A compendium of novel genomics technologies provides a chromosome-scale assembly and insights into the sex determining system of the Greenland Halibut

Ferchaud A-L^{*}, Mérot C^{*}, Normandeau E^{*}, Ragoussis I[†], Babin C^{*}, Djambazian H[†], Bérubé P[†], Audet C[‡], Treble M[§], Walkusz W[§], & Bernatchez L^{*}.

^{*} Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, G1V 0A6 Canada

[†] McGill Genome Centre and Department for Human Genetics, McGill University, Montreal, Quebec Canada

[‡] Institut des sciences de la mer de Rimouski, Université du Québec à Rimouski, 310 allée des Ursulines, Rimouski, QC G5L 3A1, Canada

[§] Fisheries and Oceans Canada, Winnipeg Department, Arctic Aquatic Research Division

14 **RUNNING TITLE:**

15 Genome assembly and sex in Greenland Halibut

16 **KEY WORDS:**

17 Long-reads PacBio, Optical mapping, Low-coverage whole genome sequencing, sex-determination, Y-autosomal
18 fusion, flatfish

19

20 **CORRESPONDING AUTHOR:**

21 Anne-Laure Ferchaud,

22 Département de Biologie
23 Laboratoire du Dr Louis Bernatchez
24 Institut de Biologie Intégrative et des Systèmes (IBIS)
25 Pavillon Charles-Eugène Marchand 1030 rue de la Médecine
26 Université Laval
27 Québec (Québec)
28 G1V 0A6
29 Canada

30
31 Phone : +1 418 656 2131 (#408455)

32 annelaureferchaud@gmail.com

33

34

ABSTRACT

Despite the commercial importance of Greenland Halibut (*Reinhardtius hippoglossoides*), important gaps still persist in our knowledge of this species, including its reproductive biology and sex determination mechanism. In this study, we combined single molecule sequencing of long reads (Pacific Sciences) with Chromatin Conformation Capture sequencing (Hi-C) data to provide the first chromosome-level genome reference for this species. The high-quality assembly encompassed more than 598 Megabases (Mb) assigned to 1 594 scaffolds (scaffold N50 = 25 Mb) with 96 % of its total length distributed among 24 chromosomes. The investigation of its syntenic relationships with other economically important flatfish species revealed a high conservation of synteny blocks among members of this phylogenetic clade. Sex determination analysis revealed that flatfishes do not escape the rule applied to other teleost fish and exhibit a high level of plasticity and turnover in sex-determination mechanisms. A whole-genome sequence analysis of 198 individuals allowed us to draw a full picture of the molecular sex determination (SD) system for Greenland Halibut, revealing that this species possesses a very nascent male heterogametic XY system, with a putative major effect of the *sox2* gene, also described as the main SD driver in two other flatfishes. Interestingly, our study also suggested for the first time in flatfishes that a putative Y-autosomal fusion could be associated with a reduction of recombination typical of early steps of sex chromosome evolution.

INTRODUCTION

The comprehension of sex-determining (SD) mechanisms is of high importance in biology, as they affect developmental processes, the evolution of genomes, and are responsible for determining sex ratio in a population, a key demographic parameter for its viability and stability (Bull 1983, Uller et al 2007). In recent years, genetic analyses in non-model species have revealed a greater degree of sex chromosome diversity and turnover than previously appreciated (Abbott et al 2017). In fact, some lineages, including lizards, fish, amphibians, insects, and plants show frequent changes in the location of SD genes and high rates of turnover of sex chromosomes. This led Furman et al (2020) to propose a new conceptual framework, in which sex chromosome evolution is more cyclical than linear and in which exceptions are the rules in sex determination (McGrath 2020). In this cycle, a new master sex-determining locus arises on an autosome, leading to sex chromosome formation generally characterized by recombination suppression between the Y and X, or the Z and W. As under a classic sex chromosome evolution model, this may progress towards increased divergence between male and female chromosomes, up to chromosome heteromorphism characterized by genomic degradation including rearrangements and loss of genetic material (Charlesworth et al 2005; Bachtrog 2013). However, sex chromosomes and sex-determining regions can also undergo turnover with either the evolution of a new sex-determining gene or the transposition of the sex-determining gene to a new location in the genome, leading to another cycle (Bachtrog 2006).

The increasing availability of high-quality genome assemblies in non-model taxonomic groups is now providing exciting opportunities to revisit the evolution of genetic sex-determination.

Fish form a particularly relevant group to investigate the evolutionary dynamics of sex chromosomes, as they exhibit a remarkable gonad development plasticity and a high evolutionary SD turnover (Kitano and Peichel 2012). Fish display highly diverse chromosome SD systems, including models analogous to the XX/XY (male heterogametic) and ZZ/ZW (female heterogametic) systems of mammals and birds respectively, but also others involving multiple sex chromosomes (Capel 2017). In accordance with this rapid turnover, chromosome heteromorphisms are rare in fish, which generally show limited sex chromosome differentiation (Chen et al 2014), characteristic of early sex-determination mechanisms. Until recently, the DM-domain gene on the Y chromosome (*dmy*) gene was the only known gene coding for a functional protein in the Y-specific chromosome region (Nanda et al 2002, Kondo et al 2006) and was described as the master SD gene in a wide range of invertebrates and vertebrates (including fish). However, thanks to new genomics tools, the genomic features underlying SD have been more intensively studied and several other strong candidates of master SD genes have been found in fish (*amhy*, *gsdf*, *amhr2*, *sdY*, Guerrero-Estévez and Moreno-Mendoza 2010, Yano et al 2013). These novel master SD genes highlight the importance of non-transcriptional factors in sex determination and differentiation. In fact, three of these genes (*amhy*, *amhr2*, and *gsdf*) are involved in the transforming growth factor beta (TGF- β) signaling pathway, while the *sdY*, *dmrt1*, and *dmy* genes code for transcription factors (Rastetter et al 2015). Insights into SD genomics goes beyond master genes in fish, as recent studies have documented mixed determination

involving both genetic and environmental factors, polygenic SD systems, extensive within-species variation in SD system, and complex rearrangements of sex chromosomes favouring species diversification (Capel 2017, Martinez et al 2014). Yet, with such dynamism and variability in fish sex chromosome evolution, we need improved genomic resources in non-model species, including chromosome-scale reference genomes and whole-genome sequences from both sexes.

In exploited species, improved genomic resources and the understanding of the genetic basis for SD is also relevant to manage fisheries in order to meet the need for seafood around the world, and to secure the economies and livelihoods that depend on them (Bernatchez et al 2017). Genomics approaches allow for the investigation of genome wide diversity and to accurately identify fish stock within commercial species (Ovenden et al 2013). Moreover, accounting for sex-differentiation and sex-linked loci is necessary to avoid unexpected biases due to sex-associated markers (Benestan et al 2017) as well as to improve the management of sex-ratio. This is particularly true for species harbouring a size-related sexual dimorphism like flatfish species, generally favouring females in their size dimorphism (e.g. European Plaice (*Pleuronectes platessa*; Rijnsdorp and Ibelings 1989), Turbot (*Scotphthalmus maximus*; Imsland et al 1997), Atlantic Halibut (*Hippoglossus hippoglossus*; Hagen et al 2006). In such species, sex control and/or identification is of high importance either for flatfish production in aquaculture or sustainable fisheries. The selection of large individuals in fisheries is a well-established concept in fisheries management that increases yield per recruit. However, in the case of sex-related sexual dimorphism, fisheries targeting large individuals may reduce the stock's reproductive potential, produce bias in the assessment of the number of reproductive females, and feed erroneous data to fisheries management decision makers (Keyl et al 2015). Overall, applied and fundamental

evolutionary biology thus benefits from moving beyond genetic markers towards contiguous reference genomes annotated for functional elements and documented for sex-determination.

Greenland Halibut (*Reinhardtius hippoglossoides*) is one of the main demersal fish exploited by fisheries in Eastern Canada (Gulf of St. Lawrence, Newfoundland and the Arctic), representing around 4 000 tones of landings per year across the two last decades (DFO, 2018). However, it suffers from a lack of genomic resources, since no karyotype, linkage map nor genome assembly has been developed so far, thus hampering the use of genomics for improved management. Since juvenile growth potential and size are sexually dimorphic in this species (Ghinter et al 2019), it is valuable to better understand the sex-determining mechanisms to control sex imbalance in catches. Moreover, the recent genome assemblies of flatfish species (Figure 1, Table1, Chen et al 2014, Figueras et al 2016, Shao et al 2017, Einfeldt et al 2021) together with studies on sex-determining loci in this group (Chen et al 2014, Liang et al 2018, Drinan et al 2018, Einfeldt et al 2021) provides an opportunity to investigate the evolutionary turnover of sex chromosomes and sex-determination systems. Flatfishes studied so far are characterized by a variable heterogametic system and different sex-determining genes. For instance, a similar female heterogametic system (ZZ/ZW) is used both by the Tongue Sole (*Cynoglossus semilaevis*) and the Turbot, but the candidate SD drivers differ, being *sox2* in the latter (Martinez et al 2021, Table1), and *dmrt1* in the former species (Chen et al 2014). Japanese Flounder, *Paralichthys olivaceus* (also called the Olive flounder,) has an XX/XY system (Liang et al 2018) but no sex determination gene has been found (Shao et al 2017) and environment is known to alter the gender (Ospina-Alvarez and Piferrer 2008). Within the righteye flounder family (*Pleuronectidae*), the *Hippoglossinae* subfamily harbors variable genetic SD systems. The

Pacific Halibut (*Hippoglossus stenolepis*) has a ZZ/ZW system and several sex-linked loci aligned in three different linkage groups (Drinan, Loher, & Hauser 2018) while its sister-species, the Atlantic Halibut (*H. hippoglossus*), has a XX/XY system and a putative SD factor called gsdf on chromosome 12 (Palaikostas et al. 2013; Einfeldt et al 2021).

In this context, the first goal of this study was to provide high-quality genomic resources and to investigate the genetic basis of sex determination in Greenland Halibut. We produced the first chromosome-scale scaffolded genome assembly for this species, combining single molecule sequencing of long reads (Pacific Sciences) with Chromatin Conformation Capture sequencing (Hi-C) data. Furthermore, we also identified and characterized the molecular SD system by mapping whole genome sequencing data of hundreds of individuals onto the assembly to screen for divergent genetic markers between sexes. Finally, we analyzed our results in a comparative genomic framework by assessing synteny (i.e. block of genes in the same relative positions in the genome) with related species of flatfish and targeting candidate sex-determining genes in this group.

METHODS

Reference genome assembly

We built a reference genome assembly for the Greenland Halibut through two steps. First, we created a draft assembly using PacBio sequencing approach (Pacific Biosciences, Menlo Park, California) and then we used DovetailTM Hi-C (Lieberman-Aiden et al 2009) – from Dovetail Genomics to increase contiguity, break up mis-joins, and orient and join scaffolds into chromosomes.

- *Sampling*

The individual used for the reference was a juvenile female caught in the St. Lawrence River Estuary (48°39'11"N, 68°28'37"W) with a Comando-type trawl aboard the CCGS Leim (Fisheries and Oceans Canada [DFO] survey) at the end of May 2016 and kept alive in common garden environment at the Maurice Lamontagne Institute (Ghinter et al 2019). The female was anaesthetized for 5 min in a solution of MS 222 (tricaine methane sulfonate 0.18 g L⁻¹; Sigma - Aldrich Co, Missouri, USA). Blood was then sampled from the caudal artery using a 23-gauge needle in a 1ml TB syringe (Becton, Dickinson & co., New Jersey, USA), both of which were previously heparinized in a 100 U ml⁻¹ heparin solution. Blood was immediately flash frozen in liquid nitrogen and then stored at -80°C until further manipulation, as well as multiple tissues including liver, muscle, heart, brain, eye, gonads, gall bladder and kidney that were excised from the fish.

- *DNA extraction, sequencing and assembly*

High Molecular Weight DNA was extracted using Qiagen MagAttract kit, (Valencia, CA). Then a library was created using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, California), which includes a DNA Damage Repair step after size selection. Size selection was made for >10 kb using a Blue Pippin instrument (Sage Science, Beverly, Massachusetts) according to the manufacturer recommended protocol for 20 kb template preparation. Five µg of concentrated DNA was then used as input for the library preparation reaction. Library quality and quantity were assessed using the Pippin Pulse field inversion gel electrophoresis system (Sage Science), as well as with the dsDNA Broad Range Assay kit and Qubit Fluorometer (Thermo Fisher). Single-Molecule Real Time sequencing was performed on the Pacific Biosciences RS II instrument at McGill University using an on-plate concentration of 125 pM, P6-C4 sequencing chemistry, with magnetic bead loading, and 360-minute movies. A total of 14 SMRT Cells were run, generating a mean sequence coverage of ~70-fold, with half of the 522 Megabases (Mb) of data contained in reads longer than 10.6 kb. An initial assembly was carried out on the PacBio long reads using Wtdbg v2.5 (-x sq gg 0.7g; Ruan and Li (2020), resulting in 2 037 contigs (N50 = 31.7 Mb), for a total assembly size of 598 MB. One Hi-C proximity ligation library was generated from a flash frozen piece of muscle of the same individual using the Dovetail's TM Hi-C Kit v.1.0 to aid with assembly scaffolding. This kit employs the restriction enzyme *DpnII* which recognizes the sequence 5' /GATC 3'. The long-range library was then sent to Dovetail Genomics to be sequenced and run with the initial PacBio assembly through Dovetail's HiRiseTM scaffolding pipeline. The completeness of the final assembly was assessed with BUSCO v 4.0.5 (Simão et al 2015), the eukaryota_odb10 database (created in October 2020) and the -m genome option.

Occurrence of repeat elements in our assembly as well as in Atlantic and Pacific halibut was assessed using RepeatMasker v.4.0.6 (Smit, Hubley and Green, 2013).

- *Synteny with other flatfish species*

Synteny blocks between the genomes of Greenland Halibut and other flatfish species for which a reference genome is available (Tongue Sole, Turbot and Japanese Flounder as well as Atlantic and Pacific Halibut) were computed using SyMAP v4.2 (Soderlund, Bomhoff and Nelson 2011). Genomic sequences were first aligned using promer/MUMmer (Kurtz et al 2004). Raw anchors resulting from MUMmer were clustered into (putative) gene anchors, filtered using a reciprocal top-2 filter and used as input to the synteny algorithm (Soderlund et al 2006). The algorithm constructs maximal-scoring anchor chains based on a given gap penalty. It also searches a range of gap penalties to generate the longest chains subject to several quality criteria, which are based on the Pearson correlation coefficient applied to the anchors in the chain as well as the anchors in its bounding box. The chains are not required to be entirely colinear and may incorporate local inversions relative to the overall chain orientation. Synteny analysis results were visualized using the Chromosome Explorer and the Circle view options in SyMAP v4.2. Finally, we named scaffolds by mapping our assembly to the Pacific Halibut assembly (IPHC_HiSten_1.0, Genbank accession: GCA_013339905.1).

- *Transcriptome assembly and annotation of the assembly*

In order to annotate the reference assembly, total RNA from 8 different tissues (liver, muscle, heart, brain, eye, gonads, gall bladder and kidney) from the same individual was extracted using the RNeasy Mini kit from Quiagen then treated using DNase I, Amplification Grade (1unit per μ l; Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. RNA concentration of each tissue was quantified with QuantiT Picogreen dsDNA Assay Kit 669 (Invitrogen) and fragment size distribution was estimated with an Agilent BioAnalyzer before being shipped to the Genome Québec Centre d'Expertise et de Services (Montréal, Canada) to prepare libraries and perform sequencing. One RNA-seq library was constructed for each tissue using the NEBNext[®] Ultra[™] II Directional RNA Library Prep Ki (Illumina, San Diego, Calif., USA) and sequenced with a 2×100bp (paired-end) read module using the NovaSeq 6000, generating 50M reads.

For each of the 8 paired-end tissue-specific RNAseq datasets, raw reads were trimmed with trimmomatic (v0.36, ILLUMINACLIP:univec.fasta:2:20:7, LEADING:20, TRAILING:20, SLIDINGWINDOW:30:30, MINLEN:80). The tissue-specific paired-end datasets were then each assembled separately using Trinity (v2.1.1, --seqType fq, --min_contig_length 200) to produce 8 assembled transcriptomes that contained between 68 481 and 276 168, with a mean of 131 038, assembled expressed sequences. The 1 048 307 assembled expressed sequences were used to annotate the assembled genome using the GAWN pipeline v0.3.4 (<https://github.com/enormandeu/gawn>). The annotations were then simplified by using the simplify_genome_annotation_table.py script, which removes consecutive almost-identical annotations from the genome annotation table.

251

252 *Genomic sex determination*

253

- 254 • *Sampling of individuals, DNA extraction, libraries and sequencing*

255

256 A total of 432 Greenland Halibut were collected aboard the R/V Paamiut, from October to
 257 November 2016, as part of the annual multispecies survey conducted by the Department of
 258 Fisheries and Oceans in Northwest Atlantic Fisheries Organization Subarea 0 in Davis Strait,
 259 Canada. Each fish collected has been measured and weighed as well as phenotypically sex
 260 identified. From those individuals, 99 females and 99 males have been selected for sex
 261 determination analysis (see sample selection details in Supplementary Material).

262 Genomic DNA was extracted using a salt-extraction protocol (Aljanabi and Martinez 1997)
 263 with a RNase A treatment 645 (Quiagen). DNA quality of each extract was evaluated with
 264 nanodrop and on a 1% agarose gel electrophoresis. Following Therkildsen and Palumbi 2017, we
 265 removed DNA fragments shorter than 1kb by treating each extract with Axygen magnetic beads
 266 in a 0.4:1 ratio, and eluted the DNA in 10mM Tris-Cl, pH 8.5. We measured DNA concentrations
 267 with Accuclear ultra high sensitivity dsDNA quantification kit (Biotium) and normalized all
 268 samples at a concentration of 5ng/ μ L. Then, sample DNA extracts were randomized, distributed
 269 in plates (96 -well see details about randomization and number of plates used in Supplementary
 270 Material) and re-normalized at 2ng/ μ L.

271 Whole-genome high-quality libraries were prepared for each sample according to the
 272 protocol described in (Baym et al. 2015; Therkildsen and Palumbi 2016). Briefly, a tagmentation

reaction using enzyme from the Nextera kit, which simultaneously fragments the DNA and incorporates partial adapters, was carried out in a 2.5 µl volume with approximately 2 ng of input DNA. Then, we used a two-step PCR procedure with a total of 12 cycles (8+4) to add the remaining Illumina adapter sequence with dual index barcodes and to amplify the libraries. The PCR was conducted with the KAPA Library Amplification Kit and custom primers derived from Nextera XT set of barcodes A, B, C and D (total 384 combinations). Amplification products were purified from primers and size-selected with a two-steps Axygen magnetic beads cleaning protocol, first with a ratio 0.5:1, keeping the supernatant (medium and short DNA fragments), second with a ratio 0.75:1, keeping the beads (medium fragments). Final concentrations of the libraries were quantified with Accuclear ultra high sensitivity dsDNA quantification kit (Biotium) and fragment size distribution was estimated with an Agilent BioAnalyzer for a subset of 10 to 20 samples per plate. Finally, libraries were pooled (see details in Supplementary Material) for sequence lanes of paired-end 150 bp reads on an Illumina HiSeq 4000 at the Norwegian Sequencing Center at the University of Oslo. Given our multiplexing (up to 96 individuals per lane) and genome size (~ 600 Mb), we targeted to a low coverage sequencing (~ 1.26 X).

- *Sequencing filtering and processing*

Raw reads were trimmed, filtered for quality, mapped to the reference genome (obtained above), cleaned for mapping quality and duplicates reads and then re-aligned using the pipeline available at https://github.com/enormandeau/wgs_sample_preparation, inspired by (Therkildsen and Palumbi 2016) and fully described in Mérot et al (2021). Given the high

percentage of the assembly comprised into the anchored 24 chromosomes (> 96 %, see results section), raw reads were mapped against a reduced version of the genome including only the 24 chromosomes and not the unplaced scaffolds.

For low coverage whole genome sequencing (lcWGS) data, the recommended practice is to avoid basing downstream analysis on raw counts of sequenced bases or called genotypes (Nielsen et al 2011), but instead use a probabilistic approach based on the use of Genotype Likelihoods. Several models for computing genotype-likelihood-based on read data have been implemented in the program ANGSD (Korneliussen, Albrechtsen and Nielsen et al 2014), which is currently the most widely used and versatile software package for the analysis of lcWGS (Lou et al 2021). Therefore, the program ANGSD v0.931 was used for most of our subsequent analyses, according to the pipeline documentation available at https://github.com/clairemerot/angsd_pipeline. For all analyses, input reads were filtered to remove reads with a samtools flag above 255 (not primary, failure and duplicate reads, tag - remove_bads = 1), with mapping quality below 30 (-minMapQ 30) and to remove bases with quality below 20 (-minQ 20). We also filtered in order to keep only SNPs covered by at least one read in at least 30 % of individuals (-minInd 60) and remove SNPs in putative repeated regions allowing a maximum depth of 3 times the number of individuals (-setMaxDepth 594). Finally, for most of the subsequent analysis (unless mentioned otherwise) we kept SNPs with minor allele frequency above 5%.

First, we ran ANGSD to estimate genotype likelihoods (GL) with the GATK model (-doGlf 2 -GL 2 -doCounts 1), the spectrum of allele frequency (-doSaf 1) and minor allele frequency (-doMaf

1). The major allele was based on the genotype likelihood and was the most frequent allele (-doMajorMinor 1). From this first analysis, we generated (i) a beagle file with GL estimates and (ii) a list of variants passing those filters and their respective major and minor alleles that were used for most subsequent analysis. The R program (R Core Team 2020) was employed for graphic output in subsequent analyses, *via* the package ggplot (Wickham 2016).

- Genome-wide coverage difference between sexes

Total depth along the entire genome was calculated using the function -doDepth 1 in ANGSD, excluding high coverage (-setMaxDepth 594 and -maxDepth 1000) and low coverage regions (-minInd 60), along with the following arguments for depth: -dumpCounts 2 -doCounts 1 and then averaged across sliding-windows of 25Kb with a step of 5 Kb using the winScan() function of the R package windowscanr (Tavares 2021). The sliding approach allowed us to visualize individual coverage along the genome, while a ratio of total depth of females above total depth of males was also plotted using ggplot(). This ratio is expected to deviate from 1 in genomic regions in which a sex-bias in coverage exists.

- *Genome wide variation between sexes*

Genome-wide variation across samples was explored using PCAngsd (Meisner and Albrechtsen 2018) on the genotype likelihoods. This program extracts a covariance matrix that is then decomposed into principal component analysis (PCA) with R using a scaling 2

transformation adding an eigenvalues correction, to obtain the individuals PC scores (Legendre and Legendre 1998).

- *Identify genomic regions implied in sex determination*

In order to identify the genomic regions involved in the sex differentiation, two analyses were performed. First, we ran PCAngsd on genotype likelihoods in non-overlapping windows of 1 000 SNPs in each chromosome to extract local covariance matrices and obtained local PCAs of genomic variation (as detailed above). For each local PCA, we analyzed the correlation between individuals PC1 scores and PC scores from the global PCA performed on the entire genome. The resulting correlation indices were then plotted across their relative position in each chromosome. Second, genome-wide F_{st} was estimated between males and females. To do this, allele frequency spectrum (-doSaf 1) and minor allele frequencies were calculated for each sex with the previous list of variant positions (-sites) and their polarisation as major or minor alleles (-doMajorMinor 3). Genome-wide F_{st} was estimated using the realSFS function in ANGSD between males and females and then summarized across sliding-windows of 25 Kb with a step of 5 Kb.

Observed heterozygosity was computed for each sex and each SNP using the function -doHWE in ANGSD and then averaged across sliding-windows of 25Kb with a step of 5 Kb using the winScan() function of the R package windowscanr.

- *LD estimation*

For the two chromosomes associated with sex determination (see results below), intra-chromosomal as well as inter-chromosomal linkage disequilibrium was calculated within each sex for a reduced number of SNPs, filtered with more stringent criteria (MAF > 10%, at least one read in 75 % of the samples). Pairwise R^2 were calculated using ngsLD (Fox et al 2019) based on genotype likelihood reported in the beagle file obtained previously. The 10th percentile of pairwise R^2 values was averaged across windows of 250 Kb and plotted for each sex using custom scripts available at https://github.com/claimeerot/angsd_pipeline. Then changes in LD along the sex identified chromosomes were analyzed by calculating mean R^2 for males and females separately.

- *Mapping known sex locus*

To localize homolog genes of known sex-associated genes in our assembly, we aligned all sex-associated markers from Drinan et al (2018) as well as others previously described to have a sex-determination effect in flatfish or teleost species, to our reference assembly using LAST v.1179 (Kielbasa et al 2011; see Table S1 for details about those genes and their accession number).

DATA AVAILABILITY:

Genome assembly is available upon the GCA_006182925.2 accession. Raw data for the 8 RNA-seq lanes are available on NCBI upon following accession numbers: SAMN18221283 to

SMAN18221290 and 198 individual whole genome sequences are submitted under the BioProject ID PRJNA737354 on NCBI SRA and will be available upon publication.

RESULTS AND DISCUSSION

- A high-quality reference genome serving flatfish management and research

We generated a final genome assembly of 598.5 Mb, with overall high contiguity (N50=25MB, N90=20MB), of which 96.35% was comprised into 24 chromosome-length scaffolds (GenBank accession: GCA_006182925.3). These 24 scaffolds are consistent with the number of diploid chromosomes (2n=48) for Atlantic Halibut (Einfeldt et al 2021) and Pacific Halibut (GCA_013339905.1). The Hi-C data supported a high degree of accuracy in the overall assembly into these 24 scaffolds, as indicated by the strong concentration of data points along the diagonal rather than elsewhere in the contact maps (Figure S1). Altogether, this suggests that these 24 scaffolds can be considered as chromosomes. The completeness of the assembly based on single-copy orthologs across eukaryotes was 98.04%, with 96.47% complete and single copy orthologs and 1.57% duplication.

There were 60 313 transcripts detected from multi-tissue transcriptomics data, with 97.7% of them (58 908 transcripts) within the 24 anchored chromosomes. A total of 19 252 genes and pseudogenes were annotated, of which 19 053 (99%) were located within the 24 chromosomes. Repeated elements accounted for 8.20% of chromosome length (49.1 Mb). These measures of completeness and assessments of repeat elements in Greenland Halibut were highly concordant

with the statistics obtained from Atlantic and Pacific Halibut as well as the three other flatfish species with reference genome and important for fisheries and/or aquaculture production (Table 1). SyMap revealed shared syntenic blocks with other flatfish species, allowing the identification of homologous chromosomes (Table 2 and Figure 1). Most of the chromosomes harbor a single syntenic block compared with the most closely-related *Hippoglossinae*, the Pacific Halibut (16 out of 24 chromosomes) and the Atlantic Halibut (15/24). In contrast, mapping with the three other flatfish species displayed several syntenic blocks for all chromosomes (Table 2). As a result, syntenic blocks are larger with Pacific and Atlantic Halibut than with the three other flatfishes (Table 2, and Figure 1). Figure 1 displays this negative relationship between the size of syntenic blocks and phylogenetic distance.

By generating this chromosome-level genome assembly for Greenland Halibut, among the most contiguous of fish assemblies to date, we provide a genomic resource for this unique species of the genus *Reinhardtius*. This reference genome will now become a fundamental tool to understand the genomic variability of Greenland Halibut in the context of comparative genomics and fisheries applications.

- *Nascent XY sex determination system*

After mapping the reads of 99 male and 99 female Greenland Halibut samples to the reference genome, cleaning for quality and processing for SNP identification using GL approach,

we identified 7 301 470 Single Nucleotide Polymorphisms (SNPs) with a minor allelic frequency (MAF) above 5%. These SNPs were then used for further differentiation analyses.

Genome-wide variation analyzed by a global PCA displayed two clusters corresponding to males and females on PC1, which explained 5.7 % of the genetic variance (Figure S2). Local PCAs on windows of 1 000 SNPs along the genome revealed that this sex differentiation was mainly explained by genetic variation on two chromosomes (Chr10 and Chr21 Figure 2A). The correlation between the 1st PC of the global PCA and each local PCA was very low at the very beginning of each of those two chromosomes but it increased rapidly along both chromosomes. A large portion of each of these two chromosomes explains the total genetic variation observed between the two sexes in PC1. Particularly, the last 66.5 % of Chr10 (16.5 Mb out of 24.8 Mb) and 52.2 % of Chr21 (10.5 Mb out of 20.1 Mb) explained more than 90 % the sex differentiation observed on global PC1 (Figure 2B). Genome-wide *F_{st}* estimations between males and females corroborate this result revealing high differentiation on Chr10 and Chr21 (Figure 2C). While the mean genome-wide *F_{st}* between the sexes is as low as 0.01, Chr10 and Chr21 harbored a large region with higher *F_{st}* values with a mean *F_{st}* of 0.21 and 0.15, respectively (Figure 2D). Few localized peaks of sex differentiation were also observed on Chr9, 11, 12, 13 and 15 (Figure 2B).

The observed proportion of heterozygotes followed the same pattern, with similar values between males and females across most of the genome (Figure 2E), but with a much higher fraction of heterozygotes in males than females in the two major regions previously described. This difference between males and females was more pronounced in Chr10 than in Chr21. Both Chr10 and Chr21 were also characterized by strong linkage disequilibrium, which was markedly

higher in males than in females (Figure 3 and Figure S4), suggesting that recombination events are less frequent in males than females for these genomic regions.

In contrast with differentiation and heterozygosity, coverage did not differ between males and females, neither genome-wide nor in regions of sex differentiation (Figure S3), suggesting that Greenland Halibut is not harboring sex chromosome heteromorphism and that the putative sex chromosomes Chr10 and Chr21 have not undergone degradation, which is typical of early SD systems.

Altogether, these results point towards a major role of Chr10 and Chr21 in genetic sex determination. The higher heterozygosity and LD in males suggests that Greenland Halibut has a XX/XY sex determination mechanism. The putative male Y chromosome exhibits two characteristics of sex chromosome evolution; it is divergent from the X chromosome and exhibits a reduction of recombination. However, it does not appear degraded, suggesting very nascent sex chromosomes at an early evolutionary stage. This is corroborated by the syntenic relationships with closely related flatfish species which indicate that homologues of Chr10 and Chr21 are autosome chromosomes. Chr10 and Chr21 neither correspond to locations where sex-determining regions have been defined in other species (Figure 1, Table1). Overall, the resolved sex-determining system in Greenland Halibut is highly consistent with the rapid turnover of sex chromosomes in fish, particularly in flatfish, given that the Atlantic Halibut has also undergone a recent evolution of an XY sex-determination mechanism (Einfeldt et al 2021).

Observing high sex differentiation on two different chromosomes was nevertheless unexpected and puzzling. The Hi-C contact map very strongly supported that Chr10 and Chr21

are distinct chromosomes in the specimen used for the assembly, a juvenile female (Figure S1). Inter-chromosomal linkage disequilibrium was also low in females, confirming the independent segregation of Chr10 and Chr21. Conversely, LD between Chr10 and Chr21 was very high in males (Figure 3), suggesting that the male haplotypes of those two chromosomes segregate together. One possible mechanism explaining such pattern would be a male-restricted fusion between the Y chromosome and one autosome. Chromosome fission and fusion events are frequent in the evolution of vertebrates, particularly in fish, in which they could explain the enormous species-richness and rapid evolution of karyotypes (Cheng et al 2020; Sutherland et al 2017). Fusions may also be polymorphic in fish and play an adaptive role, as reported in Atlantic salmon (*Salmo salar*), in which fused and unfused karyotypes have been shown to coexist (Lehnert et al 2019; Wellband et al 2019). Moreover, fusions between a Y chromosome and an autosome are known to occur more frequently than autosomal fusions or W-autosome fusions, particularly in fishes Pennell et al (2015). The most probable explanation for the preponderance of Y-autosome fusions involved both selection and sex biases, as the smallest effective size (N_e) of the Y makes it more likely to fix rearrangements including those with partially deleterious effects. This process may be even more prevalent in species with polygynous mating system like fish, which increases the variance of male reproductive success, and in species with skewed sex-ratio, as observed in the halibut. Because of its significant impact on recombination, chromosomal fusion may be one of the key factors underlying the evolution of differentiated sex chromosomes. Fusions are a peculiar class of rearrangements that not only link chromosomes that previously segregated independently but also reduce recombination rates within each arm and prevent recombination between fused and unfused homologous chromosomes in

polymorphic populations (Dumas and Britton-Davidian 2002). This typically leads to gradual differentiation from the fusion point as observed here for Greenland Halibut or the young sex chromosomes found in three-spined stickleback (*Gasterosteus aculeatus*) (Kitano et al 2009). This represents a typical step in the birth and establishment of new sex chromosomes (McAllyster 2003, Kirkpatrick 2010), which is often mediated *via* large chromosomal rearrangements such as inversions or fusions. One of the main determinants of this process is selection against recombination, because linkage between sex-determination loci and alleles beneficial to one sex has the potential to resolve sexual conflict (Charlesworth et al 2005). As discussed below, we identified three sex-associated loci spread on two chromosomes and one may speculate whether epistatic interactions may occur between them. Further work would be needed to better understand gene interactions in this region and more formally confirm the putative male-limited fusion (Guerrero and Kirkpatrick 2014; Wright et al 2016).

- *Sox2 gene as the major driver of Greenland Halibut SD?*

Alignment of DNA or mRNA sequences of previously described sex-associated markers revealed that those genes were widely distributed throughout Greenland Halibut assembly (on seven chromosomes, Table S1). Most of those genes are located in regions of low differentiation between males and females: *amhr2*, *cyp19a1A* on chr05; *cyp11b2* and *rspo1* on chr08; *dmrt1* on chr09, *foxl2* on chr10 and *gsdf* on chr18, making them unlikely to underlie sex differentiation. However, three candidate genes mapped to the putative sex chromosomes: *gdf6* and *sox2* on

Chr10, and *sox9a2* on Chr21 (Figure 2, Table S1), and are thus possible drivers of sex-determination in Greenland Halibut.

Growth differentiation factor 6 (*Gdf6*) is a member of the TGF- β family involved in vertebral segmentation and cell differentiation. While no gonadal function was reported previously, *gdf6Y* emerged recently as a putative SD gene in the Turquoise Killifish (*Nothobranchius furzeri*) (Reichwald et al 2015) because sex-linked allelic variations affect the *Gdf6* protein function. The Y chromosome allele, *gdf6Y*, also exhibits sex-linked expression during testicular differentiation in the Turquoise Killifish, supporting its SD role but functional evidence of its role as a SD gene is yet not available.

Sox2 is a member of the SOX (SRY-related HMG-box) family with a high-mobility-group (HMG) DNA-binding domain. It is an important transcription factor, which is involved in developmental processes of vertebrates, such as embryogenesis, maintenance of stem cells, neurogenesis. In the Rohu Carp (*Labeo rohita*), *sox2* mRNA is expressed in various organs, as well as in the culture proliferating spermatogonial stem cells (Patra et al (2015). Sox2 is also a candidate SD gene in two aquaculture species, the mollusc Zhikong scallop (*Chlamys varia*) (Liang et al 2019) and the crustacean black Tiger Shrimp (*Penaeus monodon*) (Guo et al 2019). In flatfish, *sox2* is a strong candidate for SD, in Japanese Flounder, *sox2* transcripts is expressed in gonadal tissues and the transcript abundance in ovaries is higher than in testis (Gao et al 2014), and, in Turbot, *sox2* is defined as a consistent candidate gene putatively driving SD (Martinez et al (2021).

Sox9 gene has been related to sexual differentiation in several teleost species: Rainbow Trout, (*Oncorhynchus mykiss*, Takamatsu et al. 1997; Vizziano et al. 2007), Tilapia (*Oreochromis niloticus*, Ijiri et al 2008), Zebrafish (*Danio rerio* Chiang et al 2001; von Hofsten and Olsson 2005), medaka, *Oryzias latipes* (Yokoi et al 2002), Guppy (*Poecilia reticul*, Shen et al 2007), the Common Carp (*Cyprinus carpio*, Du et al. 2007) and the Asian Swamp Eel (*Monopterus albus*, Zhou et al 2003). In the Rainbow Trout, two sox9 genes have been identified: sox9a1 and sox9a2. Sox9a1 shows higher expression levels in males than in females before sexual differentiation (Vizziano et al 2007). In Tilapia, sox9 expression is similar in XX and XY gonads, but, after sexual differentiation, it is positively regulated in XY gonads (Ijiri et al 2008). In Medaka, testicular sox9 (sox9a2) expression in somatic cells is similar in both sexes during early gonadal differentiation and becomes positively regulated in males during the testicular lobe formation stage (Nakamoto et al. 2005; Nakamura et al. 2008). Overall, these observations support the hypothesis of a significant role for sox9 in either testicular differentiation or testicular development. This gene may also be at play in flatfish since, in Atlantic Halibut, sox9a2 is included in a cascade of gene activation starting with the dmrt1 gene which is activated by the putative SD factor, the gsdf gene (Norris and Carr 2013, Einfeldt et al 2021).

In Greenland Halibut, sox2, gdf6, and sox9a2 represent three possible candidate genes for sex determination. One can think that these three genes may be interacting in shaping the determination of sex and/or the subsequent development of sex organs. However, several lines of evidence suggest that sox2 might have a major role in Greenland Halibut SD. First, sox2 is located in a region that is more highly differentiated between sexes than gdf6 and sox9a2 genes. Furthermore, no functional evidence in sexual determination is available yet, neither for gdf6 nor

sox9a2 genes, in other fish. Conversely, sox2 has been reported as the main SD driver for several species including another flatfish (Turbot, Martinez et al 2021).

In contrast with sox2 identified in Turbot, SD genes identified in the two most-closely related flatfish species, Atlantic and Pacific Halibut (Table1), do not appear to differentiate males and females in the Greenland Halibut. This is somewhat unsurprising given that even within the same genus (*Hyppoglossus*), not only different species have different genes driving SD, but they harbor a different heterogametic sex (XY for Atlantic Halibut and ZW for Pacific Halibut, Drinan, Loher, & Hauser 2018, Einfeldt et al 2021).

The nascent SD system with low differentiation between sex chromosomes revealed in Greenland Halibut might reflect the observed complex and still enigmatic sex maturation. For example, in some Greenland Halibut individuals, the maturation cycle can be interrupted by a sudden degeneration of the oocytes (Fedorov, 1971). More recently, it has been suggested that this species utilizes a very unusual oocyte development pattern containing two simultaneously developing cohorts of vitellogenic oocytes, not seen in any related species (Rideout et al 2012; Dominguez-Petit et al 2017). Also, unusual and still unexplained, Greenland Halibut present high densities and regular occurrence of rodlet cells in male and female gonads (Rideout et al 2015). Furthermore, other layers may be added on the top of such an SD mechanism; environmental factors such as temperature, pH, population density, and social interactions have been found to influence the sex-ratio in fish (Crews 1996; Nakamura et al. 1998; Baroiller et al. 1999; Baroiller and D’Cotta 2001). More research would be needed to better disentangle the relative contribution of environmental factors and genetic factors in sex-determination in Greenland Halibut.

Conclusion:

Despite the commercial importance of Greenland Halibut, important gaps still persist in our knowledge of this species, including gaps about its reproduction cycle and sex determination mechanisms. In this study, we used a compendium of novel technologies (single molecule sequencing of long reads (Pacific Sciences) with Chromatin Conformation Capture sequencing (Hi-C) data and whole-genome sequencing of hundreds of individuals) to provide the very first chromosome-level genome reference and elucidate the sex determination mechanism of this demersal fish. The chromosome-level assembly and the investigation of its syntenic relationships with other economically important flatfish species highlight the high conservation of synteny blocks within the flatfish phylogenetic clade. Our sex determination analysis revealed that flatfishes do not escape this rule and exhibit a high plasticity and turnover in sex-determination mechanisms. In fish, the early evolutionary stage of most sex chromosomes and the subsequent small differences between those chromosomes make it extremely difficult to distinguish diagnostic differences between them. A whole-genome sequencing approach of hundreds of individuals allowed us to draw a full picture of the molecular SD system in Greenland Halibut. Specifically, our results suggested that Greenland Halibut has a very nascent male heterogametic XY system, with a putative major effect of the *sox2* gene, also described as the main SD driver in the Turbot (*S. maximus*) and to a lesser extent in the Japanese Flounder. Interestingly, our study also suggested for the first time in flatfish a putative Y-autosomal fusion that could be associated with a reduction of recombination typical of early steps of sex chromosome evolution.

598

599

600 ACKNOWLEDGMENTS:

601 We are grateful to Renée Gagné and Léopold Ginther, for assistance while sampling the
602 individual used for the genome assembly and transcriptomic RNAseq and collecting its tissues.

603 We thank all biologists and technicians of Fisheries and Oceans Canada for their implication in
604 the project and their field assistance. We also thank Cécilia Hernandez for help gathering all the
605 samples in the lab and Justine Létourneau for help in DNA extraction. Brian Boyle and Nina
606 Therkildsen provided key advice about sequencing and libraries preparation. The genome
607 assembly was supported by the Canada 150 Sequencing Initiative (Canseq150), CFI and Genome
608 Canada Technology Platform grants to J R. This research was funded by the Canadian Research
609 Chair in Genomics and Conservation of Aquatic Resources, Ressources Aquatiques Québec (RAQ)
610 as well as by a Strategic Project Grant from the Natural Sciences and Engineering Research
611 Council of Canada (NSERC) to L. Bernatchez.

612

613

References

- Abbott JK, Nordén AK, Hansson B (2016) Sex chromosome evolution: historical insights and future. *Proc Biol Sci*, **284**, 20162806.
- Aljanabi SM, Martinez I (2017) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.
- Bachtrog D (2006) A dynamic view of sex chromosome evolution. *Current Opinion in Genetics and Development*, **16**, 578–585.
- Bachtrog D (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, **14**, 113–124.
- Baroiller JF, DCotta H Environment and sex determination in farmed fish. *comparative Biochemistry and physiology, Part D*, **30**, 399–409.
- Baroiller JF, Guiguen Y, Fostier A (1999) Endocrine and environmental aspects of sex differentiation in fish. *Cell Mol Life Sci*, **55**, 910–931.
- Baym M, Kryazhimskiy S, Lieberman TD *et al.* (2015) Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PloS one*, 1–15.
- Benestan L, Moore J-S, Sutherland BJG *et al.* (2017) Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *molecular ecology*, **26**, 6767–6783.
- Bernatchez L, Wellenreuther M, Araneda C *et al.* (2017) Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends in Ecology & Evolution*, **32**, 665–680.

636 Betancur R, Ortí G (2014) Molecular evidence for the monophyly of flatfishes (Carangimorpharia:
637 Pleuronectiformes). *Molecular phylogenetics and evolution*, **73**, 18–22.

638 Boje J, Hareide N-R Trial deepwater longline fishery in the Davis Strait, May-June 1992. *NAFO SCR*
639 *Doc*, **93**, 1–6.

640 Bowering WB, Nedreaas KH A comparison of Greenland halibut (*Reinhardtius hippoglossoides*
641 (Walbaum)) fisheries and distribution in the Northwest and Northeast Atlantic. *Sarsia*, **85**,
642 61–76.

643 Bowering WB, Power D Spatula distribution of Greenland halibut (*Reinhardtius hippoglossoides*
644 (Walbaum)) in the Labrador - eastern Newfoundland area of the Canadian Northwest Atlantic.
645 *NAFO Sci. Counc. Stud.*, **22**, 51–61.

646 Bulls JJ *Evolution of sex determining mechanisms* (M Park, Ed,). Benjamin/Cummings Publishing
647 Co, CA.

648 Capel B (2017) Vertebrate sex determination: evolutionary plasticity of a fundamental switch.
649 *Nature Reviews Genetics*, **18**, 675–689.

650 Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex
651 chromosomes. *Heredity*, **95**, 118–128.

652 Chen S, Zhang G, Shao C *et al.* (2014) Whole-genome sequence of a flatfish provides insights into
653 ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*, **46**,
654 253–260.

655 Cheng Y, Shang D, Luo M *et al.* (2020) Whole genome-wide chromosome fusion and new gene
656 birth in the *Monopterus albus* genome. **10**, 67.

657 Crews D (1976) Temperature-dependent sex determination: the interplay of steroid hormones
658 and temperature. *Zoological Sciences*, **3**, 1–13.

659 Crews D, Bulls JJ (2009) Mode and tempo in environmental sex determination in vertebrates.
660 *Seminars in Cell & Developmental Biology*, **20**, 251–255.

661 Czerwinski M, Natarajan A, Barske L, Looger LL, Capel B (2016) A time course analysis of systemic
662 and gonadal effects of temperature on sexual development of the red-eared slider turtle
663 *Trachemys scripta elegans*. **420**, 166–177.

664 DFO (Fisheries and Oceans Canada), 2018. Assessment of the Greenland Halibut stock in the Gulf
665 of St. Lawrence (4RST) in 2017. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2018/035.

666 Dominguez-Petit R, Rideout RM, Garabana D *et al.* (2018) Original Article Evaluating the use of
667 the autodiametric method for estimating fecundity of *Reinhardtius hippoglossoides*, a
668 species with an unusual oocyte development strategy. *ICES Journal of Marine Science*, **75**,
669 831–839.

670 Drinan DP, Loher T, Hauser L (2017) Identification of Genomic Regions Associated With Sex in
671 Pacific Halibut (S Grant, Ed.). *Journal of Heredity*, **109**, 326–332.

672 Du QY, Wang FY, Hua HY (2007) Cloning and study of adult-tissue-specific expression of Sox9 in
673 *Cyprinus carpio*. *J Genet*, **86**, 85–91.

674 Dumas D, Britton-Davidian J (2002) Chromosomal rearrangements and evolution of
675 recombination: comparison of chiasma distribution patterns in standard and Robertsonian
676 populations of the house mouse. *Genetics*, **162**, 1355–1366.

Dwyer KS, Treble MA, Campana SE (2016) Age and growth of Greenland Halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic: A changing perception based on bomb radiocarbon analyses. *Fisheries Research*, **179**, 342–350.

Einfeldt AL, Kess T, Messmer A *et al.* (2021) Chromosome level reference of Atlantic halibut *Hippoglossus hippoglossus* provides insight into the evolution of sexual determination systems. *Molecular ecology resources*, 1–31.

Fan Z, You F, Wang L *et al.* (2014) Gonadal Transcriptome Analysis of Male and Female Olive Flounder (*Paralichthys olivaceus*). *BioMed research international*, **2014**, Article ID 291067.

Fedorov KE (1971) The state of the gonads of the Barents sea Greenland halibut, *Reinhardtius hippoglossoides* (Walb.) in connection with failure to spawn. *Vop, Ichtiol.*, **11**, 673–682.

Figueras A, Robledo D, Corvelo A *et al.* (2016) Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. **23**, 181–192.

Fox EA, Wright AE, Fumagalli M, Vieira FG (2019) ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, **35**, 3855–3856.

Furman BLS, Metzger DCH, Darolti I *et al.* (2020) Sex Chromosome Evolution: So Many Exceptions to the Rules (B Fraser, Ed.). *Genome Biology and Evolution*, **12**, 750–763.

Gao J, Wang Z, Shao K *et al.* (2014) Identification and characterization of a Sox2 homolog in the Japanese flounder *Paralichthys olivaceus*. *Gene*, **544**, 165–176.

Ghinter L, Lambert Y, Audet C (2019) Influence of sex on growth of juvenile Greenland halibut (*Reinhardtius hippoglossoides*) in the Estuary and Gulf of St. Lawrence. *Fisheries Research*, **219**, 105332.

- 698 Gibson RN, Nash RDM, Geffen AJ, Van der Veer HW *Flatfishes: Biology and Exploitation*, 2nd
699 *edition* (RN Gibson, RDM Nash, AJ Geffen, HW Van der Veer, Eds.).
- 700 Guerrero RF, Kirkpatrick M (2014) Local adaptation and the evolution of chromosome fusions.
701 *Evolution*, **68**, 2747–2756.
- 702 Guerrero-Estévez S, Moreno-Mendoza N (2010) Sexual determination and differentiation in
703 teleost fish. *Reviews in Fish Biology and Fisheries*, 101–121.
- 704 Guo L, Xu Y-H, Zhang N *et al.* (2019) A High-Density Genetic Linkage Map and QTL Mapping for
705 Sex in Black Tiger Shrimp (*Penaeus monodon*). *fgene-10-00326.tex*, **10**, 326.
- 706 Hagen O, Solberg C, Johnston IA (2006) Sexual dimorphism of fast muscle fibre recruitment in
707 farmed Atlantic halibut (*Hippoglossus hippoglossus* L.). *Aquaculture*, **261**, 1222–1229.
- 708 Hayes TB (1998) Sex Determination and Primary Sex Differentiation in Amphibians: Genetic and
709 Developmental Mechanisms. *The Journal of Experimental Biology*, 373–399.
- 710 Ijiri S, Kaneko H, Kobayashi T *et al.* (2008) Sexual dimorphic expression of genes in gonads during
711 early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. *Biol Reprod*, **78**,
712 333–341.
- 713 Ismland AK, Folkvord A, Grang GL, Stefanson SO (1997) Sexual dimorphism in growth and
714 maturation of turbot, *Scophthalmus maximus* (Rafinesque, 1810). *Aquac Res*, **28**, 101–114.
- 715 Kamiya T, Kai W, Tasumi S *et al.* (2012) A Trans-Species Missense SNP in Amhr2 Is Associated
716 with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, **8**,
717 e1002798.
- 718 Keyl F, Kempf AJ, Sell AF (2015) Sexual size dimorphism in three North Sea gadoids. *Journal of*
719 *Fish Biology*, **86**, 261–275.

720 Kielbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence
721 comparison. *Genome Research*, **21**, 487–493.

722 Kirkpatrick M How and why chromosome inversions evolve. *PLOS Biology*, **8**, e1000501.

723 Kitano J, Peichel CL (2012) Turnover of sex chromosomes and speciation in fishes. *Environmental*
724 *Biology of Fishes*, **94**, 549–558.

725 Kitano J, Ross J, Mori S (2009) A role for a neo-sex chromosome in stickleback speciation. *Nature*,
726 **461**, 1079–1083.

727 Kondo M, Hornung U, Nanda I *et al.* (2006) Genomic organization of the sex-determining and
728 adjacent regions of the sex chromosomes of medaka. *Genome Research*, 815–826.

729 Korneliussen TS, Albrechtsen A, Nielsen R ANGSD: Analysis of Next Generation Sequencing Data.
730 *BMC Bioinformatics*, **15**.

731 Kurtz S, Phillippy A, Delcher AL *et al.* (2004) Versatile and open software for comparing large
732 genomes. *Genome biology*, **5**, R12.

733 Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier Science.

734 Lehnert SJ, Bentzen P, Kess T *et al.* (2019) Chromosome polymorphisms track trans-Atlantic
735 divergence and secondary contact in Atlantic salmon. *Molecular ecology*, **28**, 2074–2087.

736 Liang D, Fan Z, Zou Y *et al.* (2018) Characteristics of Cyp11a during Gonad Differentiation of the
737 Olive Flounder *Paralichthys olivaceus*. *International Journal of Molecular Sciences*, **19**, 2641.

738 Liang S, Liu D, Li X *et al.* (2019) SOX2 participates in spermatogenesis of Zhikong scallop *Chlamys*
739 *farreri*. *Scientific Reports*, **9**, 76.

740 Lieberman-Aiden E, van Berkim NI, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR,
741 Sabo PJ, Dorschner MO et al. (2009). Comprehensive mapping of long-range-interactions
742 reveals folding principles of the human genome. *Science*. **326** (5950):289–293.

743 Lou RN, Jacobs A, Wilder AP, Therkildsen NO (2021) A beginner’s guide to low-coverage whole
744 genome. *authorea preprints*, 1–68.

745 Martinez P, Robledo D, Taboada X *et al.* (2021) A genome-wide association study, supported by a
746 new chromosome-level genome assembly, suggests sox2 as a main driver of the
747 undifferentiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). 1–54.

748 Martinez P, Viñas AM, Sánchez L *et al.* (2014) Genetic architecture of sex determination in fish:
749 applications to sex ratio control in aquaculture. *Frontiers in Genetics*, **5**, 1–13.

750 McAllyster BF (2003) Sequence Differentiation Associated With an Inversion on the Neo-
751 X Chromosome of *Drosophila americana*. *Genetics*, **165**, 1317–1328.

752 McGrath C (2020) Highlight: Exceptions Are the Rule in Sex Determination (G Zhang, Ed.).
753 *Genome Biology and Evolution*, **12**, 764–765.

754 Meisner J, Albrechtsen A (2018) Inferring Population Structure and Admixture Proportions in
755 Low-Depth NGS Data. *Genetics*, **210**, 719–731.

756 Mérot C, Berdan E, Cayuela H *et al.* (2021) Locally adaptive inversions modulate genetic variation
757 at different geographic scales in a seaweed fly. *Molecular Biology and Evolution*, msab143.

758 Nakamoto M, Suzuki A, Matsuda M, Nagahama Y, Shibata N (2005) Testicular type Sox9 is not
759 involved in sex determination but might be in the development of testicular structures in the
760 medaka, *Oryzias latipes*. *Biochem Biophys Res Commun*, **333**, 729–736.

761 Nakamura M, Kobayashi T, Chang XT, Nagahama Y (1998) Gonadal sex differentiation in teleost
762 fish. *Journal of Experimental Zoology*, **281**, 362–372.

763 Nanda I, Kondo M, Hornung U *et al.* (2002) A duplicated copy of DMRT1 in the sex-determining
764 region of the Y chromosome of the medaka, *Oryzias latipes*. *PNAS*, **99**, 11778–11783.

765 Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation
766 sequencing data. 1–9.

767 Norris DO, Carr JA (2013) Comparative Aspects of Vertebrate Reproduction. In: *Vertebrate*
768 *Endocrinology Fifth Edition* (eds Norris DO, Carr JA), pp. 375–441.

769 Ospina-Alvarez N, Piferrer F (2008) Temperature-Dependent Sex Determination in Fish Revisited:
770 Prevalence, a Single Sex Ratio Response Pattern, and Possible Effects of Climate Change. *PloS*
771 *one*, **3**, e2837.

772 Ovenden JR, Berry O, Welch DJ, Buckworth RC, Dichmont CM (2013) Ocean's eleven: a critical
773 evaluation of the role of population, evolutionary and molecular genetics in the
774 management of wild fisheries. *Fish and Fisheries*, **16**, 125–159.

775 Palaikostas C, Bekaert M, Davie A *et al.* (2013) Mapping the sex determination locus in the
776 Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics*, **14**.

777 Patra SK, Chakrapani V, Panda RP *et al.* (2015) First evidence of molecular characterization of
778 rohu carp Sox2 gene being expressed in proliferating spermatogonial cells. *Theriogenology*,
779 **84**, 268–276.e1.

780 Pennel MW, Kirkpatrick M, Otto SP *et al.* (2015) Y Fuse? Sex Chromosome Fusions in Fishes and
781 Reptiles. *PLoS Genetics*, 1–17.

782 Piferrer F, Cal R, Gomez C *et al.* (2004) Induction of gynogenesis in the turbot (*Scophthalmus*
783 *maximus*): Effects of UV irradiation on sperm motility, the Hertwig effect and viability during
784 the first 6 months of age. **238**, 403–419.

785 Rastetter RH, Smith CA, Wilhelm D (2015) The role of non-coding RNAs in male sex
786 determination and differentiation. *Reproduction*, **150**, R93–R107.

787 R Core Team (2020). R: A language and environment for statistical computing. R Foundation for
788 Statistical computing, Vienna, Austria. URL <https://www.R-project.org/>

789 Reichwald K, Petzold A, Koch P *et al.* (2015) Insights into Sex Chromosome Evolution and Aging
790 from the Genome of a Short-Lived Fish. *Cell*, **163**, 1527–1538.

791 Rideout RM, Morgan MJ, Lambert Y *et al.* Oocyte development and vitellogenin production in
792 Northwest Atlantic Greenland halibut *Reinhardtius hippoglossoides*. *Journal of Northwest*
793 *Atlantic Fishery Science*, **44**, 15–29.

794 Rideout RM, Smith SA, Morgan MJ (2015) High-density aggregations of rodlet cells in the gonads
795 of Greenland halibut *Reinhardtius hippoglossoides*, a deep-water marine flatfish. *Journal of*
796 *Fish Biology*, **86**, 1630–1637.

797 Rijnsdorp AD, Ibelings B (1989) Sexual dimorphism in the energetics of reproduction and growth
798 of the North Sea plaice, *Pleuronectes platessa* L. *Journal of Fish Biology*, **401-415**.

799 Robledo D, Hermida M, Rubiolo JA *et al.* (2017) Integrating genomic resources of flatfish
800 (Pleuronectiformes) to boost aquaculture production. *Comparative Biochemistry and*
801 *physiology, Part D*, **21**, 41–55.

802 Ruan J, Li H (2020) Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, **17**, 155–
803 158.

804 Scott WB, Scott MG Atlantic fishes of Canada 219. *Canadian Bulletin of fisheries aquatic Science*.

805 Shao C, Bao B, Xie Z *et al.* (2017) The genome and transcriptome of Japanese flounder provide
806 insights into flatfish asymmetry. *Nature Genetics*, **49**, 119–128.

807 Shen XY, Cui JZ, Gong QL, Nagahama Y (2007) Cloning of the full-length coding sequence and
808 expression analysis of Sox9b in guppy (*Poecilia reticulata*). *Fish Physiol Biochem*, **33**, 195–
809 202.

810 Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing
811 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*,
812 **31**, 3210–3212.

813 Smit, AFA, Hubley, R & Green, P. 2013-2015 *RepeatMasker Open-4.0*.
814 <http://www.repeatmasker.org>

815 Soderlund C, Bomhoff M, Nelson W (2011a) SyMAP v3.4: a turnkey synteny system with
816 application to plant genomes. *Nucleic Acids Research*, **39**, e68.

817 Soderlund C, Bomhoff M, Nelson WM (2011b) SyMAP v3.4: a turnkey synteny system with
818 application to plant genomes. *Nucleic Acids Research*, **39**, e68.

819 Soderlund C, Nelson W, Shoemaker A, Paterson A (2006) SyMAP: A system for discovering and
820 viewing syntenic regions of FPC maps. *Genome Research*, **16**, 1159–1168.

821 Sutherland BJG, Rico C, Audet C, Bernatchez L (2017) Sex Chromosome Evolution,
822 Heterochiasmy, and Physiological QTL in the Salmonid Brook Charr *Salvelinus fontinalis*. *G3*:
823 *GenesGenomesGenetics*, **7**, 2749–2762.

824 Takamatsu N, Kanda H, Ito M *et al.* (1997) Rainbow trout SOX9: cDNA cloning, gene structure
825 and expression. *Gene*, **202**, 167–170.

826 Tavares H (2021). windowscanr: Apply functions using sliding windows. R package version 0.1.

827 Therkildsen NO, Palumbi SR (2016) Practical low-coverage genomewide sequencing of hundreds
828 of individually barcoded samples for population and evolutionary genomics in nonmodel
829 species. *Molecular ecology resources*, 1–15.

830 Tvedt H, Benfey TJ, Martin-Robichaud DJ, McGowan C, Reith M (2006) Gynogenesis and sex
831 determination in Atlantic Halibut (*Hippoglossus hippoglossus*). *Aquaculture*, **252**, 573–583.

832 Uller T, Pen I, Wapstra E, Beukeboom LW, Kombeur J (2007) The evolution of sex ratios and sex-
833 determining systems. *Trends in Ecology & Evolution*, **22**, 292–297.

834 Vinnikov KA, Thomson RC, Munroe TA (2018) Revised classification of the righteye flounders
835 (Teleostei: *Pleuronectidae*) based on multilocus phylogeny with complete taxon sampling.
836 *Molecular phylogenetics and evolution*, **125**, 147–162.

837 Vizziano D, Randuineau G, Baron D, Cauty C, Guiguen Y (2007) Characterization of early
838 molecular sex differentiation in rainbow trout, *Oncorhynchus mykiss*. *Dev Dyn*, **236**, 2198–
839 2206.

840 VonHofsten J, Olsson PE (2005) Zebrafish sex determination and differentiation: involvement of
841 FTZ-F1 genes. *Rep Biol Endocrinol*, **3**, 63.

842 Wellband K, Mérot C, Linnansaari T *et al.* (2019) Chromosomal fusion and life history-associated
843 genomic variation contribute to within-river local adaptation of Atlantic salmon. *molecular*
844 *ecology*, **28**, 1439–1459.

845 Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

846 Wilkins AS (2004) Moving up the hierarchy: A hypothesis on the evolution of a genetic sex
847 determination pathway. *BioEssays*, **17**, 71–77.

- Wright A, Dean R, Zimmer F, Mank JE (2016) How to make a sex chromosome. *Nature communications*, **7**, 12087.
- Yao HH-C, Capel B (2005) Temperature, genes, and sex: a comparative view of sex determination in *Trachemys scripta* and *Mus musculus*. *Journal of Biochemistry*, **138**, 5–12.
- Yokoi H, Kobayashi T, Tanaka M *et al.* (2002) Sox9 in a teleost fish, medaka (*Oryzias latipes*): evidence for diversified function of Sox9 in gonad differentiation. *Mol Reprod Dev*, **63**, 5–16.
- Yano A, Nicol B, Jouanno E *et al.* (2013) The sexually dimorphic on the Y chromosome gene (sdY) is a conserved male specific Y chromosome sequence in many salmonids. *Evolutionary Applications*, **6**, 486–496.
- Zhang B, Zhao N, Peng K *et al.* (2020) A combination of genome-wide association study screening and SNaPshot for detecting sex-related SNPs and genes in *Cynoglossus semilaevis*. *Comparative Biochemistry and physiology, Part D*, **35**, 100711.
- Zhou RJ, Liu L, Guo LQ *et al.* (2003) Similar gene structure of two Sox9a genes and their expression patterns during gonadal differentiation in a teleost fish, rice field eel (*Monopterus albus*). *Mol Reprod Dev*, **66**, 211–217.

FIGURES:

Figure 1: Phylogenetic relationships and syntenic distances between important flatfish species in fisheries and aquaculture. Phylogenetic trees were retraced according to previous studies, Betancour et al (2014) for the top one and Vinnikov et al (2018) for the zoom into the sub-family *Hippoglossinae* framed in red. Only the names of relevant species compared in our studies are reported as well as the SD system, either “XY” for a male heterogametic system (XX/XY) or “ZW” for a female heterogametic system (ZZ/ZW). The XY system for Greenland Halibut has been assessed in the present study. Circos plots showing synteny between the Greenland Halibut (displayed in low half circle) and the Tongue Sole, Turbot, Japanese Flounder, Atlantic and Pacific Halibut. Black arrows in each circos plot refers to the sex chromosomes or the location of sex-related markers. A positive relationship between the mean block size and phylogenetic distance is also reported. Colors used for reporting the standard deviation and the mean of block size are indicated next to each species name respectively. Note that the phylogenetic distance reported on the x-axis is qualitative only and that species have been positioned from the most closely related species to the most distant species. The same color code as the one near the circus plots is used.

Figure 2: Two genomic regions implied in sex differentiation. (A) Correlation between PC1 scores of the local PCAs performed on windows of 1000 SNPs and PC1 scores of the global PCA along their genomic position in each chromosome. **(C)** Fst differentiation estimated between males and females in sliding-windows of 25 Kb within each chromosome. **(E)** Mean observed heterozygosity in each sex smoothed for visualization. **(B), (D), and (F)** are zooms on the two

892 differentiated chromosomes (chr 10 and 21) of the previous values, respectively (A), (C), and (E).

893 Vertical dashed grey lines indicate the positions of detected sex-related candidate genes.

894 **Figure 3: Intra- and Inter- chromosomal Linkage Disequilibrium (LD) among males and females,**

895 **revealing a putative Y-autosomal fusion.** Male values are displayed above the diagonal and female

896 values are displayed below the diagonal. The color scale corresponds to the 10th highest

897 percentile of the R² value of each SNP pairwise summarized by windows of 250 Kb. Positions are

898 represented in megabases.

899

900 Table 1: Statistics of genome assemblies of the important flatfish species in fisheries and aquaculture. n.a. is used for data not available

901 ⁽¹⁾ This study, ⁽²⁾ Drinan, Loher, & Hauser, 2018, ⁽³⁾ Einfeldt et al 2021, ⁽⁴⁾ Liang et al 2018, ⁽⁵⁾ Martinez et al 2021, ⁽⁶⁾ Cheng et al

902 2014

903

904

	Greenland Halibut	Pacific Halibut	Atlantic Halibut	Japanese flounder	Turbot	Tongue sole
Species latin name	<i>Reinhardtius hippoglossoides</i>	<i>Hippoglossus stenolepis</i>	<i>Hippoglossus hippoglossus</i>	<i>Paralichthys olivaceus</i>	<i>Scophthalmus maximus</i>	<i>Cynoglossus semilaevis</i>
total sequence length	598,525,902	594,145,200	596,792,615	545,775,252	524,979,463	470,199,494
number of scaffolds	1,594	120	57	9,525	22	31,181
max scaffolds lengh	31,982,681	32,413,155	30,753,830	27,810,759	31,901,587	34,528,841
scaffold N50	25,015,410	24,985,257	26,312,018	3,817,360	24,811,384	509,861
scaffold L50	11	11	11	40	10	289
total number of chromosomes	24	24	24	24	22	21
Genbank accession number	GCA_006182925.3	GCA_013339905.1	GCA_009819705.1	GCA_001904815.2	GCA_003186165.1	GCA_000523025.1
Reference	This study	International	Einfeldt et al 2021	Shao et al 2017	Martinez et al	Chen et al 2014

		Pacific Halibut Commission			2021	
Name of the assembly	UL_REHI_2.0	IPHC_HiSten_1.0	fHipHip1	ParOli_1.1	ASM318616v1	Cse_v1.0
Complete BUSCO	98.04%	na	96.90%	na	na	na
Repeat elements	8.20%	7.28%	7.14%	4.92%	6.35%	6.08%
Genetic sex-determinism	XX//XY	ZZ/ZW	XX/XY	XX/XY	ZZ/ZW	ZZ/ZW
SD genes or putative SD genes	sox2, sox9a2, gdf6 (1)	no gene has been specifically identified (2)	gsdf (3)	Cyp11a (4)	sox2 (5)	dmrt1 (6)
Homolog of sex chromosome in Greenland Halibut	-	GH06, GH12, G20	GH18	GH19	GH14	GH09

Table 2: Syntenic relations between Greenland Halibut and related commercial flatfish species.

	Greenland Halibut / Pacific Halibut	Greenland Halibut / Atlantic Halibut	Greenland Halibut / Japanese Flounder	Greenland Halibut / Turbot	Greenland Halibut / Tongue sole
number of blocks	40	45	191	333	745
number of chr with single block	16	15	0	0	0
min size	560,659	537,880	39,610	23,190	25,486
max size	28,459,961	30,361,793	21,363,777	24,314,634	16,925,444
median size	24,561,162	11,813,728	1,877,480	507,912	482,179
nb of block > 10Mb	24	24	11	23	6
percent of blocks > 10Mb (%)	60	53	6	7	1

910 Figure 1

911

914 Figure 2

915

916

917

918

919

920

921

922

923

Figure 3