Heterochiasmy facilitated the establishment of *gsdf* as a novel sex determining gene in Atlantic halibut

Edvardsen, R.B*[1]., Wallerman, O[3]., Furmanek, T[1]., Kleppe, L[1]., Jern, P[3], Wallberg, A[3]., Kjærner-Semb, E[1]., Mæhle, S[1]., Olausson, S. K[2]., Sundström, E[3]., Harboe, T[2]., Mangor-Jensen, R[2]., Møgster, M[2]., Perrichon, P[2]., Norberg, B.[2], Rubin, C-J*[3]

*Corresponding authors

rolf.brudvik.edvardsen@hi.no

carl-johan.rubin@imbim.uu.se

Affiliations:

[1] Institute of Marine Research, P.O. Box 1870, Nordnes, NO-5817, Bergen, Norway.

[2] Institute of Marine Research, Austevoll Research Station, NO-5392, Storebø, Norway.

[3] Department of Medical Biochemistry and Microbiology, Uppsala University, 75123 Uppsala, Sweden.

Keywords: aquaculture, nucleotide diversity, transposable elements, genome assembly, meiotic recombination, sex chromosome

Abstract

Atlantic Halibut (Hippoglossus hippoglossus) has a X/Y genetic sex determination system, but the sex determining factor is not known. We produced a high-quality genome assembly and identified parts of chromosome 13 as the Y chromosome due to sequence divergence between sexes and segregation of sex genotypes in pedigrees. Linkage analysis revealed that all

chromosomes exhibit heterochiasmy, i.e. male- and female restricted meiotic recombination intervals (MRR/FRR). We show that FRR/MRR intervals differ in nucleotide diversity and repeat class content and that this is true also for other Pleuronectidae species. We further show that remnants of a Gypsy-like transposable element insertion on chr13 promotes early male specific expression of gonadal somatic cell derived factor (gsdf). Less than 4 MYA, this male-determining element evolved on an autosomal FRR segment featuring pre-existing male meiotic recombination barriers, thereby creating a Y chromosome. We propose that heterochiasmy may facilitate the evolution of genetic sex determination systems.

Introduction

The Atlantic halibut (*Hippoglossus hippoglossus*) is the largest teleost species in the North Atlantic and can reach more than 50 years of age. In recent years, the demand for Atlantic halibut has been larger than the supply of wild fish. In some areas the stocks have been defined as vulnerable and sensitive to overfishing, due to relatively slow growth and high age at maturity. In combination with a high market price, this has led to an increased interest in halibut aquaculture. Female halibuts can weigh more than 350 kg and become more than 3.6 m long, while males rarely weigh more than 50 kg. A problem encountered in Atlantic halibut aquaculture is that males mature earlier and at a smaller size than females[1], and efficient elimination of male halibut in production of market fish would therefore constitute a particularly great advantage in halibut compared to other aquaculture fish species. Atlantic halibut has an XX/XY, male heterogametic, sex determination system[2]. The most common and practical method to achieve production of all female populations is sex reversal of F1 female juveniles, to create phenotypic and functional males with an XX karyotype[3,4]. The offspring of these sex-reversed genetic females will be 100% female, but identifying and separating XX males from XY males involves raising offspring sibling groups until their sex can be visually

determined by histology or microscopy of gonads[4,5], a time- and resource-consuming process that can take up to four years. A reliable and practical method for determination of genetic sex based on identification of Y chromosome markers is instrumental for the development of efficient protocols for all-female production.

In contrast to the highly conserved mammalian and avian sex determination systems, fish generally display much more plasticity and turnover when it comes to sex chromosomes and sex determination systems, with many already described systems being young in comparison to their mammalian and avian counterparts[6]. Furthermore, teleosts display a rich diversity of sex-determining mechanisms, both environmental and genetic[7]. As for genetic control of sex, both XX/XY and ZZ/ZW systems have been described in teleosts, sometimes with closely related species having differing systems, such as for sticklebacks[8], tilapias[9], swordtail fish[10] and mosquitofish[11]. Righteye flounders (*Pleuronectidae*) have variable genetic sex determinations systems with both XX/XY and ZZ/ZW systems represented[12]. The Pacific Halibut (*Hippoglossus stenolepis*), which diverged from the Atlantic halibut about 3.7 million years ago[2], has a ZZ/ZW system[13]. Several master sex determining (MSD) genes have been identified in teleosts, most of which belong to one of three protein families (DMRT, SOX, and TGF-ß and its signaling pathway)[14], and these have originated either from sub- or neo-functionalization of duplicated genes, or by allelic diversification[6]. Exceptions to these protein families are found for salmonids, where *sdy* is the MSD gene[15] and recently the *breast cancer anti-estrogen resistance protein 1* (*BCAR1*) and *bone morphogenetic protein receptor type-1B* (*bmpr1b*) have been reported as MSD genes in catfish (*Ictalurus punctatus*)[16] and Atlantic herring (*Clupea harengus*)[17], respectively. Thus, a wide range of systems have evolved for genetic control of sex determination in fish and additional complexity is added by environmental variables contributing to sex determination in some species.

76

77 Single molecule sequencing of long reads has transformed the way the scientific community
78 produces genome assemblies[18]. In relation to short-read sequencing technologies, methods
79 capable of sequencing long reads can greatly increase the contiguity of genome assemblies due
80 to long reads bridging repetitive parts of the genome. In this project we used an Oxford
81 Nanopore Technologies (ONT) MinION device to sequence long native DNA molecules from
82 a single male Atlantic halibut and used the assembled contigs in conjunction with Chromatin
83 Conformation Capture sequencing (HiC) data to produce a chromosome-scale scaffolded
84 genome assembly. To identify and characterize the sex chromosome we mapped short read data
85 from male and female DNA pools onto the assembly to screen for divergent genetic markers
86 between the sexes. The differing systems within the *hippoglossus* genus indicates that at least
87 one of Atlantic- and Pacific halibut must have evolved a novel sex determination system after
88 their 3.7 MYA split and that the mechanism for genetic sex determination in Atlantic halibut
89 may be recent on an evolutionary time scale. Thus, our high-quality Atlantic halibut genome
90 assembly offers great potential to study the emergence of a young genetic sex determination
91 system and evolution of sex chromosomes.

92

93 **Results**

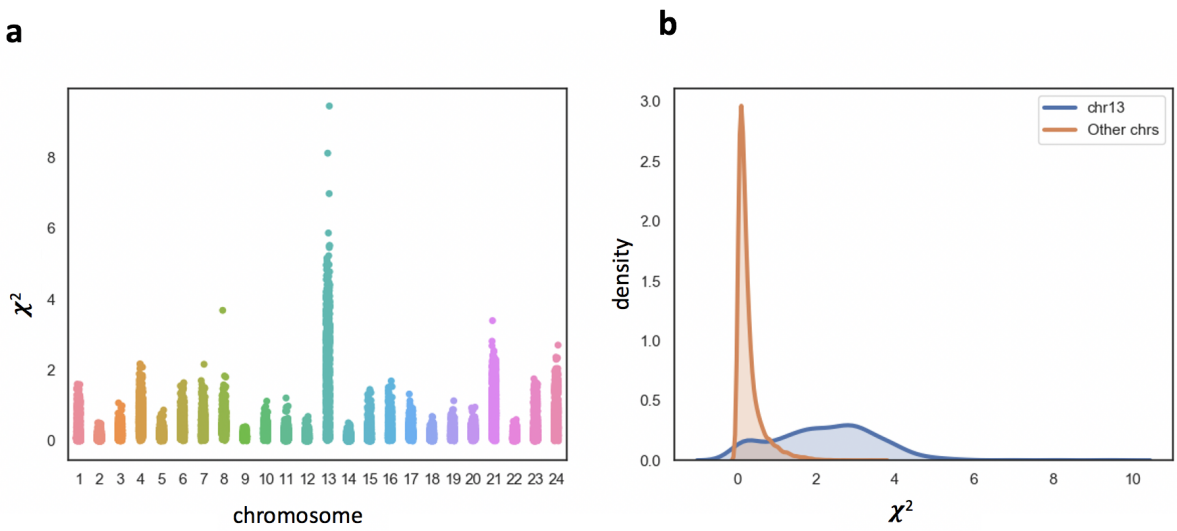
94

95 **Genome assembly and annotation**

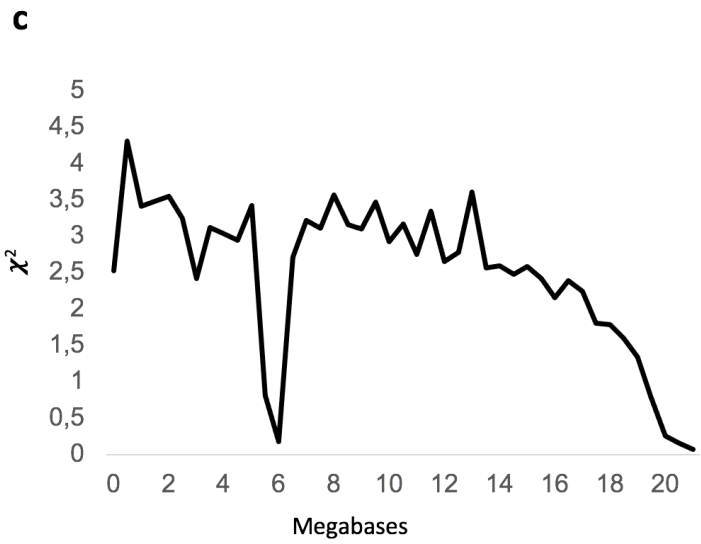
96 We used nanopore sequencing in combination with Illumina sequencing of 10X Genomics
97 Chromium-, paired-end- and HiC libraries to generate a chromosome-scale assembly of the
98 Atlantic halibut. The final scaffolded assembly (IMR_hipHip.v1) had a scaffold N50 of 27 Mb
99 (Supplementary Fig. 1a,b) and this assembly was evaluated using BUSCO using 4584 genes in
100 *Actinopterygii* odb9 data set, annotating 95.4% of analyzed genes as "Complete" in the genome

101 assembly. The assembly is organized into 24 larger scaffolds containing 97.6% of the
102 assembled sequence (Supplementary Fig. 2), corresponding to the chromosomes inferred from
103 karyotyping[19] and linkage analysis[20]. A number of smaller contigs and scaffolds, together
104 comprising only 2.4 % of the assembled sequence were left as unplaced. We named the
105 scaffolds by mapping microsatellite primer sequences from a published genetic linkage map for
106 Atlantic halibut[20] to IMR_hipHip.v1. Genome to genome alignment of IMR_hipHip.v1 to a
107 recently finished reference genome of the Pacific halibut (NCBI accession GCF_013339905.1)
108 revealed a largely co-linear genome organization, with few large-scale disagreements
109 (Supplementary Fig. 1c) as expected due to the recent split of the two species. To generate
110 RNA-seq data for genome annotation and to obtain transcriptomic data describing the early
111 stages of development, we used samples from different time points throughout embryonic
112 development from newly fertilized to 54 days post fertilization (dpf). For the first four stages
113 (1 hour post fertilization, 5 dpf, 8 dpf and 12 dpf), four pools of eight individuals were collected
114 and subjected to RNA extraction and Illumina HiSeq mRNA-sequencing. For the 28 dpf, 43
115 dpf and 54 dpf, four individual fish were sampled and sequenced for each stage. Genome
116 annotation based on the RNA-sequencing identified 18,314 genes, of which 18,092 (98.8%)
117 significantly matched with proteins in public databases.
118
119 **Identification of the Atlantic halibut sex chromosome**
120 In order to identify and characterize the sex chromosomes we sampled and extracted and pooled
121 DNA from ten full siblings of each sex and sequenced these using Illumina HiSeq sequencing.
122 The sequencing data was mapped against the IMR_hipHip.v1 assembly and the sequence
123 divergence between the male and female pools were determined for non-overlapping windows
124 of 100 SNPs along each chromosome using χ^2-tests comparing male- and female allele
125 frequencies. This screen revealed chromosome 13 as the most diverged between males and

126 females (Fig. 1a), in agreement with a previously published linkage study using RAD-
127 sequencing[21]. It should be noted that the haploid nature of the IMR_hipHip.v1 assembly makes
128 it possible that chr13 in this assembly is a hybrid between an X and a Y, which is what we find
129 when comparing the pool-seq genotypes of male vs. female divergent SNPs. The divergence
130 between phenotypic males and females on chr13 is not uniform, but overall, gradually declines
131 along the length of the chromosome, with the end being no more divergent than autosomes (Fig
132 1b and c).
133



134



135

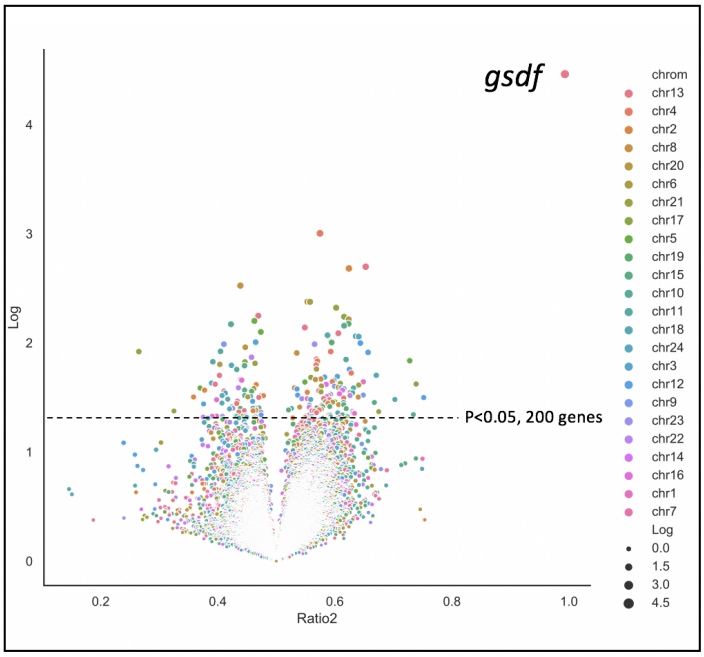
136 **Fig. 1**: Genetic differentiation between phenotypic males and females indicates chr13 as the Y-chromosome. **a**
137 Average χ^2-values for 100 SNPs along each assembled chromosome in the Atlantic halibut genome. **b** Distribution
138 of average χ^2-values for chr13 compared to the rest of the genome. **c** Average $\chi 2$-values for 500 kb windows along
139 chr13.
140

141 **Identification of the Atlantic halibut sex determining gene**

142 RNA-seq data was generated from 12 individuals covering three developmental stages with the
143 assumption that both genetic sexes would be represented. As we were unable to infer
144 phenotypic sex for these samples, an unbiased screen could involve searching for the most
145 variably expressed genes among the 12 individuals, but after conducting a Principal Component
146 Analysis we concluded that the samples clustered strongly by development stage
147 (Supplementary Fig. 3). An alternative strategy could involve screening for variable gene
148 expression between known phenotypic sexes during gonadal differentiation. As known
149 phenotypic sex was unavailable, we determined the genetic sex of the RNA-seq samples (XY
150 or XX). We first mapped the DNA-seq pools to the genome and extracted SNPs that had
151 approximately 50% allele frequency in the male pool-seq while being fixed for one allele in the
152 female pool. We next mapped the RNA-seq samples to the genome, extracted RNA-seq alleles
153 at the extracted divergent positions and performed hierarchical clustering (Supplementary Fig.
154 4) showing clearly that 7 samples were genetically male (XY) and 5 were genetically female
155 (XX). We then produced a Volcano plot showing the global genetic male vs. female expression
156 ratio plotted against the statistical significance of a differential expression (DE) between sexes
157 (Fig. 2a). The most significantly DE gene was *gsdf*, previously shown to be the male-
158 determining MSD gene in Luzon ricefish and sablefish[22,23]. Furthermore, among genes
159 previously identified as MSD genes in other fish species[14], only *gsdf* resides on chr13 in Atlantic
160 halibut (chr13:8506700-8510213). Among 12 RNA-seq individuals the 5 XX fish had almost
161 no *gsdf* expression, while the 7 XY fish all had high expression, consistent with Gsdf being a

162 male determining factor. To verify sex specific expression of *gsdf* during development and to
163 obtain better temporal resolution of expression levels, we performed quantitative PCR (qPCR)
164 on PCR sex-determined samples (Fig. 2b and c). RNA was extracted from individual embryos
165 from 11 development stages, ranging from 1 hour post fertilization (hpf) to 13 dpf. A PCR assay
166 based on SNPs in the 3'UTR of *brx* (chr13:9125004-9125007) was used to determine genetic
167 sex of samples from 48 hpf and onwards (Supplementary Fig. 5). The qPCR results show that
168 the expression of *gsdf* starts in males already at 48 hpf, and from 7 dpf onwards there is a clear
169 increase in expression compared to the earlier stages. For one individual sampled at 6 dpf
170 determined as a genetic male, we could not detect any expression, possibly due to non-genetic
171 factors having the chance to overrule the genotype in the determination of phenotypic sex in
172 the species. *in silico* searches for conserved genes located up- and downstream of *gsdf*
173 confirmed the *gsdf* identity. (Supplementary Fig. 6).
174

a



175
176

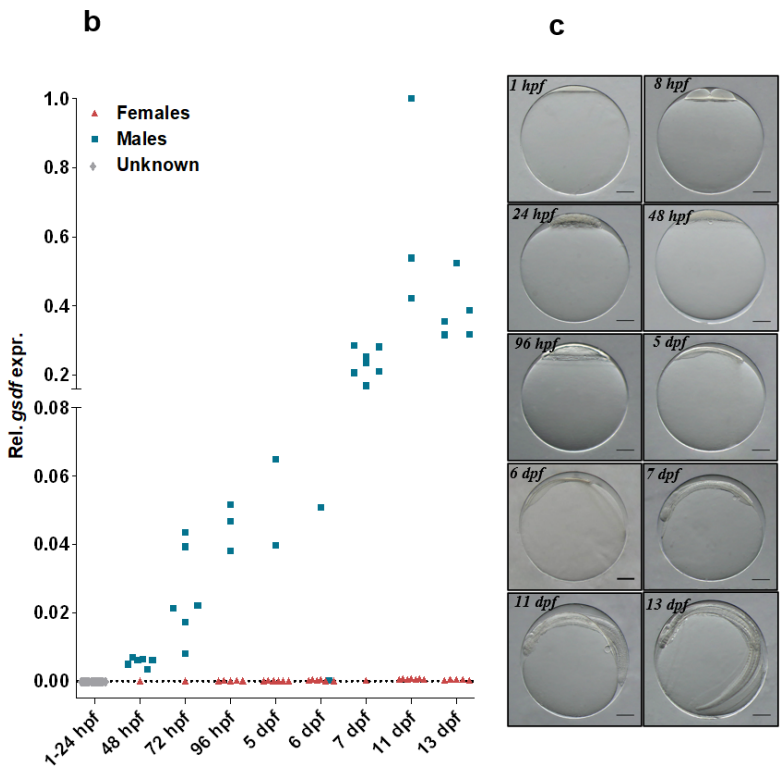


Fig. 2: Identification of *gsdf* as the Atlantic halibut sex determining gene. **a** Volcano plot representation of differential expression (DE) between individuals stratified by chr13 genotyping-based sex assignment. Dot colors indicate the chromosome from which genes are expressed and dot sizes were made proportional to the significance of DE (higher significance=larger dots). The most highly differentially expressed gene, *gsdf*, is indicated in the upper right corner. **b** qPCR result showing the sex specific *gsdf* expression (relative to *gtf3c6*) across developmental stages. **c** Pictures of fish representing the developmental stages sampled for gene expression analyses; 1 hpf to 13 dpf.

Most of the Atlantic halibut genome is subject to male or female restricted meiotic recombination

In order to validate chr13 as the Atlantic halibut sex chromosome we downloaded the pedigree RAD-sequencing data previously used to assign linkage group 13 as X/Y[21], aligned the data to our assembly IMR_hipHip.v1, called SNPs and investigated genotypes by sex phenotype in the pedigrees. The pedigree analysis verified our divergence-based assignment of IMR_hipHip.v1 chr13 as the sex chromosome and 90/94 F1 individuals were in agreement with regard to the

193 association between phenotypic sex and chr13 genotype. A similar observation was made for
194 the embryonal qPCR analysis where one out of 34 genetic males (*brx*-assay) had no *gsdf*
195 expression. The reason for the non-perfect association between genetic- and phenotypic sex
196 could be erroneous annotation of the downloaded RAD-seq dataset or a logical consequence of
197 non-perfect penetrance of young genetic sex determination systems. In our genome assembly-
198 guided linkage analysis, we detected suppression of meiotic recombination along a large part
199 of chr13 in male germ cells as would be expected in an X/Y system. Surprisingly, the majority
200 of the sex chromosome pair exhibited strict male- or female restricted meiotic recombination
201 (MRR/FRR) (Fig. 3a, Supplementary Fig. 7a). The male- and female genetic maps both sum
202 up to 49 cM but the recombining intervals are nonoverlapping. None of the 90 F1 offspring
203 used for analysis carried a male meiotic recombination in the first ~15 Mb, while all detected
204 female meiotic recombination events on chr13 occurred in this region. Thus, the first ~15 Mb
205 of chr13 behaves like a X/Y chromosome, with high male vs. female divergence (Fig. 1c) and
206 nonexistent meiotic recombination in males**.** The later part of chr13 (~15 Mb until chromosome
207 end) only recombines in males (X/Y recombination) and has lower male vs. female divergence,
208 suggesting that sequence diversity in this part of the chromosome is shared between X and Y.
209 Even more surprising than the meiotic recombination barrier observed for the of chr13 in
210 females was the observation that all chromosomes clustered into more or less strict MRR/FRR
211 regions (Fig. 3b, Supplementary Table 1 and Supplementary Fig. 7a). We limited our analysis
212 to the placed chromosome part of the genome (chromosomes 1-24), spanning 596 Mb out of
213 the total 611 Mb, and defined FRRs and MRRs according to our classification outlined in
214 Supplementary Table 2. Genome-wide, this annotation defined 407 Mb (68% of the sequence)
215 as FRRs and 154 Mb as MRRs (26% of the sequence) and only ~6% of the sequence fell outside
216 of this classification, i.e. either both sexes or neither sex showed meiotic recombination**.**

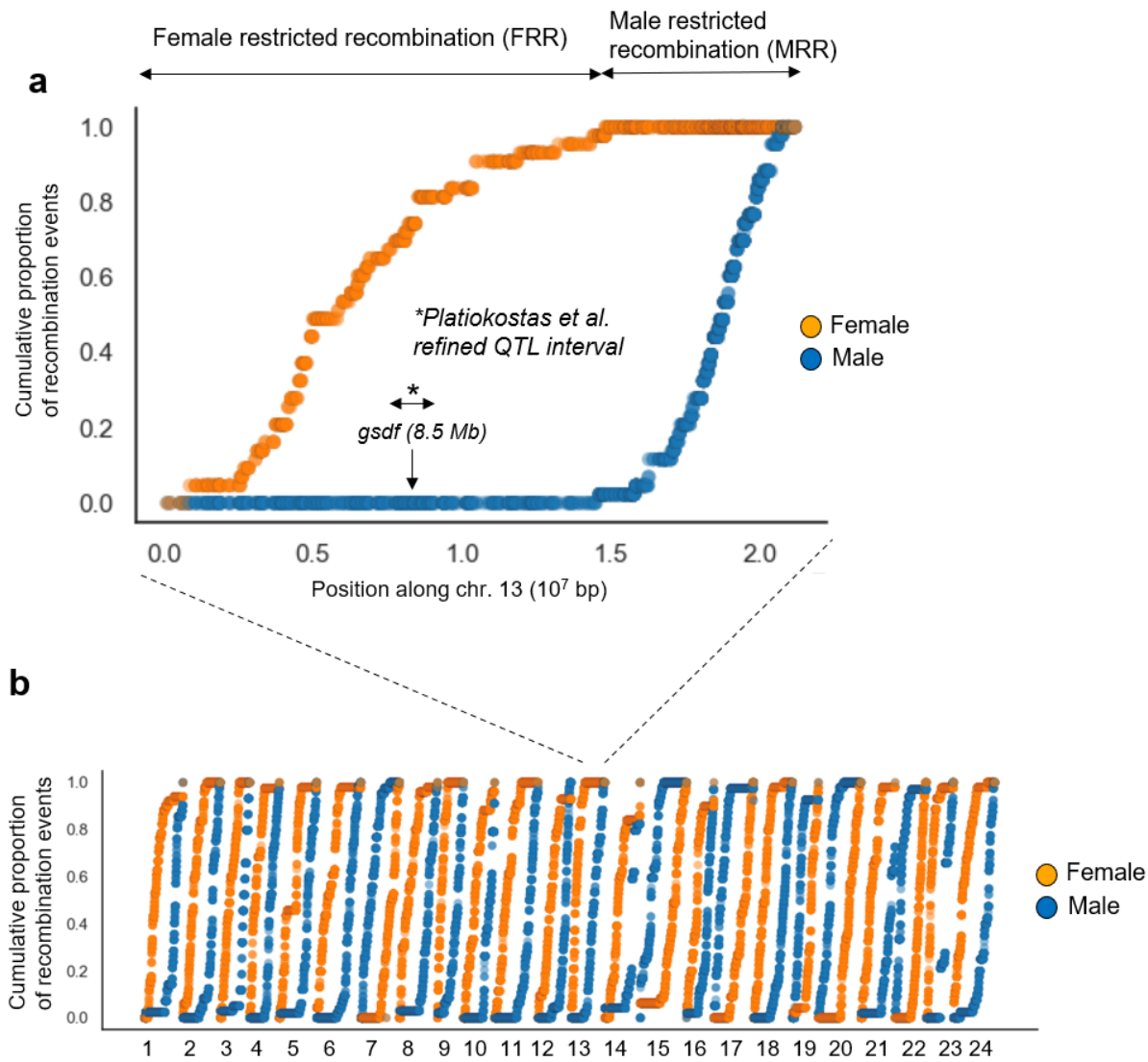


Fig. 3: Cumulative proportion of all recombination events occurring during male and female meiosis in a two-pedigree intercross. **a** chr13 with *gsdf* location indicated. *Mapping interval of markers most highly associated with sex in a previous study[21]. **b** Male and female linkage maps differ strongly. All 24 chromosomes exhibit clear sex restricted recombination intervals.

Nucleotide diversity and repeat content is correlated with sex-restricted recombination

We investigated what distinguishes an MRR from an FRR and concluded that for all chromosomes except chr9 MRRs are much smaller than FRRs while retaining approximately the same number of recombination events (Supplementary Fig. 7b). Furthermore, MRRs, again with the exception of chr9, exhibit higher nucleotide diversity (π) than their FRR counterparts

228 (Fig. 4a). Recombination rate is correlated with π (Supplementary Fig. 8) and chr13 is an outlier
229 with the largest MRR over FRR π ratio among all chromosomes, presumably due to the limited
230 effective population size of its FRR, the herein identified Y-chromosome. Typically, MRRs
231 occupy one end of a chromosome and FRRs start from the other end. Atlantic halibut has been
232 shown to have subtelomeric or acrocentric centromeres[19] and for 20 out of 21 chromosomes
233 where we could infer centromere placement by mapping centromeric markers[20] to
234 IMR_Hiphip.v1, the female end coincided with the centromere (Supplementary Table 3). We
235 next sought to determine whether meiotic recombination is sex restricted also in species closely
236 related to Atlantic halibut but could not find any publicly available pedigree datasets for other
237 Righteye flounders. Since Atlantic halibut π was found to be higher for MRRs than FRRs, we
238 decided to investigate whether π is correlated with MRR/FRR classification also in other
239 Righteye flounder species. We could not find any publicly available genome sequencing data
240 for Pacific halibut and instead used RNA-sequencing data to estimate π for this species. For
241 Greenland halibut (*Reinhardtius hippoglossoides*) and European plaice (*Pleuronectes platessa*)
242 we estimated π from publicly available RAD-seq- and whole genome shotgun sequencing data,
243 respectively. The across-species comparison of π distribution in MRRs and FRRs revealed a
244 striking agreement among the four species analyzed, with regions corresponding to Atlantic
245 halibut MRR having higher π (Fig. 4a-c), indicating that heterochiasmy has been conserved
246 among righteye flounders.

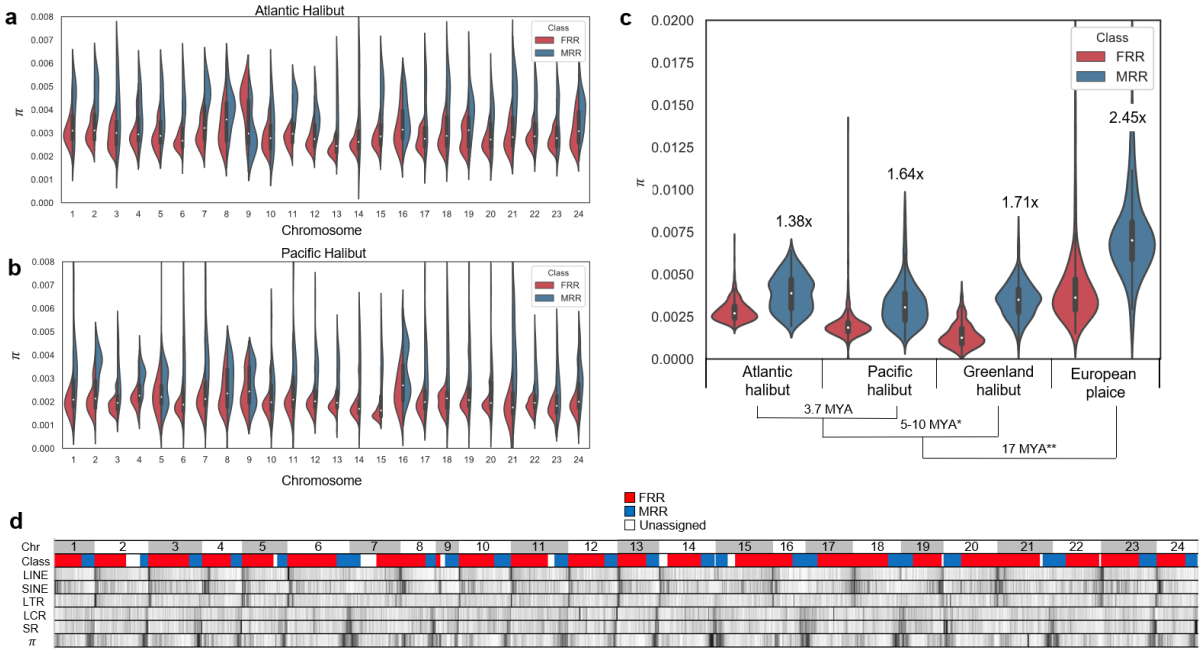


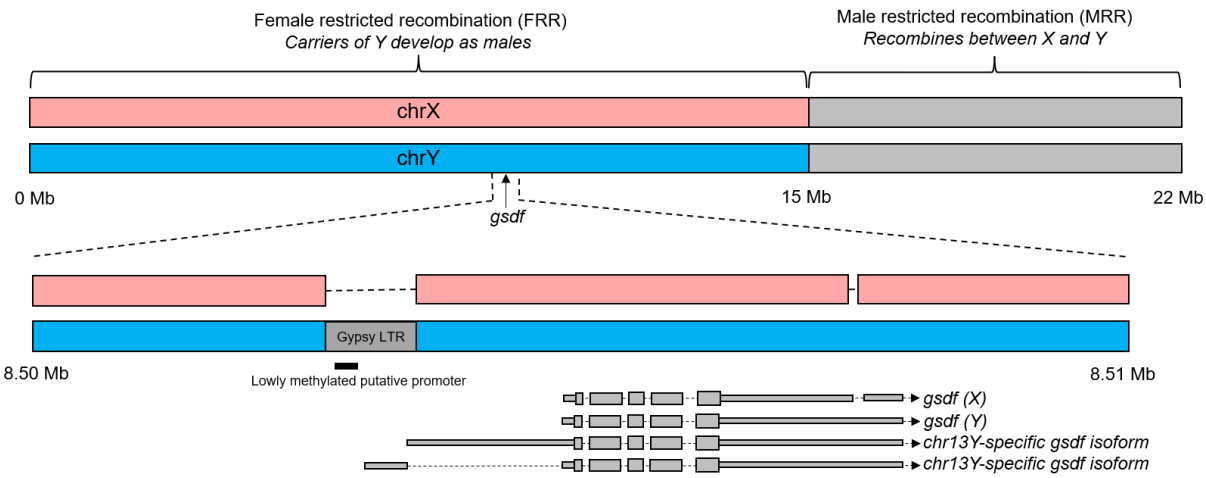
Fig. 4: Nucleotide diversity (π) and repeat content is correlated with sex-restricted recombination. **a** π distribution in the Atlantic halibut genome, contrasting regions of male- vs. female restricted recombination, MRRs and FRRs, respectively. **b** Distribution of Pacific Halibut π estimates for regions mapped back to the Atlantic halibut genome, classified as either MRRs or FRRs. Chromosomal and MRR/FRR assignments were extracted from the Atlantic halibut genome assembly coordinates. **c** Genome-wide distributions of estimated π for Atlantic-, Pacific- and Greenland halibut and European plaice in regions corresponding to herein defined Atlantic halibut MRRs and FRRs. Numbers above MRR distributions indicate, within species, how much higher π is in MRR regions compared to FRR regions. * The time since divergence of Greenland halibut from *Hippoglossus* was estimated from[24]. ** Divergence estimate for European plaice was obtained from TimeTree[25]. **d** Densities of repeat classes and π for 500 kb windows along each Atlantic halibut chromosome. Class color indicates whether a region was annotated as FRR, MRR or was unassigned. LINE and SINE represent Long- and short interspersed nucleotide elements, respectively. LTR indicates Long Terminal Repeats. LCR and SR indicate low complexity- and simple repeats, respectively. π indicates nucleotide diversity. The white to grey scale in heatmaps is proportional to the min. and max. values within each class.

We investigated whether MRRs and FRRs differed for repeat content and annotated repeats in IMR_Hiphip.v1 by means of RepeatMasker[26] using the *Actinopterygii* database. RepeatMasker raw output was parsed by repeat subclass and the density per subclass was investigated for each

266 chromosome (Supplementary Table 4). Overall, Long- and Short Interspersed Nucleotide
267 Element (LINE and SINE) densities increase towards the FRR end, while simple repeats and
268 low complexity repeats show the opposite pattern, being enriched towards the MRR end (Fig.
269 4d). Chr13 and chr10 (Supplementary Fig. 9 and 10) are both typical with respect to clustering
270 of major repeat classes into MRRs/FRRs. Chr10 is the only example where the MRR end
271 overlaps the predicted centromere, suggesting that the clustering and accumulation of repeats
272 is shaped by evolutionary forces associated with sex restricted recombination rather than
273 centromere location.
274
275 **The insertion on chromosome Y is of transposon origin and contains a *gsdf* promoter**
276 In an effort to identify candidate Structural Variants (SVs) between Y and X that may influence
277 *gsdf* expression, we aligned all the ONT reads from the male individual used for genome
278 assembly to the IMR_hipHip.v1 assembly and used the software Sniffles[27] to call SVs along
279 the genome. This analysis revealed a large set of putative SVs (Supplementary Table 5), two of
280 which coincided with the *gsdf* locus, both detected as heterozygous deletions in relation to the
281 reference at chr13:8503901-8505128 (1,227 bp) and chr13:8509688-8509767 (79 bp). Manual
282 inspection of these hits revealed that both ONT reads and linked read pseudo-haplotypes from
283 the male assembly individual validate the assignments made by Sniffles. We used alignments
284 of the male and female pool sequencing data and the phases of ONT long reads to conclude that
285 the insertion allele at these two indels is present on chrY. The HiC data used for scaffolding the
286 assembly was used to screen for large-scale SVs that may differ between X and Y. This analysis
287 indicated that no large inversions are present on chr13 and that neither Y nor X are part of any
288 large translocations with other chromosomes, as such events would appear as off-diagonal
289 contacts in the contact map (Supplementary Fig. 11).
290

291 The 1.2 kb insertion upstream of *gsdf* (INS1) is not found in the corresponding Pacific halibut
292 locus, supporting that this region is derived on the Y chromosome in Atlantic halibut. INS1
293 (IMR_Hiphip.v1 chr13:8503902-8505127) is highly similar to 24 other loci in the genome
294 assembly. Sequence alignments showed high sequence similarity starting with TG and ending
295 with CA di-nucleotides in 19/24 loci. The remaining 5 loci showed strong, further extending
296 similarities to each other, and we therefore extracted additional sequence flanking these five
297 loci to perform a 7-8 kb alignment, which also terminated in conserved TG/CA di-nucleotides
298 in four out of the five sequences. The TG/CA-ends are hallmarks of retrotransposon long
299 terminal repeat (LTR) boundaries, and we could also identify 4-6 nucleotide sequence
300 duplications immediately flanking the insertions, indicating target site duplications (TSDs) that
301 confirm the transposable element (TE) insertions as retrotransposons. Although these TE
302 hallmarks were obscured at the INS1 site upstream if *gsdf*, the sequence similarity and
303 alignment to the related loci in the genome support its TE origin (Supplementary File 1). One
304 of the five longer sequences (IMR_Hiphip.v1 chr14:3028070-3045206) was used as query to
305 search the DFAM database[28]. This search revealed a 5254 nt long hit against the 5410 nt long
306 *Danio rerio* Gypsy transposable element (Gypsy6-I_DR, Bit score 2342), supporting by
307 sequence similarity association that INS1 was derived from a Gypsy retrotransposon that
308 inserted in antisense orientation relative to *gsdf*. Strong similarities between the full-length
309 Gypsy LTR sequences and INS1 revealed that INS1 is indeed an LTR remnant of an ancient
310 Gypsy insertion. This association is further supported by RepeatMasker analysis of the five
311 putative full-length hits partially matching INS1, as they all contained Gypsy sequence
312 annotations.
313
314 From the ONT data used for genome assembly we could assess DNA methylation at the *gsdf*
315 locus. Interestingly, we observed low DNA methylation within INS1, in a region coinciding

316 with a predicted CpG-island (Fig. 5 and Supplementary Fig. 12). Further inspection of the lowly
317 methylated CpG-island revealed predicted Sp1 binding sites indicating that this locus may act
318 as a promoter region for *gsdf*. It is clear from both ONT read alignment and the linked read
319 pseudo-haplotype alignments that the chrY haplotype indeed carries INS1. We developed a
320 PCR assay to amplify into INS1 and observed that this assay gave clear bands of the expected
321 size for 10 adult phenotypic males while no bands were observed for 10 phenotypic females
322 (Supplementary Fig. 13). Expression of embryonal *gsdf* (28-54 dpf) from the Y-allele (*gsdf*-Y)
323 extends further upstream of the transcription start site than gene models both in adult Atlantic
324 halibut testis and in male- and female reproductive tissues in adult Pacific halibut. The longer
325 *gsdf-Y* isoforms contain 5'UTRs extending into the Gypsy LTR INS1 (Fig. 5 and
326 Supplementary Fig. 12), which supports promoter activity of INS1.
327



328

329 **Fig. 5:** Chr. Y carries a transposon-derived LTR upstream of *gsdf*. Schematic representation of findings. Male and
330 female chr13 haplotypes differ for a 1.2 kb transposable element derived segment 2 kb upstream of the *gsdf*
331 transcription start site. Males carry the insertion allele which contains a lowly methylated CpG-island inferred to
332 act as a derived promoter of *gsdf* uniquely in males. The shorter *gsdf* isoforms show the extent of *gsdf* expression
333 observed in adult testis in Atlantic halibut as well as adult ovary and testis of Pacific halibut. The two *gsdf* gene
334 models labeled chr13Y-specific indicate isoforms detected only in males from three separate developmental stages.
335 An IGV-plot of this region is presented in Supplementary Fig. 12.

336

337 **Analyzing chrY decay**

338 In addition to the ONT genome assembly, we also sequenced the same male using a Chromium
339 Genome linked read library and used the resulting data to generate two separate pseudo-
340 haplotype assemblies which successfully phased the first 10 Mb of chr13 into separate scaffolds
341 for the X and the Y chromosome. Loss of X/Y recombination is expected to lead to chrY decay,
342 and we investigated whether we could find evidence of this using the phased chrX (X^{PH}) and
343 chrY (Y^{PH}) along the first 10 Mb of chr13. This analysis, which involved quantification of
344 possibly deleterious variants along the first 10 Mb of X^{PH} and Y^{PH}, revealed a 30% excess of
345 variants predicted to be deleterious (HIGH classification in SNPEff) on Y^{PH} (n=35) compared
346 to X^{PH} (n=27). That no large-scale loss of genetic material has occurred on chrY is supported
347 by the observation that super-scaffolds of X^{PH} and Y^{PH} were both 20.4 Mb in size.

348

349 **Discussion**

350 In this study we show that *gsdf* is most likely the master sex determining gene. Our
351 interpretation is that following an ancient Gypsy element insertion on the autosomal progenitor
352 of chr13, this element degraded and left a remnant LTR, which now acts as a promoter driving
353 the expression of *gsdf* ectopically during development of XY individuals, making carriers
354 develop into phenotypic males. Interestingly, In Luzon ricefish, where *gsdf* is also the MSD
355 gene, the region immediately upstream of *gsdf* was shown to be most important for modulating
356 *gsdf* expression[22]. The region tested corresponds approximately to the site of TE insertion in
357 Atlantic halibut. A role of *gsdf* in early gonadal development is supported by experiments in
358 medaka where targeted homozygous deletion of *gsdf* in genetic (XY) males triggers an early
359 development of ovaries, with the majority of these *gsdf* KO homozygote fish still developing
360 testes later in life[29]. It would be interesting to use CRISPR/Cas9 to remove the Gypsy LTR

361 from XY Atlantic halibut (X/Y^{LTR-}) and insert the Gypsy LTR into XX individuals X/X^{LTR+}),
362 as we have done in Atlantic salmon[30,31]. Our predictions are that X/Y^{LTR-} individuals would
363 develop as females and X/X^{LTR+} individuals as males. Furthermore, in accordance with previous
364 studies[32], we observe *gsdf* expression in adult testes and ovaries in Pacific halibut and in
365 Atlantic halibut testis, indicating a function of Gsdf in mature gonads. In medaka ovaries and
366 testes, *gsdf* is expressed by somatic cells surrounding germ cells, with higher expression in
367 testis[29]. In Atlantic salmon *gsdf* expression has been detected in Granulosa and Sertoli in ovaries
368 and testes, respectively[33]. This supports a role for Gsdf in the mature gonad and suggests that
369 high expression of *gsdf* from Atlantic halibut chrY during development is expected to trigger
370 testis development. TEs are known to be involved in sex-chromosome evolution[34], e.g. TEs are
371 probably responsible for translocation of *sdY* in salmonides[35] and in Medaka a LINE insertion
372 at the *dmrt1b* locus has been implicated as the sex-determining variant[36]. It has been repeatedly
373 observed that TEs are repurposed for genomic innovation by domestication[37], as might be
374 expected for a dynamic part of the genome.
375
376 In this study we observed that meiotic recombination is dramatically different between male
377 and female gonads, and this can also be seen when revisiting a previous Atlantic halibut linkage
378 study[20], where this phenomenon was not strongly emphasized. Interestingly, the guppy
379 (*Poecilia reticulata*) has genome-wide heterochiasmy, while no significant difference was
380 observed in the closely related genus *Xiphophorus*[38], suggesting that the mechanism underlying
381 heterochiasmy may be easily modified over short evolutionary time-spans. A common feature
382 for observations of heterochiasmy in teleosts is enrichment of male crossovers towards
383 chromosome ends[38-40], which is also the case in Atlantic halibut. We found that four analyzed
384 righteye flounder species all had higher π in the regions orthologous to Atlantic halibut MRRs
385 than in corresponding FRRs orthologous regions (Fig. 4c). This likely reflects conservation of

386 sex restricted recombination across *Pleuronectidae*. The reason for this interpretation is that
387 few polymorphic sites are expected to be shared between species over the millions of years
388 separating the four analyzed species. Thus, the diverging π distributions must have been
389 established independently, presumably by selective forces associated with heterochiasmy.
390
391 In medaka, sex restricted recombination between sex chromosomes has been shown to follow
392 the XY male pattern also in XX neomales[41], which supports that a mechanism associated with
393 the gonadal environment governs recombination suppression in meiosis. We propose to study
394 the mechanisms of heterochiasmy in sex-reversed genetic female Atlantic halibut, which are
395 used as breeding males in aquaculture. Regions of TE-accumulation in genomes have been
396 implicated as putative heterochromatin sinks[42], and both TEs and heterochromatin are
397 positively correlated with suppressed recombination[43]. It is difficult to disentangle whether TEs
398 are involved in the establishment of meiotic recombination suppression or if their differing
399 distributions in MRRs/FRRs is merely a consequence of how the genome evolves due to
400 unknown selective pressures imposed by heterochiasmy already being in place. One possibility
401 could be that Atlantic halibut male and female gonads differ for which regions of the genome
402 become heterochromatin, perhaps as a defense against TE transposition, and that male and
403 female gonads direct recombination differently in response to variable TE content along the
404 chromosomes. Possibly, transposable elements are more likely to accumulate in FRRs, for
405 example if regions only recombining in males (MRRs) exist in a state less accessible for
406 transposition during female and/or male meiotic recombination. Accumulation of local TE
407 clusters in FRRs may also contribute to the further homogenization of this chromosome
408 compartment by gene conversion of homologous TE insertions. It should be noted that we have
409 analyzed repeats broadly, and future studies aiming to investigate repeat classifications in
410 relation to heterochiasmy may benefit from more detailed classifications. We show that π is

411 higher in MRRs than in FRRs and hypothesize that the higher recombination rate on the
412 physically smaller MRRs increase the probability of retaining diversity at linked sites during
413 purifying selection[44], or that differences in purifying selection due to the differing TE-
414 composition landscapes of FRR and MRRs has been involved in shaping π.
415
416 Sex-restricted meiotic recombination, as we describe here, suppresses recombination for large
417 segments along autosomes. Due to genome-wide, pre-existing sex-restricted recombination
418 being in place, this system is expected to facilitate the survival of novel sex determining factors
419 which in turn could give rise to new, and even competing genetic sex determination systems
420 (Supplementary Fig. 14). We propose that heterochiasmy has the potential to facilitate evolution
421 of sex determination systems, and presumably speciation as postulated in stickleback[45]. It has
422 been reported that female is the heterogametic sex (ZZ/ZW) in both Pacific halibut and
423 European plaice[12] as well as the half-smooth tongue sole[46], with the main candidate MSD gene
424 being *dmrt1*. This makes it conceivable, but not certain, that the Atlantic halibut X/Y system
425 replaced a Z/W system shared across different flatfish species.
426
427 Regardless of heterochiasmy being a facilitating factor in the evolution of sex determination
428 systems or not, we conclude that the TE insertion in the *gsdf* locus occurred after the split
429 between Atlantic halibut and Pacific halibut around 3.7 MYA. Sex chromosomes originate from
430 autosomes that acquire a sex determination locus and then lose recombination with the former
431 autosomal pair, which typically involves structural variation in the non-recombining region of
432 the sex chromosome pair, while other regions may still recombine and act as pseudo-autosomal
433 regions. For mammalian Y, as well as avian W, loss of purifying selection over millions of
434 years has led to pseudogenization, gradual loss of genes and a reduction in chromosome size[47],
435 where remaining genes and regulatory elements determine male and female phenotypes,

436 respectively. Our analysis of deleterious alleles identified a non-significant 30% increase in
437 predicted deleterious alleles on chrY compared to chrX, which may indicate an early stage of
438 chrY decay. We observe a 1 Mb segment (chr13:5.5-6.5 Mb) (Fig. 1c) showing differentiation
439 equal to the genome average, interpreted as X/Y recombination, although rare, not being fully
440 suppressed in the FRR of chr13. This is supported by previous fine mapping of the sex
441 determining locus to a smaller marker interval[21] (Fig. 3a) in a region corresponding to the herein
442 defined chrY. A retrotransposon insertion into the *gsdf* locus started the transformation of an
443 autosome into a sex chromosome in Atlantic halibut less than 4 million years ago. Incomplete
444 recombination barriers and little, if any, chrY decay both suggest that Atlantic halibut chrY is
445 in the early stages of the evolutionary trajectory of a sex-chromosome.
446
447
448 **Methods**
449
450 **Ethics statement**
451 For this study the halibut were reared under conditions similar to standard commercial fish
452 farming conditions. Such conditions are listed as an exception in The Norwegian Regulation on
453 Animal Experimentation, thus approval of the experimental protocol of this experiment by
454 NARA (the governmental Norwegian Animal Research Authority) was not required.
455
456 **Halibut material**
457 All samples were taken during routine production of Atlantic halibut at the Austevoll Research
458 Station, Institute of Marine Research, Storebø, Norway. Halibut breeding individuals were
459 stripped, eggs fertilized and incubated, and juvenile halibut produced according to standard
460 aquaculture protocols developed at the Austevoll Research Station[48-50]. Briefly, females were

461 stripped according to their individual ovulatory rhythms to ensure high (>90%) fertilization.
462 Fertilized eggs were incubated in 250 l conical incubators with upwelling water for 60 day
463 degrees (i.e. 10 days at 6ºC) and transferred to larger incubation units, silos (vol. 50,000 l)
464 before hatching. Yolk sac larvae were held in the silos until ca. 265 day degrees posthatch when
465 they were moved to first feeding tanks and fed *Artemia* for 70 days until metamorphosis was
466 complete. At this stage, the larvae were weaned onto a formulated diet. Juvenile halibut were
467 moved to ongrowing tanks when they had reached an average size of ca. 10 cm and held there
468 until they had reached a harvest size of 2-5 kg. One adult, three-year old male was euthanized
469 with metacain (100 mg/l) and muscle, spleen and blood samples were taken for reference
470 genome sequencing. For sequencing of pooled family samples, tissue samples were taken from
471 ten full siblings of mixed sex (ten males and ten females). All samples were snap frozen in
472 liquid nitrogen (N_2) and stored at -80ºC before DNA extraction. RNA-sequencing was made on
473 a series of samples taken through embryonic development. Before hatching, embryo samples
474 were snap frozen in liquid N_2 and stored at -80ºC. Larvae were euthanized in metacain (100
475 mg/l), transferred into RNA later, kept at 4ºC for 24 h and stored at -20ºC until RNA isolation.
476 The samples for qPCR were snap frozen in liquid N_2 and stored at -80ºC. Digital images of all
477 embryo-larval stages were captured using video cameras Moticam 1080 (Motic) mounted on
478 Olympus SZX10 stereomicroscope using Motic Live Imaging Module. Living embryos and
479 larvae were individually immobilized in a Petri dish containing 2% methylcellylose/98%
480 seawater. Images were then scaled and processed by ImageJ software[51].
481
482 **DNA extraction and sequencing of male and female pools**
483 DNA was extracted from muscle samples using Qiagen DNeasy Blood and Tissue Kit. Paired-
484 end libraries of equimolar male (10 individuals) and female (10 individuals) DNA pools and
485 from one single individual of each sex were constructed using the Genomic DNA Sample

486 Preparation Kit (Illumina) according to manufacturer's instructions and sequenced on the
487 Illumina HiSeq2000 platform (Illumina) at the Norwegian Sequencing center
488 (https://www.sequencing.uio.no, Oslo, Norway).
489
490 **High Molecular Weight DNA extraction and sequencing**
491 Several tissues were collected from a single male Atlantic Halibut. Muscle and spleen were
492 snap frozen in liquid nitrogen and stored at -80ºC until further use. Fresh blood was used to
493 isolate erythrocyte nuclei and then prepare High Molecular Weight (HMW) DNA from these
494 nuclei according to a Phenol/Chloroform protocol (Quick). Extracted DNA was either left
495 unfragmented or sheared using either MegaRuptor (Diagenode) or by passing DNA through a
496 100µl pipette tip 10-20 times. From these DNA samples we produced MinION sequencing
497 libraries with the standard adapter ligation protocol kit (LSK-108), 1D^2 sequencing kit (LSK-
498 308) and the RAD002 and RAD003 kits (Oxford Nanopore Technologies). For libraries
499 prepared with the RAD002 and RAD003 kits we used a protocol developed to produce ultra–
500 long reads (Quick, protocols.io). In total we used 9 flow cells and sequenced 18.8 Gbp in
501 2,963,108 reads with a read N50 of 17.8 kb. Sequenced fast5 raw signal data files were
502 basecalled using albacore v.2. The size distribution of sequenced reads is presented in
503 **Supplementary Figure 1a.** We also used the collected muscle samples to produce HiC libraries
504 (Dovetail), sequenced on one lane of HiSeq X (Illumina). One of the DNA preps inferred to
505 have HMW DNA was used to prepare a linked reads Chromium library (10x genomics) and
506 was sequenced on one lane of HiSeqX (Illumina).
507
508 **Genome assembly, polishing and scaffolding**
509 Nanopore reads passing filters were used in assembly by first constructing all-vs-all alignments
510 between all reads using minimap2[52]. These all-vs-all alignments were then used in the software

511 miniasm to produce a haploid contig assembly. The contig assembly was polished using racon[53]
512 with the nanopore data followed by pilon[54] using the Illumina data from the male pool.
513 Scaffolding of the polished miniasm assembly was performed using the HiRise pipeline
514 (Dovetail Genomics). The quality of the generated assembly was assessed using BUSCO[55].
515 Initial nanopore contigs assembled with miniasm[52] had a contig N50 of 3.2Mb and the resulting
516 final assembly had a scaffold N50 of 27.2 Mb.
517
518
519
520 **Assembly of linked read data**
521 The software Supernova2[56] (10X Genomics) was used to produce a genome assembly of the
522 sequencing data resulting from the Illumina sequencing of the Chromium library. For this
523 assembly, reads were subsampled to get close to the recommended 56x genome coverage
524 recommended in the Supernova2 user manual. The two pseudo-haplotype linked read
525 assemblies featured one scaffold each which covered the whole of IMR_Hiphip.v1 chr13. We
526 used SNP markers identified as divergent between the male and female short read data to
527 investigate the pseudo-haplotypes in relation to sex divergent markers, i.e. if one pseudo-
528 haplotype could represent the Y chromosome and the other the X-chromosome. This analysis
529 revealed that for the first 10 Mb of chr13, the allele being heterozygous in males and not present
530 in females (the perceived Y-allele) was observed on pseudo-haplotype number 2 (1,954 SNPs).
531 The perceived Y-allele was observed only for 2 SNPs in this region for pseudo-haplotype
532 number 1. The corresponding counts for the rest of chr13 was 125 and 213 alleles on pseudo-
533 haplotype numbers 2 and 1, respectively.
534
535 **RNA-sequencing and gene annotation**

536 Total RNA was extracted from whole larvae from seven different stages of development using
537 miRNeasy Mini Kit (Qiagen) and treated with TURBO DNase-Free kit (Ambion). TruSeq
538 cDNA libraries were sequenced on the HiSeq 4000 sequencing platform (Illumina). The
539 average number of reads was 50,320,639 per sample. Twenty-two samples (four samples from
540 54 dpf and three samples from each other stage) were assembled individually with Trinity using
541 the Agalma pipeline[57]. The transcriptome assemblies were annotated with BLAST against
542 Swissprot (download date 20191227) and the Ensembl gene models of zebrafish (GRCz11) and
543 stickleback (BROAD S1). The hits with the best scores were kept. Transcripts from all
544 assemblies were mapped to the halibut genome with BLAT. The thresholds for mapping were
545 set to 98% match for sequences with length >=401 nt and 95% match for sequences with length
546 <=400 nt. Long sequences were mapped before short sequences. The mapped sequences were
547 clustered on genome location, selecting the sequence with the highest score for each non-
548 overlapping cluster. Only clusters with an annotation (>300 in BLAST score threshold) were
549 saved. Neighboring clusters annotated with the same gene from either Swissprot, zebrafish or
550 stickleback were merged in cases where the annotation indicated that the gene was split in
551 several clusters (the neighboring clusters were only merged in cases where the annotation was
552 from different parts of the same gene). For the merged clusters, the sequence with the highest
553 BLAT score from all merged clusters were chosen. Additionally, the transcriptome assembly
554 sequence with the highest scoring BLAT match to the merged cluster was added. We found
555 15,273 annotated genes based on the Trinity assemblies. The zebrafish and stickleback
556 ENSEMBL gene models were mapped to the halibut genome using exonerate in order to add
557 genes without RNA-seq evidence. The exonerate hits were clustered by location, using the
558 exonerate hit with the highest score for each cluster. Sequences for each exonerate based gene
559 were extracted from the halibut genome sequence. 2,269 genes were added based on the
560 exonerate analysis. We added 222 genes that had no annotation (BLAST threshold score <300)

561 but had RNA-seq expression. After running BUSCO 3.1 with the Actinopterygii odb9 dataset
562 on the non-filtered Trinity assemblies that mapped to the genome[55], we added 550 genes from
563 the BUSCO analysis that were missing from our filtered list. This resulted in a total number of
564 18,314 genes and 20,579 transcripts giving a 80.9% complete, 8.7% fragmented and 10.4%
565 missing BUSCO coverage.
566
567 **Nanopore sequencing of testis RNA**
568 Testis tissue was homogenized at 4°C in 1ml QIAzol Lysis Reagent (Qiagen, Hilden, Germany)
569 together with RNase-Free Zirconium Oxide Beads (NextAdvance, Inc., Troy, NY, USA) during
570 1 min, maximum effect, in a BulletBlender (NextAdvance) and total RNA was prepared
571 according to the QIAzol (Qiagen) protocol. Polyadenylated RNA was isolated with the
572 Dynabeads mRNA direct purification kit (Thermo Fisher Scientific), following the kit protocol.
573 For full-length cDNA Oxford Nanopore sequencing, the cDNA-PCR Sequencing SQK-
574 PCS108 kit (Oxford Nanopore Technologies) was used according to the kit protocol with minor
575 modifications. The finished cDNA library was sequenced on a MinION device using a R9.4
576 flow cell (Oxford Nanopore Technologies).
577
578 **Screening for male vs. female divergence**
579 Illumina sequences from the male- and female DNA pools were filtered by trimmomatic[58], then
580 aligned to the assembly IMR_hipHip.v1 using BWA-MEM[59]**.** Aligned reads were subjected to
581 duplicate flagging ("Picard Toolkit." 2019. Broad Institute, GitHub Repository.
582 http://broadinstitute.github.io/picard/; Broad Institute) and were then used as input for calling
583 SNPs as well as small Insertions/Deletions using Genome Analysis Toolkit (GATK)[60]
584 UnifiedGenotyper. Raw variant vcf files were subjected to variant filtration using Best Practices
585 procedure with SNP-clusters being added as an additional filter (more than 5 SNPs in a window

586 of 20 bases were filtered out). SNPs passing filters were isolated from the vcf file and allele
587 frequencies from the male- and female pools were extracted and used in χ^2-analysis examining
588 the null hypothesis; that allele frequencies were not different between the male- and female
589 pools. Prior to the χ^2-test, in order not to inflate the significance at positions with high depths
590 of coverage, SNP read depths >12 for either pool were set to 12 while retaining the observed
591 allele frequency ratio. χ^2-tests were then performed SNP-by-SNP, and average χ^2-scores of
592 non-overlapping windows of 100 adjacent SNPs along the chromosomes were extracted.
593
594 **Sex differentiating PCR assay**
595 Genetic sex of each embryo was determined by PCR using sex-specific primers (Table 1). The
596 thermal cycling conditions were as follows: 94°C for 5 min, 32 cycles with 94°C for 30 sec,
597 65°C for 1 min and 72°C for 1 min, and finally 72°C for 5 min. Each 12 µl PCR reaction
598 contained 5X GOTaq Flexi Buffer, MgCl, dNTPs, GoTaq Polymerase, PCR primers and
599 template cDNA.
600
601 **Differential gene expression between sexes**
602 We generated RNA-sequencing data from 12 Atlantic halibut whole embryos representing
603 stages 28 (n=4), 43 (n=4) and 54 (n=4) dpf. Since one cannot determine the phenotypic sex at
604 this stage we aligned the RNA-seq data using Bowtie 2[61] to IMR_hipHip.v1 and extracted
605 expressed allele counts at positions identified as divergent between males and females in the
606 pool DNA-seq. Extracted RNA-seq allele ratios were used to cluster the individuals, revealing
607 two well-separated clusters of XY and XX individuals and this assignment was used to identify
608 genetic male and female individuals for the differential expression screen. The differential
609 expression screen involved contrasting across-sample normalized gene expression levels of the
610 seven inferred males to the five inferred females in a two-tailed t-test. Principal Component

611 Analysis (PCA) was performed, using the R function "prcomp", on the normalized gene
612 expression levels from the 12 embryos.
613
614 ***gsdf* quantitative PCR**
615 Genomic DNA and total RNA was extracted from single embryos using the Allprep
616 DNA/RNA/miRNA Universal Kit (Qiagen), according to the manufacturer's instructions. The
617 RNA was DNase treated as part of the extraction process, according to the manufacturer's
618 instructions. For each sample, cDNA was synthesized from 125 ng RNA using the Superscript
619 VILO cDNA synthesis Kit (ThermoFisher Scientific). The primers used for qPCR were
620 designed using BatchPrimer3 (https://wheat.pw.usda.gov/demos/BatchPrimer3/) (Table 2). The
621 gene *gtf3c6* was used as an internal reference gene for normalization of the qPCR data due to
622 its stable expression in the developmental stages analysed in this study (Supplementary Fig.
623 15). Using different RNA concentrations, the PCR efficiencies (E = -1+10(-1/slope)) for the
624 primers were calculated to be 108.5% (*gsdf*) and 92.4% (*gtf3c6*). A melt curve analysis was
625 performed to confirm that the primers did not amplify more than one PCR product. qPCR was
626 performed in duplicates in 384-well optical plates in a QuantStudio 5 Real-Time PCR system
627 (ThermoFisher Scientific) using default settings. 2 µl cDNA (diluted 1:20) was used in a 6 µl
628 Fast SYBR Green qPCR reaction (ThermoFisher Scientific). No-template controls for each
629 gene were run in all qPCR plates. The relative gene expression level was calculated using the
630 comparative Ct (or $2^{-\Delta\Delta Ct}$) method. All values were normalized to *gtf3c6* and calibrated to the
631 sample with the lowest ΔCt.

632 **Detection of structural variants, small insertion/deletions and annotation of repeat**
633 **sequences**
634 The long-read SV identification tool Sniffles[27] was used to identify SVs from the ONT data.
635 The ONT data from the single male individual used for genome assembly was aligned to

636 IMR_hipHip.v1 using minimap2[52] with the "-x map-ont" option and the resulting bam file was
637 used in sniffles, requiring >=5 observations of a variant allele to be retained as a potential SV.
638
639 **Analysis of chrY decay**
640 The two linked-read pseudo-haplotype assemblies of the male Atlantic halibut used for the
641 genome assembly were aligned against the Pacific halibut reference genome (nucmer) and
642 SNPs and small indels were called using the "show-snps" function of the MUMmer 4.0
643 package. Identified variants were converted to vcf-format
644 (https://github.com/MatteoSchiavinato/Utilities/blob/master/my-mummer-2-vcf.py) and were
645 then annotated using SNPEff, with gene models derived from the Pacific halibut
646 IPHC_HiSten_1.0 gtf-file "GCF_013339905.1_IPHC_HiSten_1.0_genomic.gtf". SNPs and
647 indels that were unique to either pseudo-haplotype were investigated for their annotations.
648 Annotation frequencies were compared between the two pseudohaplotypes for the interval
649 where they represent the X and Y (Atlantic halibut: chr13:0-10Mb, Pacific halibut:
650 NC_048944.1:0-11Mb).
651
652 **Analysis of the evolutionary origin of the 1.2 kb insertion upstream of *gsdf***
653 The 1.2 kb insertion (IMR_Hiphip.v1 chr13:8503902-8505127) was used as a query sequence
654 in BLAT[62] searches against IMR_HipHip.v1. Full length hits were retained (n=30) and were
655 extended to include 2kb up- and downstream sequences and these approximately 5.3 kb
656 sequences were used as input for multiple sequence alignment using MUSCLE[63]. The resulting
657 alignments were visualized in AliView[64] and abrupt drops in sequence similarities were
658 observed for 24 of the sequences on each flank. The extent of highly similar sequences started
659 and ended at perfectly conserved TG and CA di-nucleotides, the hallmark motifs of LTR

660 boundaries, and the insertion was immediately flanked by 4-6 nucleotide duplications of loci
661 TE target sites (TSD) confirming the LTR-retrotransposon insertion.
662
663 **DNA methylation calling from ONT data and CpG island prediction**
664 Basecalled ONT reads previously used for the genome assembly were aligned back to the
665 IMR_Hiphip.v1 assembly using minimap2. The software nanopolish[65] was used to index the
666 raw ONT fast5-files to enable extraction of likelihoods of CpG site methylation along
667 individual mapped reads in genome assembly coordinates. Next, the nanopolish script
668 "calculate_methylation_frequency.py" was used to extract the methylation frequency at each
669 adequately covered CpG cluster in the assembly. CpG islands were predicted for the
670 IMR_Hiphip.v1 assembly using gcluster[66] employing default settings.
671
672 **Estimation of nucleotide diversity**
673 For Atlantic Halibut we estimated π by creating chromosome-specific samtools mpileup files
674 from the aligned data previously used for variant calling, using only positions with a mapping
675 quality >=20 and read depths ranging between 10 and 50. The script "WinNuclDivmPileup.py"
676 (https://github.com/cjrubinlab/python_scripts2) was used to estimate π in 500kb windows
677 along the genome for the male- and female pools. For Pacific halibut no DNA-sequencing data
678 was available in public sequence repositories, so we devised a way to estimate π from RNA-
679 sequencing data. Sequencing data from 12 RNA-seq experiments of various tissues and
680 individuals was downloaded from the SRA (SRR11826178-SRR11826189) and these data were
681 aligned to the Oxford nanopore derived Pacific Halibut genome assembly (GCF_013339905.1)
682 using the STAR aligner[67]. The resulting alignments were randomly subsampled 20 times and
683 then subjected to duplicate-pair removal using picard-tools MarkDuplicates, retaining 70% of
684 the reads after filtering. From the bam file containing filtered reads an mpileup file was

685 generated using samtools (mapping quality exclusion filter < 20). For the Greenland Halibut
686 (*Reinhandtius Hippoglossoides*) we downloaded the genome assembly from NCBI
687 (GCA_006182925.2_MU_Rhippo_v1) as well as RAD-sequencing data from 20 individuals
688 (SRA accessions: SRR11351808-SRR11351827), aligned the short-read data to the assembly
689 using minimap2 with the "-x sr" option. From mpileup files derived from samtools mpileup for
690 Pacific Halibut and Greenland halibut the estimated heterozygosity (He) was calculated for
691 positions having coverage>=10 and mapping quality >=20 using the WinNuclDivmPileup.py"
692 script. For European plaice (*Pleuronectes platessa*), a species lacking a genome assembly, we
693 downloaded paired genome resequencing Illumina reads (SRA accession: SRX8240416),
694 aligned these reads to the IMR_Hiphip.v1 assembly using bwa and proceeded with the same
695 pipeline as was used for estimation of Atlantic halibut π. In order to lift over genomic intervals
696 annotated as MRRs and FRRs in the Atlantic halibut to the Pacific- and Greenland halibut
697 assemblies we used the satsuma[68] commands "Chromosemble" to generate a detailed list of lift-
698 over intervals and "BlockDisplaySatsuma" to summarize the detailed list into coherent blocks.
699 We then converted these coordinates to the bed-format and used the bedtools[69] "intersectbed"
700 command to link coordinates in the other assemblies to the corresponding coordinates in the
701 Atlantic halibut assembly and MRRs and FRR classification.
702
703 **Linkage analysis**
704 RAD-sequencing data from an Atlantic halibut half-sib pedigree altogether comprising three
705 parents and 90 offspring were downloaded from the SRA (accession=SRX193134). The
706 downloaded individual files were split by barcodes and were then connected with annotation
707 information from the original publication regarding phenotypic sex, parent/offspring status and
708 family B/C[21]. The paired-end data from each individual in the pedigree was then aligned to
709 IMR_hipHip.v1 using BWA-MEM. Resulting bam-files were subjected to duplicate removal

710 using picard-tools “MarkDuplicates”, then raw variant (SNP and indel) calling using GATK
711 UnifiedGenotyper. Raw SNPs were filtered using GATK “VariantFiltration” with the following
712 arguments, also with a list of raw indel positions given to the “InDel” mask; “--maskName
713 InDel”, “--maskExtension=5”, “--clusterSize=5”, “--clusterWindowSize=15”,
714 "BaseQRankSum<-7.0", "BaseQRankSum > 7.0", "MQRankSum < -7.0", "MQRankSum >
715 7.0", "QD < 5.0","FS > 50.0","ReadPosRankSum > 7.0", "ReadPosRankSum < -7.0" and "DP
716 > 40". SNPs remaining after filtration were saved in a new vcf file which was given as input to
717 Lep-MAP3[70] to generate recombination maps for the two families. The commands used in Lep-
718 MAP were: “ParentCall2”, with the options (removeNonInformative=1 XLimit=2 halfSibs=1),
719 then “Filtering2”, with the option data=- (dataTolerance=0.01), then “SeparateChromosomes2”
720 with the options (lodLimit=10 distortionLod=1) to identify linkage groups in the pedigrees.
721 SeparateChromosomes2 with lodLimit=10 generated a map of 31 linkage groups, out of which
722 20 had 964–3063 markers assigned to them (24 were expected based on karyotype). We
723 continued with the output from SeparateChromosomes2 (lodLimit=10) to run “OrderMarkers2”
724 to, for each identified linkage group, order the markers without having given the software access
725 to genome assembly coordinates. The output files from the first iteration of “OrderMarkers2”
726 were then subjected to a second round of “OrderMarkers2”, this time giving the Lep-MAP3
727 access to genome coordinates of markers on each linkage group using the option
728 “evaluateOrder”. The resulting linkage group specific files containing male- and female
729 centiMorgan (cM) positions of each marker were filtered to prune obviously inflated cM
730 positions in the very beginning and the very end of assembled chromosomes and by splitting
731 four linkage groups where two assembled chromosomes had been merged into one linkage
732 group, thereby increasing the observed linkage group count from 20 to the expected 24.
733 Resulting maps from this stage were retained for downstream analyses.
734

735 **Pacific halibut RNA-seq contrasts**

736 Sequencing data from 12 RNA-seq experiments of various tissues and individuals
737 (SRR11826178-SRR11826189) was downloaded from SRA. Four of these tissues (testis,
738 ovary, white muscle and brain) were used for characterization of *gsdf* gene expression levels
739 and validation of gene models from the gff-file accompanying the Pacific halibut assembly on
740 NCBI. Alignments were performed using STAR as described above for π estimation.

741

742 **RepeatMasker analysis**

743 RepeatMasker v4.1.0[26] was used to annotate both simple and interspersed repeats. We deployed
744 the RMBLASTN (v2.6.0.+) search engine to scan for known repeats among ray-finned fishes
745 ("-species actinopterygii") using a database that combined Dfam_3.1 and RepBase-20181026
746 repeat sequences. A custom Perl script was used to parse the main output table from
747 RepeatMasker and compute the densities of major repeat classes (i.e. "Simple_repeat", "LINE",
748 "LTR", "Low_complexity", "DNA", "RC", "SINE", "Satellite", "Unknown", "Retroposon")
749 in windows of 50 kbp and 500 kbp, respectively.

750

751 **Supplementary data**

752 Supplementary File 1: Integration site analysis of the 1.2 kb insertion upstream of *gsdf*, which
753 controls male development, and related insertions in the Atlantic halibut genome.

754

755 **References**

756

757 1 Jakupsstovu, S. H. & Haug, T. Growth, Sexual-Maturation, and Spawning Season of
758 Atlantic Halibut, Hippoglossus-Hippoglossus, in Faroese Waters. *Fish Res* **6**, 201-215
759 (1988).

760 2 Tvedt, H. B., Benfey, T. J., Martin-Robichaud, D. J., McGowan, C. & Reith, M.
761 Gynogenesis and sex determination in Atlantic halibut (Hippoglossus hippoglossus).
762 *Aquaculture* **252**, 573-583, doi:10.1016/j.aquaculture.2005.06.042 (2006).
763 3 Hendry, C. I., Martin-Robichaud, D. J. & Benfey, T. J. Hormonal sex reversal of
764 Atlantic halibut (Hippoglossus hippoglossus L.). *Aquaculture* **219**, 769-781, doi:Pii
765 S0044-8486(02)00344-7
766 Doi 10.1016/S0044-8486(02)00344-7 (2003).
767 4 Babiak, J. *et al.* Induced sex reversal using an aromatase inhibitor, Fadrozole, in Atlantic
768 halibut (Hippoglossus hippoglossus L.). *Aquaculture* **324**, 276-280,
769 doi:10.1016/j.aquaculture.2011.11.001 (2012).
770 5 Hendry, C. I., Martin-Robichaud, D. J. & Benfey, T. J. Gonadal sex differentiation in
771 Atlantic halibut. *J Fish Biol* **60**, 1431-1442, doi:10.1006/jfbi.2002.1950 (2002).
772 6 Kikuchi, K. & Hamaguchi, S. Novel sex-determining genes in fish and sex chromosome
773 evolution. *Dev Dynam* **242**, 339-353, doi:10.1002/dvdy.23927 (2013).
774 7 Heule, C., Salzburger, W. & Bohne, A. Genetics of sexual development: an evolutionary
775 playground for fish. *Genetics* **196**, 579-591, doi:10.1534/genetics.114.161158 (2014).
776 8 Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D. & Peichel, C. L. Turnover of sex
777 chromosomes in the stickleback fishes (gasterosteidae). *Plos Genet* **5**, e1000391,
778 doi:10.1371/journal.pgen.1000391 (2009).
779 9 Cnaani, A. *et al.* Genetics of sex determination in tilapiine species. *Sexual development*
780 *: genetics, molecular biology, evolution, endocrinology, embryology, and pathology of*
781 *sex determination and differentiation* **2**, 43-54, doi:10.1159/000117718 (2008).
782 10 Franchini, P. *et al.* Long-term experimental hybridisation results in the evolution of a
783 new sex chromosome in swordtail fish. *Nat Commun* **9**, 5136, doi:10.1038/s41467-018-
784 07648-2 (2018).

785 11 Kottler, V. A. *et al.* Independent Origin of XY and ZW Sex Determination Mechanisms
786 in Mosquitofish Sister Species. *Genetics* **214**, 193-209,
787 doi:10.1534/genetics.119.302698 (2020).
788 12 Luckenbach, J. A., Borski, R. J., Daniels, H. V. & Godwin, J. Sex determination in
789 flatfishes: Mechanisms and environmental influences. *Semin Cell Dev Biol* **20**, 256-263,
790 doi:10.1016/j.semcdb.2008.12.002 (2009).
791 13 Drinan, D. P., Loher, T. & Hauser, L. Identification of Genomic Regions Associated
792 With Sex in Pacific Halibut. *J Hered* **109**, 326-332, doi:10.1093/jhered/esx102 (2018).
793 14 Herpin, A. & Schartl, M. Plasticity of gene-regulatory networks controlling sex
794 determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO*
795 *reports* **16**, 1260-1274, doi:10.15252/embr.201540667 (2015).
796 15 Yano, A. *et al.* An immune-related gene evolved into the master sex-determining gene
797 in rainbow trout, Oncorhynchus mykiss. *Current biology : CB* **22**, 1423-1428,
798 doi:10.1016/j.cub.2012.05.045 (2012).
799 16 Bao, L. *et al.* The Y chromosome sequence of the channel catfish suggests novel sex
800 determination mechanisms in teleost fish. *BMC biology* **17**, 6, doi:10.1186/s12915-019-
801 0627-7 (2019).
802 17 Rafati, N. *et al.* Reconstruction of the birth of a male sex chromosome present in
803 Atlantic herring. *Proc Natl Acad Sci U S A* **117**, 24359-24368,
804 doi:10.1073/pnas.2009925117 (2020).
805 18 Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate
806 species. *bioRxiv*, doi:10.1101/2020.05.22.110833 (2020).
807 19 Brown, N. P., Bromage, N. R., Penman, D. J. & Shields, R. J. The karyotype of the
808 Atlantic halibut, Hippoglossus hippoglossus (Linnaeus). *Aquac Res* **28**, 489-491 (1997).

809 20 Reid, D. P. *et al.* A genetic linkage map of Atlantic halibut (Hippoglossus hippoglossus
810 L.). *Genetics* **177**, 1193-1205, doi:10.1534/genetics.107.075374 (2007).
811 21 Palaiokostas, C. *et al.* Mapping the sex determination locus in the Atlantic halibut
812 (Hippoglossus hippoglossus) using RAD sequencing. *BMC genomics* **14**, doi:Artn 566
813 10.1186/1471-2164-14-566 (2013).
814 22 Myosho, T. *et al.* Tracing the emergence of a novel sex-determining gene in medaka,
815 Oryzias luzonensis. *Genetics* **191**, 163-170, doi:10.1534/genetics.111.137497 (2012).
816 23 Rondeau, E. B. *et al.* Genomics of sablefish (Anoplopoma fimbria): expressed genes,
817 mitochondrial phylogeny, linkage map and identification of a putative sex gene. *BMC*
818 *genomics* **14**, 452, doi:10.1186/1471-2164-14-452 (2013).
819 24 Vinnikov, K. A., Thomson, R. C. & Munroe, T. A. Revised classification of the righteye
820 flounders (Teleostei: Pleuronectidae) based on multilocus phylogeny with complete
821 taxon sampling. *Mol Phylogenet Evol* **125**, 147-162, doi:10.1016/j.ympev.2018.03.014
822 (2018).
823 25 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for
824 Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **34**, 1812-1819,
825 doi:10.1093/molbev/msx116 (2017).
826 26 Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in
827 genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4 10,
828 doi:10.1002/0471250953.bi0410s25 (2009).
829 27 Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single-
830 molecule sequencing. *Nat Methods* **15**, 461-+, doi:10.1038/s41592-018-0001-7 (2018).
831 28 Hubley, R. *et al.* The Dfam database of repetitive DNA families. *Nucleic Acids Res* **44**,
832 D81-D89, doi:10.1093/nar/gkv1272 (2016).

833 29 Imai, T., Saino, K. & Matsuda, M. Mutation of Gonadal soma-derived factor induces
834 medaka XY gonads to undergo ovarian development. *Biochem Bioph Res Co* **467**, 109-
835 114, doi:10.1016/j.bbrc.2015.09.112 (2015).
836 30 Edvardsen, R. B., Leininger, S., Kleppe, L., Skaftnesmo, K. O. & Wargelius, A.
837 Targeted mutagenesis in Atlantic salmon (Salmo salar L.) using the CRISPR/Cas9
838 system induces complete knockout individuals in the F0 generation. *PloS one* **9**,
839 e108622, doi:10.1371/journal.pone.0108622 (2014).
840 31 Straume, A. H. *et al.* Indel locations are determined by template polarity in highly
841 efficient in vivo CRISPR/Cas9-mediated HDR in Atlantic salmon. *Sci Rep* **10**, 409,
842 doi:10.1038/s41598-019-57295-w (2020).
843 32 Sawatari, E., Shikina, S., Takeuchi, T. & Yoshizaki, G. A novel transforming growth
844 factor-beta superfamily member expressed in gonadal somatic cells enhances primordial
845 germ cell and spermatogonial proliferation in rainbow trout (Oncorhynchus mykiss).
846 *Dev Biol* **301**, 266-275, doi:10.1016/j.ydbio.2006.10.001 (2007).
847 33 Kleppe, L. *et al.* Transcriptomic analysis of dead end knockout testis reveals germ cell
848 and gonadal somatic factors in Atlantic salmon. *BMC genomics* **21**, 99,
849 doi:10.1186/s12864-020-6513-4 (2020).
850 34 Dechaud, C., Volff, J. N., Schartl, M. & Naville, M. Sex and the TEs: transposable
851 elements in sexual development and function in animals. *Mob DNA* **10**, 42,
852 doi:10.1186/s13100-019-0185-0 (2019).
853 35 Faber-Hammond, J. J., Phillips, R. B. & Brown, K. H. Comparative Analysis of the
854 Shared Sex-Determination Region (SDR) among Salmonid Fishes. *Genome Biol Evol*
855 **7**, 1972-1987, doi:10.1093/gbe/evv123 (2015).

856 36 Herpin, A. *et al.* Transcriptional rewiring of the sex determining dmrt1 gene duplicate
857 by transposable elements. *Plos Genet* **6**, e1000844, doi:10.1371/journal.pgen.1000844
858 (2010).
859 37 Jangam, D., Feschotte, C. & Betran, E. Transposable Element Domestication As an
860 Adaptation to Evolutionary Conflicts. *Trends Genet* **33**, 817-831,
861 doi:10.1016/j.tig.2017.07.011 (2017).
862 38 Bergero, R., Gardner, J., Bader, B., Yong, L. & Charlesworth, D. Exaggerated
863 heterochiasmy in a fish with sex-linked male coloration polymorphisms. *Proc Natl Acad*
864 *Sci U S A* **116**, 6924-6931, doi:10.1073/pnas.1818486116 (2019).
865 39 Sardell, J. M. *et al.* Sex Differences in Recombination in Sticklebacks. *G3-Genes*
866 *Genom Genet* **8**, 1971-1983, doi:10.1534/g3.118.200166 (2018).
867 40 Sutherland, B. J. G., Rico, C., Audet, C. & Bernatchez, L. Sex Chromosome Evolution,
868 Heterochiasmy, and Physiological QTL in the Salmonid Brook Charr Salvelinus
869 fontinalis. *G3 (Bethesda)* **7**, 2749-2762, doi:10.1534/g3.117.040915 (2017).
870 41 Matsuda, M., Sotoyama, S., Hamaguchi, S. & Sakaizumi, M. Male-specific restriction
871 of recombination frequency in the sex chromosomes of the medaka, Oryzias latipes.
872 *Genet Res* **73**, 225-231, doi:Doi 10.1017/S0016672399003754 (1999).
873 42 Lemos, B., Branco, A. T. & Hartl, D. L. Epigenetic effects of polymorphic Y
874 chromosomes modulate chromatin components, immune response, and sexual conflict.
875 *Proc Natl Acad Sci U S A* **107**, 15826-15831, doi:10.1073/pnas.1010383107 (2010).
876 43 Kent, T. V., Uzunovic, J. & Wright, S. I. Coevolution between transposable elements
877 and recombination. *Philos T R Soc B* **372**, doi:ARTN 20160458
878 10.1098/rstb.2016.0458 (2017).
879 44 McGaugh, S. E. *et al.* Recombination Modulates How Selection Affects Linked Sites in
880 Drosophila. *Plos Biol* **10**, doi:ARTN e1001422

881 10.1371/journal.pbio.1001422 (2012).

882 45 Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**,
883 1079-1083, doi:10.1038/nature08441 (2009).

884 46 Chen, S. L. *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex
885 chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* **46**, 253-+,
886 doi:10.1038/ng.2890 (2014).

887 47 Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-
888 chromosome degeneration. *Nat Rev Genet* **14**, 113-124, doi:10.1038/nrg3366 (2013).

889 48 Harboe, T., Mangor-Jensen, A., Naas, K. E. & Naess, T. A tank design for first feeding
890 of Atlantic halibut, Hippoglossus hippoglossus L., larvae. *Aquac Res* **29**, 919-923,
891 doi:DOI 10.1046/j.1365-2109.1998.29120919.x (1998).

892 49 Harboe, T., Tuene, S., Mangorjensen, A., Rabben, H. & Huse, I. Design and Operation
893 of an Incubator for Yolk-Sac Larvae of Atlantic Halibut. *Prog Fish Cult* **56**, 188-193,
894 doi:Doi 10.1577/1548-8640(1994)056<0188:Daooai>2.3.Co;2 (1994).

895 50 Norberg, B., Valkner, V., Huse, J., Karlsen, I. & Grung, G. L. Ovulatory Rhythms and
896 Egg Viability in the Atlantic Halibut (Hippoglossus-Hippoglossus). *Aquaculture* **97**,
897 365-371, doi:Doi 10.1016/0044-8486(91)90328-5 (1991).

898 51 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of
899 image analysis. *Nat Methods* **9**, 671-675, doi:10.1038/nmeth.2089 (2012).

900 52 Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long
901 sequences. *Bioinformatics* **32**, 2103-2110, doi:10.1093/bioinformatics/btw152 (2016).

902 53 Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome
903 assembly from long uncorrected reads. *Genome research* **27**, 737-746,
904 doi:10.1101/gr.214270.116 (2017).

905 54 Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant
906 Detection and Genome Assembly Improvement. *PloS one* **9**, doi:ARTN e112963
907 10.1371/journal.pone.0112963 (2014).
908 55 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M.
909 BUSCO: assessing genome assembly and annotation completeness with single-copy
910 orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).
911 56 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct
912 determination of diploid genome sequences. *Genome research* **27**, 757-767,
913 doi:10.1101/gr.214874.116 (2017).
914 57 Dunn, C. W., Howison, M. & Zapata, F. Agalma: an automated phylogenomics
915 workflow. *BMC bioinformatics* **14**, 330, doi:10.1186/1471-2105-14-330 (2013).
916 58 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
917 sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170
918 (2014).
919 59 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
920 MEM. *arXiv* **1303.3997** (2013).
921 60 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the
922 Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10
923 11-11 10 33, doi:10.1002/0471250953.bi1110s43 (2013).
924 61 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
925 *Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
926 62 Kent, W. J. BLAT - The BLAST-like alignment tool. *Genome research* **12**, 656-664,
927 doi:10.1101/gr.229202 (2002).
928 63 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
929 throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

930 64 Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large
931 datasets. *Bioinformatics* **30**, 3276-3278, doi:10.1093/bioinformatics/btu531 (2014).
932 65 Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing.
933 *Nat Methods* **14**, 407-+, doi:10.1038/nmeth.4184 (2017).
934 66 Li, X. Y., Chen, F. & Chen, Y. P. Gcluster: a simple-to-use tool for visualizing and
935 comparing genome contexts for numerous genomes. *Bioinformatics* **36**, 3871-3873,
936 doi:10.1093/bioinformatics/btaa212 (2020).
937 67 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
938 doi:10.1093/bioinformatics/bts635 (2013).
939 68 Grabherr, M. G. *et al.* Genome-wide synteny through highly sensitive sequence
940 alignment: Satsuma. *Bioinformatics* **26**, 1145-1151, doi:10.1093/bioinformatics/btq102
941 (2010).
942 69 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
943 genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033
944 (2010).
945 70 Rastas, P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome
946 sequencing data. *Bioinformatics* **33**, 3726-3732, doi:10.1093/bioinformatics/btx494
947 (2017).
948
949
950 **Table 1. Brx PCR primers**

PCR primer	5'-3' sequence
brx_fwd	CAAGCCTGTTCACATCGAAA
brx_malerev	ATGGCAACAAACGCTTATCAA
brx_femalerev	TGGCAACAAACGCTTAACAC

951

952 **Table 2. Primers used for qPCR**

QPCR primer	5'-3' sequence
gsdf forward	TGCTGGACTCGTCCACATAG
gsdf reverse	GTCCATCATCCCACACCAAT
gtf3c6 forward	CACCGTGAAGAAGCTGATGA
gtf3c6 reverse	ACTTCCTGCTGTGCTGGTTT

953

954 **Acknowledgements**

955 We would like to thank the National Genomics Infrastructure/SNP&SEQ Technology Platform
956 in Uppsala for preparing and sequencing the 10X Genomics Chromium Genome library and
957 thank the Norwegian Sequencing Centre (Oslo) for Illumina sequencing of DNA and RNA.
958 Computations and data handling were enabled by resources provided by the Swedish National
959 Infrastructure for Computing (SNIC) at Uppmax, Uppsala, Sweden, with these resources being
960 partially funded by the Swedish Research Council through grant agreement no. 2018-05973.
961 We would like to thank Leif Andersson for comments on an earlier version of the manuscript.
962 We are grateful for financial support from Sterling White Halibut for the sex determining factor
963 part of this work. This study was co-funded by the Norwegian Ministry of Trade, Industry and
964 Fisheries. This research was supported by grants from the Swedish Research Councils Formas
965 (2012-00740 to C.J.R.; 2018-01008 to P.J.; 2017-00413 to A.W) and Vetenskapsrådet 2018-
966 03017 to P.J.; 2018-04444 to A.W.).

967

968 **Author contributions**

969 R.B.E., C.J.R and B.N. designed the project. T.H., R.M.J. and M.M. contributed with rearing
970 of fish. R.M.J., P.P. and B.N. collected the tissue samples. S.K.O. extracted the RNA. L.K.

971 performed the qPCR. E.K.S. and C.J.R performed the structural variant bioinformatic analysis.
972 C.J.R. and O.W. did HMW DNA isolation, nanopore sequencing and genome assembly. E.S.
973 generated RNA-sequencing data. A.W. and C.J.R. performed analysis of RepeatMasker data.
974 C.J.R performed and P.J. contributed to the TE analysis. C.J.R performed variant calling,
975 comparative genomics- and divergence analyses. R.B.E. T.F and O.W. performed annotation.
976 R.B.E. designed, and L.K. and S.M. tested the PCR sexing assay. R.B.E., C.J.R and B.N.
977 analyzed the results and wrote the paper. All authors read and approved the final manuscript.
978

979 **Competing interests**

980 The authors declare that they have no competing interests.
981

982 **Data Availability**

983 The authors declare that the data supporting the findings of this study are available within the
984 article and its Supplementary Information. All datasets will be made publicly available under
985 bioproject PRJNA680301 upon publication.
986