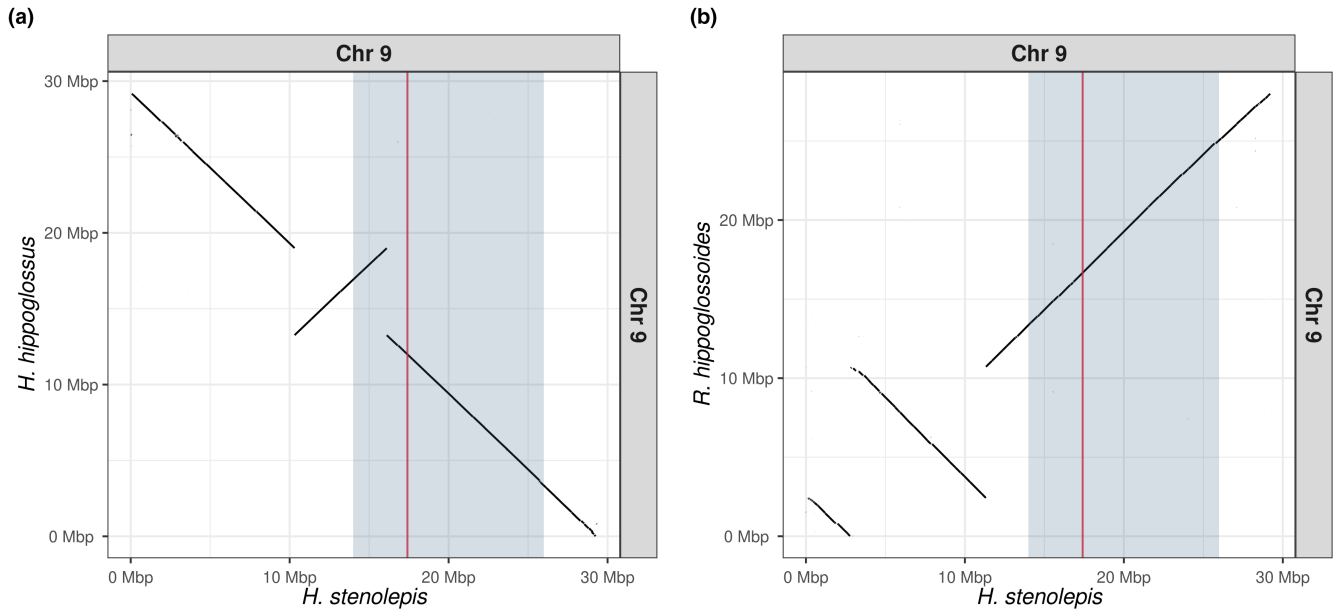
*bmpr1ba* mRNA expression between ovaries and testes from adult fish (Figure S5). These results are not surprising given that sex-related differences in expression levels of candidate MSD genes are most likely to occur during early gonad development at the sex determination



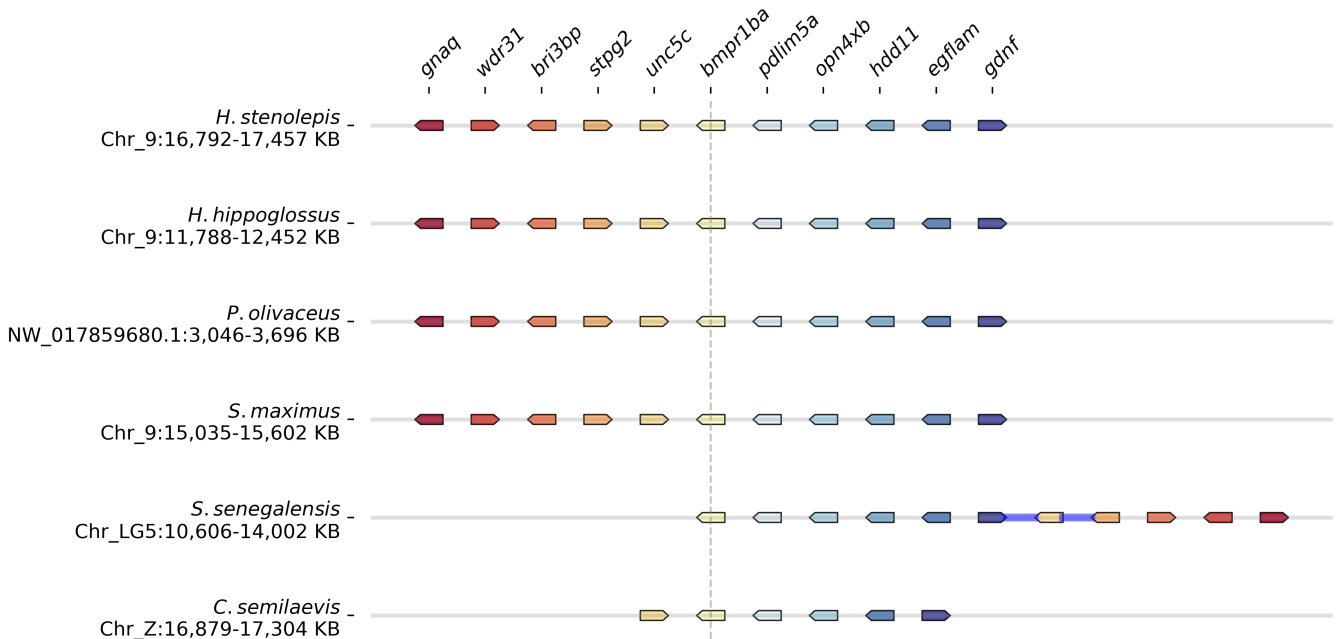
**FIGURE 5** Whole-genome macrosyntentic relationships between Pacific halibut (*H. stenolepis*) and Atlantic halibut (*H. hippoglossus*) (a), Greenland halibut (*R. hippoglossoides*) (b), turbot (*S. maximus*) (c), Senegalese sole (*S. senegalensis*) (d) and half-smooth tongue sole (*C. semilaevis*) (e). Orthologous relationships to chromosome 9 of Pacific halibut are highlighted in yellow and orthologous relationship to the sex determining region are highlighted in red. Red asterisks indicate the chromosomes linked to sex in the other flatfish species

stage. Unfortunately, the lack of Pacific halibut captive broodstock to produce early embryos prevents us from testing this hypothesis. However, we searched for female-specific SNPs in non-coding regions of the *bmpr1ba* gene for indications of potential sex-specific gene transcriptional differences. We identified 12 SNPs in the 5 kb region immediately upstream of the *bmpr1ba* gene that overlap with 57 putative TF binding sites (Table S5). Despite the limited information on the mechanisms involved in the transcriptional regulation of the *bmpr1ba* gene, the presence in the promoter region of SNPs in binding

sites for TFs that interact with BMP signalling, such as MEIS1 (Erickson et al., 2010), GSX1 (Ma et al., 2020), Nkx2-5 (Smith et al., 2000; Theodosiou & Tabin, 2005) and RUNX-2 (Ito & Miyazono, 2003), is suggestive of possible sex-specific differences in the transcription of the *bmpr1ba* gene during the sex determination stage. As with functional alterations of *Amhr2* that promote a female phenotype in medaka (Morinaga et al., 2007) and the tiger pufferfish (*Takifugu rubripes*) (Kamiya et al., 2012), given the required association between *Bmpr1b* and *Amhr2* for *Amh* signalling (Gouédard et al., 2000), it is tempting



**FIGURE 6** Syntenic relationship of Pacific halibut (*H. stenolepis*) chromosome 9 (Chr9) and orthologues in Atlantic halibut (*H. hippoglossus*) (a) and Greenland halibut (*R. hippoglossoides*) (b). The sex associated region of Chr9 in Pacific halibut is shaded in light blue, the genomic location of the *bmpr1ba* gene is indicated in red



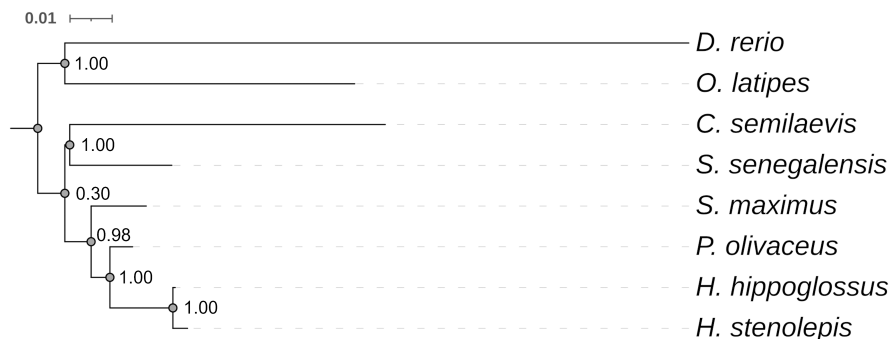
**FIGURE 7** Syntenic comparison of the genes located around the *bmpr1ba* gene in Pacific halibut (*H. stenolepis*), Atlantic halibut (*H. hippoglossus*), Japanese flounder (*P. olivaceus*), turbot (*S. maximus*), Senegalese sole (*S. senegalensis*) and half-smooth tongue sole (*C. semilaevis*). Orthologous genes are identified by colour, a blue line between genes represents a gap in the order of genes (i.e., nonorthologous genes are present in this region of the genome)

to speculate that changes in the expression levels of *bmpr1ba* could potentially result in alterations of Amh signalling in Pacific halibut. Clearly, further studies are needed to provide evidence for a possible role of *bmpr1ba* in sex determination in Pacific halibut.

Related to other Pleuronectiform species, the *bmpr1ba* gene is similarly located in Chr9 in both Atlantic halibut and turbot (Table S6). Interestingly, the *bmpr1ba* gene is located in sex

chromosome Z in half-smooth tongue sole, whereas in Senegalese sole is located in LG05 and in Japanese flounder in an unplaced scaffold (NW\_017859680.1). The syntenic arrangement of the *bmpr1ba* gene is well conserved among Pacific halibut, Atlantic halibut, Japanese flounder and turbot and partially conserved when compared to Senegalese sole and half-smooth tongue sole (Figure 7). The conserved synteny of the *bmpr1ba* gene is coupled with high





**FIGURE 8** Maximum likelihood phylogenetic reconstruction of flatfish *Bmpr1ba* protein sequences, including Pacific halibut (*H. stenolepis*), Atlantic halibut (*H. hippoglossus*), Japanese flounder (*P. olivaceus*), turbot (*S. maximus*), Senegalese sole (*S. senegalensis*) and half-smooth tongue sole (*C. semilaevis*), compared to zebrafish (*D. rerio*) and medaka (*O. latipes*). The values indicate branch support (estimated by aLRT)

homology among the flatfish *Bmpr1ba* protein sequences, particularly between the Atlantic and Pacific halibut sequences that differ only in two amino-acid residues (Figure S6), as supported by their close phylogenetic relationship (Figure 8). Unfortunately, these observations do not provide indications on how the *bmpr1ba* gene may have emerged as a potential candidate MSD gene in Pacific halibut but we attribute this to the relatively recent evolutionary history of species within the Pleuronectidae family.

In summary, we have generated the first chromosome-level assembly of the Pacific halibut genome and used this resource to identify the sex-determining genomic region, leading to the identification of a putative candidate MSD gene for Pacific halibut. The generated genomic resource will be instrumental for future research on population genomics informing management and conservation strategies for Pacific halibut.

#### AUTHOR CONTRIBUTIONS

Yann Guiguen and Josep V. Planas designed the study. Anna Simeon, Céline Roques, Carole lampietro, Jérôme Lluch, Cécile Donnadiou and Hugues Parrinello performed research. Andrew J. Jasonowicz, Margot Zahm, Cédric Cabau, Christophe Klopp, Hugues Parrinello, Daniel P. Drinan, Lorenz Hauser, Yann Guigen analysed the data. Andrew J. Jasonowicz and Josep V. Planas wrote the manuscript and Yann Guigen, Daniel P. Drinan and Lorenz Hauser edited the manuscript. All the authors read and approved the final manuscript.

#### ACKNOWLEDGEMENTS

This work was conducted with funding from the International Pacific Halibut Commission and the North Pacific Research Board (Project no. 2110). MGX and GetPlage acknowledge financial support from France Génomique National infrastructure, funded as part of “Investissement d’Avenir” program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). The authors would like to thank Anita Kroska and Dr Nathan Wolf from Alaska Pacific University for their help with providing the blood donor female Pacific halibut and to the F/V Kema Sue, skipper Lando Echevario and IPHC field samplers Levy Boitor, Kaitlin Johnson and Bruce Biffard for their help collecting the fin clips used for Pool-seq.

#### CONFLICT OF INTERESTS

The authors have no conflict of interest to declare.

#### DATA AVAILABILITY AND BENEFIT-SHARING STATEMENT

Data availability statement:

Genetic data:

The Pacific halibut whole genome sequencing data that support the findings of this study are openly available in NCBI at <https://www.ncbi.nlm.nih.gov/bioproject/622249>, under BioProject PRJNA622249. The Pacific halibut genome assembly is openly available in NCBI at [https://www.ncbi.nlm.nih.gov/assembly/GCA\\_022539355.2/](https://www.ncbi.nlm.nih.gov/assembly/GCA_022539355.2/) with GenBank assembly accession number GCA\_022539355.2. The RNA-seq data that support the findings of this study are openly available in NCBI at [https://www.ncbi.nlm.nih.gov/bioproject?LinkName=biosample\\_bioproject&from\\_uid=14989926](https://www.ncbi.nlm.nih.gov/bioproject?LinkName=biosample_bioproject&from_uid=14989926), under BioProject PRJNA634339.

Sample metadata:

The master record for the whole genome shotgun sequencing project has been deposited at DDBJ/ENA/GenBank under the accession JAKRZP000000000. The version described in this paper is version JAKRZP010000000 and is openly available in NCBI at <https://www.ncbi.nlm.nih.gov/nuccore/JAKRZP000000000>. Sample metadata is openly available in NCBI at [https://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject\\_biosample&LinkReadableName=BioSample&ordinalpos=1&IdsFromResult=622249](https://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject_biosample&LinkReadableName=BioSample&ordinalpos=1&IdsFromResult=622249), under BioSamples SAMN14503176, SAMN25516224, SAMN25600010 and SAMN25600011.

Benefit-sharing statement:

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

#### ORCID

Margot Zahm  <https://orcid.org/0000-0001-8632-0109>

Cédric Cabau  <https://orcid.org/0000-0001-8560-4003>

Christophe Klopp  <https://orcid.org/0000-0001-7126-5477>

Carole lampietro  <https://orcid.org/0000-0002-8148-4785>

Cécile Donnadieu  <https://orcid.org/0000-0002-5164-3095>  
 Hugues Parrinello  <https://orcid.org/0000-0002-1151-001X>  
 Yann Guiguen  <https://orcid.org/0000-0001-5464-6219>  
 Josep V. Planas  <https://orcid.org/0000-0002-6525-9617>

## REFERENCES

- Adolfi, M. C., Du, K., Kneitz, S., Cabau, C., Zahm, M., Klopp, C., ... Scharrtl, M. (2021). A duplicated copy of id2b is an unusual sex-determining candidate gene on the Y chromosome of arapaima (*Arapaima gigas*). *Scientific Reports*, 11(1), 21544. <https://doi.org/10.1038/s41598-021-01066-z>
- Alexa, A., & Rahnenfuhrer, J. (2020). topGO: Enrichment analysis for gene ontology. R package.
- Andersen, C. L., Jensen, J. L., & Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, 64, 5245–5250. <https://doi.org/10.1158/0008-5472.CAN-04-0496>
- Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, 55(4), 539–552. <https://doi.org/10.1080/10635150600755453>
- Asahida, T., Kobayashi, T., Saitoh, K., & Nakayama, I. (1996). Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. *Fisheries Science*, 62, 727–730. <https://doi.org/10.2331/fishsci.62.727>
- Atta, C. J., Yuan, H., Li, C., Arcila, D., Betancur-R, R., Hughes, L. C., & Ort. (2022). Exon-capture data and locus screening provide new insights into the phylogeny of flatfishes (Pleuronectoidei). *Molecular Phylogenetics and Evolution*, 166(December 2020), 107315. <https://doi.org/10.1016/j.ympev.2021.107315>
- Aury, J.-M., & Istace, B. (2021). Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads, *NAR Genomics and Bioinformatics*, 3(2), lqab034. [10.1093/nargab/lqab034](https://doi.org/10.1093/nargab/lqab034)
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., ... Vamosi, J. C. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, 12(7), e1001899. <https://doi.org/10.1371/journal.pbio.1001899>
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27(2), 573–580.
- Best, E. A. (1981). Halibut ecology. In W. Hood, Calder, J.A. (Ed.), *The eastern Bering Sea shelf: Oceanography and resources* (Vol. 1, pp. 495–508). U.S. Department of Commerce, NOAA.
- Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., ... Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1), D344–D354. <https://doi.org/10.1093/nar/gkaa977>
- Brionne, A., Juanchich, A., & Hennequet-Antier, C. (2019). ViSEAGO: A Bioconductor package for clustering biological functions using gene ontology and semantic similarity. *BioData Mining*, 12(1), 16. <https://doi.org/10.1186/s13040-019-0204-1>
- Brown, N. P., Bromage, N. R., Penman, D. J., & Shields, R. J. (1997). The karyotype of the Atlantic halibut, *Hippoglossus hippoglossus* (Linnaeus). *Aquaculture Research*, 28(7), 489–491.
- Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6, e4958. <https://doi.org/10.7717/peerj.4958>
- Capel, B. (2017). Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics*, 18(11), 675–689. <https://doi.org/10.1038/nrg.2017.60>
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 1–16. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., ... Wang, J. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*, 46(3), 253–260. <https://doi.org/10.1038/ng.2890>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21.
- Drinan, D. P., Loher, T., & Hauser, L. (2018). Identification of genomic regions associated with sex in Pacific halibut. *Journal of Heredity*, 109(3), 326–332. <https://doi.org/10.1093/jhered/esx102>
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., ... Aiden, A. P. (2017). De novo assembly of the *Aedes aegypti* genome using hi-C yields chromosome-length scaffolds. *Science*, 356(6333), 92–95.
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Systems*, 3(1), 95–98.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Edvardsen, R. B., Wallerman, O., Furmanek, T., Kleppe, L., Jern, P., Wallberg, A., ... Rubin, C. J. (2022). Heterochiasmy facilitated the establishment of *gsdf* as a novel sex determining gene in Atlantic halibut. *PLoS Genetics*, 18(2), e1010011. <https://doi.org/10.1371/journal.pgen.1010011>
- Eilbeck, K., Moore, B., Holt, C., & Yandell, M. (2009). Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics*, 10(1), 67.
- Einfeldt, A. L., Kess, T., Messmer, A., Duffy, S., Wringe, B. F., Fisher, J., ... 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>
- Erickson, T., French, C. R., & Waskiewicz, A. J. (2010). Meis1 specifies positional information in the retina and tectum to organize the zebrafish visual system. *Neural Development*, 5(1), 22. <https://doi.org/10.1186/1749-8104-5-22>
- 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(1), jkab376. <https://doi.org/10.1093/g3journal/jkab376>
- Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J. A., ... Martinez, P. (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Research*, 23(3), 181–192. <https://doi.org/10.1093/dnares/dsw007>
- Fish, T., Wolf, N., Harris, B. P., & Planas, J. V. (2020). A comprehensive description of oocyte developmental stages in Pacific halibut, *Hippoglossus stenolepis*. *Journal of Fish Biology*, 97(6), 1880–1885. <https://doi.org/10.1111/jfb.14551>
- Foissac, S., Djebali, S., Munyard, K., Vialaneix, N., Rau, A., Muret, K., ... Giuffra, E. (2019). Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biology*, 17(1), 108. <https://doi.org/10.1186/s12915-019-0726-5>

- Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranasic, D., Santana-Garcia, W., Tan, G., Cheneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W. W., ... Mathelier, A. (2019). JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 48, D87–D92. <https://doi.org/10.1093/nar/gkz1001>
- Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., ... Mank, J. E. (2020). Sex chromosome evolution: So many exceptions to the rules. *Genome Biology and Evolution*, 12(6), 750–763. <https://doi.org/10.1093/gbe/evaa081>
- Galindo, H. M., Loher, T., & Hauser, L. (2011). Genetic sex identification and the potential evolution of sex determination in Pacific halibut (*Hippoglossus stenolepis*). *Marine Biotechnology*, 13(5), 1027–1037. <https://doi.org/10.1007/s10126-011-9366-7>
- Godwin, J., & Roberts, R. (2018). Environmental and genetic sex determining mechanisms in fishes. In J. Leonard (Eds.), *Transitions between sexual systems* (pp. 311–344). Springer. [https://doi.org/10.1007/978-3-319-94139-4\\_11](https://doi.org/10.1007/978-3-319-94139-4_11)
- Gouédard, L., Chen, Y. G., Thevenet, L., Racine, C., Borie, S., Lamarre, I., Josso, N., Massague, J., & di Clemente, N. (2000). Engagement of bone morphogenetic protein type IB receptor and Smad1 signaling by anti-Müllerian hormone and its type II receptor. *The Journal of Biological Chemistry*, 275(36), 27973–27978. <https://doi.org/10.1074/jbc.M002704200>
- Grant, W. S. (1987). Genetic divergence between congeneric Atlantic and Pacific Ocean fishes. In N. Ryman, F. Utter (Eds.), *Population genetics and fishery management*. University of Washington Press.
- Grant, W. S., Teel, D. J., & Kobayashi, T. (1984). Biochemical population genetics of Pacific halibut (*Hippoglossus stenolepis*) and comparison with Atlantic halibut (*H. hippoglossus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 41, 1083–1088.
- Guerrero-Czar, I., Gomez-Garrido, J., Berbel, C., Martinez-Blanch, J. F., Alioto, T., Claros, M. G., ... Machado, M. (2021). Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish. *Scientific Reports*, 11(1), 13460. <https://doi.org/10.1038/s41598-021-92601-5>
- 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(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Harrington, R. C., Faircloth, B. C., Eytan, R. I., Smith, W. L., Near, T. J., Alfaro, M. E., & Friedman, M. (2016). Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evolutionary Biology*, 16(1), 224. <https://doi.org/10.1186/s12862-016-0786-x>
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32(5), 767–769.
- Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12, 491. <https://doi.org/10.1186/1471-2105-12-491>
- Imarazene, B., Du, K., Beille, S., Jouanno, E., Feron, R., Pan, Q., ... Guiguen, Y. (2021). A supernumerary "B-sex" chromosome drives male sex determination in the Pachon cavefish, *Astyanax mexicanus*. *Current Biology*, 31(21), 4800–4809. e4809. <https://doi.org/10.1016/j.cub.2021.08.030>
- Ito, Y., & Miyazono, K. (2003). RUNX transcription factors as key targets of TGF-beta superfamily signaling. *Current Opinion in Genetics and Development*, 13(1), 43–47. [https://doi.org/10.1016/s0959-437x\(03\)00007-8](https://doi.org/10.1016/s0959-437x(03)00007-8)
- Jasonowicz, A. J., Simchick, C., & Planas, J. V. (2022). In preparation.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- 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(7), e1002798. <https://doi.org/10.1371/journal.pgen.1002798>
- Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution. *Developmental Dynamics*, 242(4), 339–353. <https://doi.org/10.1002/dvdy.23927>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnol*, 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Kottler, V. A., Feron, R., Nanda, I., Klopp, C., Du, K., Kneitz, S., ... Scharlt, M. (2020). Independent origin of XY and ZW sex determination mechanisms in mosquitofish sister species. *Genetics*, 214(1), 193–209. <https://doi.org/10.1534/genetics.119.302698>
- Lefort, V., Longueville, J.-E., & Gascuel, O. (2017). SMS: Smart model selection in PhyML. *Molecular Biology and Evolution*, 34(9), 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100.
- 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 (SAM) format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C[T]) method. *Methods*, 25(4), 402–408. <https://doi.org/10.1006/meth.2001.1262>
- 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(3), 256–263. <https://doi.org/10.1016/j.semcdb.2008.12.002>
- Ma, C. T., Vong, K. I., & Kwan, K. M. (2020). Spatiotemporal decline of BMP signaling activity in neural progenitors mediates fate transition and safeguards neurogenesis. *Cell Reports*, 30, 3616–3624. <https://doi.org/10.1016/j.celrep.2020.02.089>
- Martinez, P., Robledo, D., Taboada, X., Blanco, A., Moser, M., Maroso, F., ... Vias, 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(4), 1705–1718. <https://doi.org/10.1016/j.ygeno.2021.04.007>
- Miyazono, K., Maeda, S., & Imamura, T. (2005). BMP receptor signaling: Transcriptional targets, regulation of signals, and signaling cross-talk. *Cytokine Growth Factor Reviews*, 16(3), 251–263. <https://doi.org/10.1016/j.cytogfr.2005.01.009>
- Morgulis, A., Gertz, E. M., Schaffer, A. A., & Agarwala, R. (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *Journal of Computational Biology*, 13, 1028–1040.
- Morinaga, C., Saito, D., Nakamura, S., Sasaki, T., Asakawa, S., Shimizu, N., Mitani, H., Furutani-Seiki, M., Tanaka, M., & Kondoh, H. (2007). The hotei mutation of medaka in the anti-Müllerian hormone receptor causes the dysregulation of germ cell and sexual development. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23), 9691–9696.
- Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M. E., Oral, M., Taggart, J. B., ... Migaud, H. (2013). Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*)

- using RAD sequencing. *BMC Genomics*, 14(1), 1–12. <https://doi.org/10.1186/1471-2164-14-566>
- Pan, Q., Anderson, J., Bertho, S., Herpin, A., Wilson, C., Postlethwait, J. H., ... Guiguen, Y. (2016). Vertebrate sex-determining genes play musical chairs. *Comptes Rendus Biologies*, 339(7–8), 258–262. <https://doi.org/10.1016/j.crvi.2016.05.010>
- Pan, Q., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., ... Guiguen, Y. (2021). The rise and fall of the ancient northern pike master sex-determining gene. *eLife*, 10, e62858. <https://doi.org/10.7554/eLife.62858>
- Pan, Q., Guiguen, Y., & Herpin, A. (2018). Evolution of sex determining genes in fish. In M. K. Skinner (Eds.), *Encyclopedia of reproduction* (Vol. 6, pp. 168–175). Academic Press, Elsevier.
- Pan, Q., Kay, T., Depince, A., Adolphi, M., Scharfl, M., Guiguen, Y., & Herpin, A. (2021). Evolution of master sex determiners: TGF-beta signalling pathways at regulatory crossroads. *Philosophical Transactions Royal Society London B Biological Science*, 376(1832), 20200091. <https://doi.org/10.1098/rstb.2020.0091>
- Pennell, M. W., Mank, J. E., & Peichel, C. L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Molecular Ecology*, 27(19), 3950–3963. <https://doi.org/10.1111/mec.14540>
- Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, 9, 304. <https://doi.org/10.12688/f1000research.23297.1>
- Portela-Bens, S., Merlo, M. A., Rodriguez, M. E., Cross, I., Machado, M., Kosyakova, N., ... Rebordinos, L. (2017). Integrated gene mapping and synteny studies give insights into the evolution of a sex proto-chromosome in *Solea senegalensis*. *Chromosoma*, 126(2), 261–277. <https://doi.org/10.1007/s00412-016-0589-2>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Statistical Foundation.
- Rafati, N., Chen, J., Herpin, A., Pettersson, M. E., Han, F., Feng, C., Wallerman, O., Rubin, C. J., Péron, S., Cocco, A., Larsson, M., Trötschel, C., Poetsch, A., Korsching, K., Bönigk, W., Körschen, H. G., Berg, F., Folkvord, A., Kaupp, U. B., ... Andersson, L. (2020). Reconstruction of the birth of a male sex chromosome present in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America*, 117(39), 24359–24368. <https://doi.org/10.1073/pnas.2009925117>
- Reid, D. P., Smith, C. A., Rommens, M., Blanchard, B., Martin-Robichaud, D., & Reith, M. (2007). A genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Genetics*, 177(2), 1193–1205. <https://doi.org/10.1534/genetics.107.075374>
- Schlotterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749–763. <https://doi.org/10.1038/nrg3803>
- Shao, C., Bao, B., Xie, Z., Chen, X., Li, B., Jia, X., ... Chen, S. (2017). The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nature Genetics*, 49(1), 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(19), 3210–3212.
- 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>
- Smith, D. M., Nielsen, C., Tabin, C. J., & Roberts, D. J. (2000). Roles of BMP signaling and Nkx2.5 in patterning at the chick midgut-foregut boundary. *Development*, 127(17), 3671–3681. <https://doi.org/10.1242/dev.127.17.3671>
- Stewart, I. J., Hicks, A. C., & Hutniczak, B. (2021). *Evaluation of directed commercial fishery size limits in 2020* (p. 28). Retrieved from International Pacific Halibut Commission. <https://www.iphc.int/uploads/pdf/am/am097/iphc-2021-am097-09.pdf>
- Stewart, I. J., Webster, R. (2021). Overview of data sources for the Pacific halibut stock assessment, harvest policy, and related analyses. International Pacific Halibut Commission. p. 54. Retrieved from <https://www.iphc.int/uploads/pdf/sa/2021/iphc-2021-sa-02.pdf>
- Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2008). Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosome Research*, 16(5), 801–811. <https://doi.org/10.1007/s10577-008-1227-5>
- Tan, G., & Lenhard, B. (2016). TFBSTools: An R/bioconductor package for transcription factor binding site analysis. *Bioinformatics*, 32(10), 1555–1556. <https://doi.org/10.1093/bioinformatics/btw024>
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science*, 320(5875), 486–488. <https://doi.org/10.1126/science.1153917>
- Tarailo-Gravoac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols Bioinformatics*, 25, 4.10.11–14.10.14.
- Theodosiou, N. A., & Tabin, C. J. (2005). Sox9 and Nkx2.5 determine the phyloric sphincter epithelium under the control of BMP signaling. *Developmental Biology*, 279(2), 481–490. <https://doi.org/10.1016/j.ydbio.2004.12.019>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., ... Pachter, L. (2010). Transcript assembly and abundance estimation from RNA-seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology*, 28(5), 511–515.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3, research0034.1. <https://doi.org/10.1186/gb-2002-3-7-research0034>
- Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, 27(5), 737–746.
- Wallis, M. C., Waters, P. D., & Graves, J. A. M. (2008). Sex determination in mammals – Before and after the evolution of SRY. *Cellular and Molecular Life Sciences*, 65(20), 3182–3195. <https://doi.org/10.1007/s00018-008-8109-z>
- Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E. D., ... Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, 346(6215), 1246338. <https://doi.org/10.1126/science.1246338>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Jasonowicz, A. J., Simeon, A., Zahm, M., Cabau, C., Klopp, C., Roques, C., Iampietro, C., Lluç, J., Donnadieu, C., Parrinello, H., Drinan, D. P., 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 genomic region. *Molecular Ecology Resources*, 22, 2685–2700. <https://doi.org/10.1111/1755-0998.13641>