DR. MELISSA K HOLBORN (Orcid ID : 0000-0002-5729-6039)
DR. ANTHONY EINFELDT (Orcid ID : 0000-0003-4588-3479)

Article type : Resource Article

Journal target: Molecular Ecology Resources

Reference genome of Lumpfish *Cyclopterus lumpus* Linnaeus provides evidence of male heterogametic sex determination through the AMH pathway

Short title: Lumpfish putative sex determining factor

Melissa K. Holborn[1], Anthony L. Einfeldt[2], Tony Kess[1], Steve J. Duffy[1], Amber M. Messmer[1], Barbara L. Langille[1], Johanne Gauthier[3], Paul Bentzen[2], Tim Martin Knutsen[4], Matthew Kent[5], Danny Boyce[6], Ian R. Bradbury[1]

[1]Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada, St. John's, NL, A1C 5X1, Canada

[2]Marine Gene Probe Laboratory, Department of Biology, Dalhousie University, Halifax, NS, B3H 4R2, Canada

[3]Maurice Lamontagne Institute, Fisheries and Oceans Canada, Mont-Joli, QC, G5H 3Z4
[4]AquaGen AS, P.O. Box 1240, 7462, Trondheim, Norway
[5]Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway
[6]Department of Ocean Sciences, Ocean Sciences Centre, Memorial University of Newfoundland, St John's, NL, A1C 5S7, Canada

*Corresponding author email: melissaholborn@gmail.com

ORCID

Melissa Holborn	0000-0002-5729-6039
Anthony Einfeldt	0000-0003-4588-3479
Tony Kess	0000-0002-1079-3791
Barbara Langille	0000-0002-7984-4593
Paul Bentzen	0000-0003-4288-7730

Abstract

Teleosts exhibit extensive diversity of sex determination (SD) systems and mechanisms, providing the opportunity to study the evolution of sex determination and sex chromosomes. Here we sequenced the genome of the Common Lumpfish (*Cyclopterus lumpus* Linnaeus), a species of increasing importance to aquaculture, and identified the SD region and master SD locus using a 70K SNP array and tissue-specific expression data. The chromosome-level assembly identified 25 diploid chromosomes with a total size of 572.89 Mb, a scaffold N50 of 23.86 Mb, and genome annotation predicted 21,480 protein-coding genes. Genome wide association analysis located a highly sex-associated region on chromosome 13, suggesting that anti-Müllerian hormone (AMH) is the putative SD factor. Linkage disequilibrium and heterozygosity across chromosome 13 support a proto-XX/XY system, with an absence of widespread chromosome divergence between sexes. We identified three copies of *AMH* in the Lumpfish primary and alternate haplotype assemblies localized in the SD region. Comparison to sequences from other teleosts suggested a monophyletic relationship and conservation within the Cottioidei. One *AMH* copy showed similarity to *AMH/AMHY* in a related species and was also the only copy with expression in testis tissue, suggesting this copy may be the functional copy of *AMH* in Lumpfish. The two other copies arranged in tandem inverted duplication were highly similar, suggesting a recent duplication event. This study provides a resource for the study of early sex chromosome evolution and novel genomic resources that benefits Lumpfish conservation management and aquaculture.

Keywords: genome assembly, Lumpfish, sex determination factors, genome wide association analysis

1. Introduction

Separate sexes have evolved independently in many Eukaryotic lineages, and the nature of sexual systems and the mechanisms underlying sex determination are remarkably diverse (Bachtrog et al., 2014; The Tree of Sex Consortium, 2014). This variety of sex determination systems begets questions of how and why transitions between different systems occur (The Tree of Sex Consortium, 2014; Pennell, Mank, & Peichel, 2018). While transitions between sex determining mechanisms are

common, the rates at which they occur vary among lineages (Capel, 2017; Quinn et al., 2011). In vertebrates, eutherian mammals and avians have highly conserved master sex determination genes located on diverged sex chromosomes (Chue & Smith, 2011; Koopman, Gubbay, Vivian, Goodfellow, & Lovell-Badge, 1991). However, in other vertebrates, sex determination is varied and more than one mechanism often exists within clades and rarely within species (reviewed in Capel, 2017; Mank, Promislow, & Avise, 2006; Sarre, Ezaz, & Georges, 2011). How and why novel sex determination systems evolve remains poorly understood (Bachtrog et al., 2014). The increasing availability of genomic tools in non-model species, such as genome assemblies and high-density molecular markers, can help elucidate the evolution of novel sex determination systems by identifying underlying sex determination factors and the genomic changes associated with transitions between systems.

Fish exhibit an exceptionally high diversity of sex determination systems and underlying mechanisms (Devlin & Nagahama, 2002; Pennell, Mank, & Peichel, 2018; The Tree of Sex Consortium, 2014), making them an excellent system to study the evolution of sex determination systems. Mechanisms of sex determination in fish include a broad mixture of systems, involving factors that are genetic, environmental with cues such as temperature, density of conspecifics, pH, and oxygen concentration, and combinations of both (Baroiller, D'Cotta, & Saillant, 2009; Devlin & Nagahama, 2002; Heule, Salzburger, & Böhne, 2014). Studies of genetic sex determination in teleost fishes have identified both male (XY) and female (ZW) heterogametic systems, with XY/XX systems being the most prevalent (The Tree of Sex Consortium, 2014), as well as transitions between these systems in some lineages (Einfeldt et al., 2021; Matsuda & Sakaizumi, 2016). Sex chromosomes are classically thought to evolve predominately from a pair of autosomes after a sex determining locus is acquired, progressing from a homomorphic to a heteromorphic state through divergence and degradation (Charlesworth, Charlesworth, & Marais, 2005). However, evidence suggests sex chromosome evolution may not be linear across all taxa; not all known sex chromosomes are heteromorphic and certain groups appear to have higher sex chromosome turnover (Bachtrog et al., 2014; Furman et al., 2020; Myosho, Takehana, Hamaguchi, & Sakaizumi, 2015; Volff, Nanda, Schmid, & Schartl, 2007). Sex chromosome turnover can occur due to the translocation of an existing sex determining locus to an autosomal chromosome (Lubieniecki et al., 2015; Traut & Willhoeft,

1990), the fusion of an ancestral sex chromosome with an autosome (Ross et al., 2009), or a gene acquiring a sex determining role through mutation (Myosho et al., 2012), after which the novel sex chromosome becomes fixed through various evolutionary forces (reviewed in Palmer, Rogers, Dean, & Wright, 2019). Taxa with sex determination systems that differ from their relatives can therefore provide novel insights into the evolution of sex chromosomes by offering a genomic snapshot of the early changes that occur during transitions between sex determination systems.

The Lumpfish *Cyclopterus lumpus* Linnaeus is a gonochoristic, commercially important teleost that inhabits the North Atlantic Ocean (Davenport, 1985). Previous fishing for Lumpfish focused on sex-specific harvest for roe as a caviar substitute, but recently harvest has increased due to their usage as a biological control for sea lice in Atlantic salmon aquaculture leading to the development of Lumpfish aquaculture and breeding programs (Davenport, 1985; Imsland et al., 2014; Imsland et al., 2015, Powell et al., 2018). Lumpfish is the only species in the *Cyclopterus* genus, which along with *Liparis* are basal to the Cycloptteriae family (Voskoboinikova et al. 2020). Although important to the conservation management and development of breeding programs, little is currently known about sex determination in Lumpfish and sex-specific markers have not been developed. Feminization experiments revealed that males are likely the heterogametic sex based on sex ratio differences in a single feminized genotypic male versus a genotypic female (Martin-Robichaud, Peterson, Benfey, & Crim, 1994), consistent with an XX/XY genetic sex determination system. Early karyotyping detected an average of 25 chromosome pairs, but no heteromorphic sex chromosomes were visibly distinguishable (Li & Clyburne, 1977). Since these studies, efforts to sequence the whole genomes of non-model organisms have increased, providing an abundance of genetic information to elucidate sex determination in combination with population genomic data.

In this study, we used a novel chromosome-level genome assembly combined with genome wide SNP array data and tissue-specific transcriptomes to identify the putative sex determining region in Lumpfish. This study had two core aims: first, to provide the first genome sequence assembly for Lumpfish and locate the sex determining region for future development of sex-specific molecular markers, and second, to identify the sex determination pathway in Lumpfish to expand our understanding of sex chromosome evolution in teleosts. This study represents the first genomic analysis of sex determination in the Cyclopteridae (Lumpfishes) family, filling a gap in knowledge of

sex chromosome evolution and turnover in the teleost phylogeny. Our results identified the anti-Müllerian hormone (*AMH*) as the putative sex determining locus, and characterized three *AMH* copies, one of which is a tandem inverted duplication. By comparing these *AMH* copies to *AMH* orthologs in other species in which it has been identified as the male sex determining region, we provide evidence of parallel evolution of AMH as a master sex determination factor through independent evolutionary events.

2. Methods

2.1 Sample collection, DNA extraction, and genotyping

A total of 142 North American Lumpfish with phenotypically determined sex via dissection were used in this study, comprising 99 males and 43 females collected from two broad sampling locations. The Lumpfish breeding program at the Ocean Science Centre (OSC) in St. John's, Newfoundland, Canada, provided 52 fin clips from individuals of known sex from the 2017 year class. These fish represent the offspring from broodstock captured at multiple collection sites in Newfoundland. Additionally, 44 and 43 wild Lumpfish fin clips were collected through Fisheries and Oceans Canada trawl surveys in the Gulf of St. Lawrence (Bourdages et al., 2020) in late summer of 2018 and 2019, respectively (Figure 1). All experimental procedures and fish handling were approved and conducted in accordance with the Canadian Council of Animal Care Guidelines and Memorial University of Newfoundland animal utilization protocols. All tissue samples were preserved in 95% ethanol before DNA extraction using DNeasy 96 Blood and Tissue kits (Qiagen) according to manufacturer's protocols. The quality of extracted genomic DNA was visualized by 1% agarose gel electrophoresis and quantified using Quant-iT PicoGreen ds-DNA Assay kits (Thermofisher) on a fluorescent plate reader. Genomic DNA was normalized to 15 ng/µl and sent to the Centre of Integrative Genomics (CIGENE, Ås, Norway) where it was genotyped on a Lumpfish custom Affymetrix array developed by Aquagen and CIGENE (manuscript in preparation) containing 69,062 SNPs.

2.2 Genome assembly and annotation

A single adult male Lumpfish from the OSC was selected for sequencing to assemble a reference genome as part of the Vertebrate Genomes Project (VGP). Muscle tissue from the heart was flash

frozen on dry ice for sequencing and assembly using the VGP standard assembly pipeline v1.6 (see Rhie et al., 2021). Briefly, multiple sequencing and scaffolding technologies were combined (PacBio continuous long reads, 10x-Genomics linked reads, Arima Genomics Hi-C chromatin conformation capture, and Bionano Genomics DLS optical mapping). The primary pseudo-haplotype and alternate haplotype were generated with FALCON-Unzip, scaffolds were generated (Scaff10x, Bionano Solve, Salsa), gap-filling and polishing was performed using the Arrow algorithm, and final short-read polishing using FreeBayes. Assembly software, details, and rationale for methods can be found in Rhie et al. 2021, which details the VGP assembly pipeline; assembler version details are available at https://www.ncbi.nlm.nih.gov/assembly/GCA_009769545.1. The assembly was then manually curated by the gEVALteam at Wellcome Sanger Institute to correct assembly errors. Genome annotation was performed with two gene prediction pipelines, the NCBI annotation pipeline v8.4 including automated gnomon gene predictions (Souvorov et al., 2010) and Ensembl annotation (Yates et al., 2020). We assessed assembly completeness using BUSCO v5.2.2 (Benchmarking Universal Single-Copy Orthologs; Manni et al., 2021) with the Actinopterygii gene set (actinopterygii_odb10, busco.ezlab.org).

The physical location for each array SNP was mapped to the Lumpfish chromosome-level genome assembly. For each SNP locus, 100 bp of flanking sequence was aligned to the Lumpfish genome using the Burrows-Wheeler algorithm implemented using BWA-MEM (Li, 2013). The top alignment for each array SNP was then used as the physical position for each locus.

The transcriptomes of brain, liver, and testis tissues were generated from the same male Lumpfish as the genome assembly using PacBio long read isoform sequencing (Iso-Seq). PacBio Iso-Seq libraries were constructed per the PacBio Iso-seq protocol. Tissue was preserved in RNAlater and frozen on dry ice and total RNA was isolated from each tissue and used for cDNA synthesis via reverse transcription. For RNA extraction, we used the QIAGEN RNAeasy Protect kit (Cat. No 74124). Reverse transcription was then carried out with the NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module (New England BioLabs, cat. no. E6421S) and Iso-Seq Express Oligo Kit (PacBio PN 10 1-737-500). The protocol followed was “Procedure & Checklist – Iso-Seq™ Express Template Preparation for Sequel® and Sequel II Systems (PN 101-763-800 Version 01)”. These libraries were then sequenced on the PacBio Sequel platform.

2.3 Association analysis

To locate regions of the genome involved in sex determination, we performed a genome wide association study (GWAS) to test for associations between SNP variants and phenotypic sex. SNP and sample level filtering was performed in PLINK v1.9 (Chang et al., 2015) with quality control thresholds of SNP and sample call rates >0.95 and MAF >0.01. The relationship inference function in KING (Manichaikul et al., 2010) was used to identify and remove any duplicated genotypes.

As Lumpfish were from wild and domestic origins, both population stratification and relatedness were accounted for to avoid spurious associations (Freedman et al., 2004; Voight & Pritchard, 2005). First, principal component (PC) analysis was performed with the *pcadapt* R package (Luu, Bazin, & Blum, 2017) using the set of filtered SNPs to obtain PC scores for each fish. Visual inspection of the scree plot indicated three PCs (K=3) was the optimal value for K. A genetic kinship matrix was computed using all available SNPs with the R package *gaston* (Dandine-Roulland & Perdry, 2018) to account for both familial and cryptic relatedness in the samples, especially the highly related OSC samples due to breeding design, however relatedness could be present in the wild samples due to limited dispersal or natal homing behaviour (Fréchet et al. 2011; Kennedy, Jónsson, Kasper, & Ólafsson, 2015). Phenotypic sex was converted to binary phenotypes (females 0, males 1) as the dependant variable in the association model. We performed GWAS tests using a mixed logistic regression (MLR) model using the *milorGWAS* package in R v3.5.3 (Milet, Courtin, Garcia, & Perdry, 2020; R Core Team, 2019). This approach is advantageous for structured populations and has the added benefit of estimating SNP variant effect size. Several GWAS models were tested, including MLR without accounting for population structure, MLR with PC1 as a covariate, MLR with both PC1 and PC2 as covariates, and MLR with the sampling location population as a covariate (Gulf of St. Lawrence or OSC), and although all models had slight genomic under-inflation indicating overcorrecting for structure, the model with the kinship matrix and the first PC scores as a covariate yielded the least amount of under-inflation and thus best goodness of fit. Multiple testing was controlled using a Bonferroni correction with an alpha value of 0.05 and Benjamini–Hochberg false discovery rate (FDR) using the p.adjust function in the *stats* R package (Benjamini–Hochberg, 1995; R Core Team, 2019).

To evaluate genomic divergence between sex chromosomes, we estimated linkage disequilibrium (LD) and heterozygosity across the identified sex chromosome separately for males and females. LD was evaluated by calculating the pairwise r^2 using PLINK v1.9 for males and females separately, as well as the genome wide r^2 medians for each sex as a baseline. The r^2 medians were calculated using a sliding window size of 1 megabase pairs (Mbp) and a step size of 0.1 Mbp. Heterozygosity was calculated using the *hierfstat* package in R (Goudet, 2005), with local regression (loess smoothing) of observed heterozygosity values for each SNP along the sex chromosome fitted to detect differences in heterozygosity between males and females.

2.4 Sex determination factor

Genes located within 100 kilobase pairs (kbp) of SNPs significantly associated with sex were identified using the NCBI and the Ensembl gene annotations. If gene annotation information was lacking, genes within this flanking region were compared to Teleost species (taxid: 324423) in the Nucleotide collection (nt) using the BLASTN 2.10.1+ algorithm with default cut-off (E > 0.05) (Zhang, Schwartz, Wagner, & Miller, 2000), and Ensembl gene names were used to determine function and thus potential involvement in sex determination.

Genes within the flanking regions that were found to be associated with sex determination in other teleost species were investigated further. Three *AMH* variants were detected, two from the primary pseudohaplotype and one from the alternative pseudohaplotype, within the Lumpfish genome. Tissue-specific expression of these *AMH* genes was investigated using the Iso-Seq transcriptomic data by aligning the transcripts from brain, liver, and testis to these genomic regions in the reference assembly using LAST v9.6.3 (Kielbasa et a. 2011). LAST was run with a minimum alignment score of 0.05 (-E flag), a generalized affine gap cost of 2 (-C flag), and a multiplicity score of 100 (-m) for the Iso-Seq transcripts from all three tissue types. We then assessed the sequence similarity for the three *AMH* variants expressed in the gonadal tissue for both the primary and alternative pseudohaplotype. The sequences for the *AMH* variants were manually inspected for break points based on their position within the genome. Sequence similarity was assessed using MUSCLE (Edgar 2004) and the transcripts from the gonadal tissue were aligned to the three *AMH* variants using

BLAT (Kent 2002). This may provide secondary evidence of *AMH* as the sex determination locus as expression of genes involved in sex determination would be expected in gonadal tissue.

2.5 Phylogenetic analysis of *AMH* sequences

The nucleotide sequences of the three detected Lumpfish *AMH* variants were compared with *AMH* sequences from other species available in GenBank. *AMH* sequences were selected from other teleost species, including some with AMH documented as the sex determination factor, and thus having male specific *AMH* sequences (*AMHY*), and some with other genes responsible for sex determination. To assess relationships with phylogenetically related species, any known *AMH* sequences for species within the Cottioidei suborder were used. Nucleotide sequences were translated into amino acids to account for multiple substitutions present at the same base pair, and multiple alignments were performed using the ClustalW algorithm (Thompson, Higgins, & Gibson, 1994) in MEGA X (version 10.1) using default parameters (Kumar, Stecher, Li, Knyaz, & Tamura, 2018; Stecher, Tamura, & Kumar, 2020). The sequences for *Ophiodon elongatus* Girard *AMH* and *AMHY* (accession numbers KP686074.1 and KP686073.1), *Oreochromis niloticus* Linnaeus *AMH* (EF512167.1), *Salmo salar* Linnaeus *AMH* (NM_001123585.1), *Anarrhichthys ocellatus* Ayres *AMH* (XM_031864052.1), *Anoplopoma fimbria* Pallas *AMH* (KC112919.1), *Odontesthes bonariensis* Valenciennes *AMH* and *AMHY* (AY763406.2 and KC847082.1), *Odontesthes hatcheri* Eigenmann *AMH* and *AMHY* (DQ441594.2 and HM153803.1), *Oryzias latipes* Temminck and Schlegel *AMH* (NM_001360941.1), *Plecoglossus altivelis* Temminck and Schlegel *AMH* (LC512015.1), and *Xenopus laevis* Daudin *AMH* (AB548671.1) as the outgroup were used. A phylogenetic tree was inferred by the Neighbor-Joining method (Saitou & Nei, 1987) with 1000 bootstrap replicates. FigTree (version 1.4.4) was used to visualize the phylogenetic tree (Rambaut, 2018).

3.0 Results

3.1 Genomic and genetic data

The final chromosome-level primary genome assembly contained 572.89 Mb with 55.96x genome coverage (where coverage reflected PacBio coverage of the final assembly) and contained 48 scaffolds, and a scaffold N50 of 23.86 Mb and N90 of 16.67 Mb (GenBank assembly accession

GCA_009769545.1, coverage per sequencing technology available at: https://vgp.github.io/genomeark/Cyclopterus_lumpus/). The genome is comprised of 25 diploid chromosomes, consistent with previous karyotyping (Li & Clyburne, 1977). Only 23 small sequences could not be placed on a chromosome (total length = 0.82 Mb). Genome annotation identified 26,630 genes and pseudogenes, of which 21,480 were identified as protein-coding (breakdown of annotation of genes and pseudogenes and mRNAs are at: NCBI *Cyclopterus lumpus* Annotation Release 100). Genome information, such as repeats, proteins, transcripts, and annotations can be downloaded at: https://projects.ensembl.org/vgp/. Genome quality assessment resulted in the recovery of 96.9% complete BUSCO's in which there were 96.0% single-copy and 0.9% duplicated complete genes. The assembly showed 92 missing genes out of a total of 3640 BUSCO's investigated (2.5%).

A total of 69,050 SNPs were successfully genotyped, of which 1018 could not be placed using the genome assembly and nine were placed to small scaffolds that have yet to be placed within the genome. A total of 53,253 SNPs and 139 fish consisting of 96 males and 43 females passed all quality control filters (Table 1).

3.2 Association analysis

The first PC explained 10.7% of the variance, with visual separation of most samples from wild and domesticated origin (Figure S1). However, some domestic samples clustered with the wild samples possibly due to recent introduction to the breeding program or catch origin. The second PC separated individuals within the domesticated samples, and likely was due to familial structure. Therefore, inclusion of the second PC was likely causing genomic under-inflation as the kinship matrix was already accounting for relatedness.

The GWAS accounting for kinship and population structure identified seven significant SNPs on chromosome 13 using a false discovery rate significance threshold (of which four were significant with a more conservative Bonferroni threshold), and one SNP on chromosome 21 (Figure 1; Table 2). The significant SNPs on chromosome 13 formed a distinct peak at ~13.6 Mb, while other SNPs on chromosome 13 showed an elevated association with sex compared to other chromosomes. The top SNP (AX-298030667) had a large difference in the allele frequencies between the sexes, consistent with a proto-XX/XY genetic sex determination system. While males had a relatively equal ratio of the

A_1 and A_2 alleles (allele frequencies of 0.516 and 0.484 respectively, approximating an XY genotype), females had an A_1 allele frequency much lower than the A_2 frequency (allele frequencies of 0.128 and 0.872 respectively, approximating an XX genotype) resulting in a large negative effect size.

LD for males and females was relatively consistent between the sexes across the length of chromosome 13 (Figure 2A). Most of the chromosome had LD levels that matched the genomic average, except for a region (~10 Mb to 13 Mb) adjacent to gene highly likely to be involved in the core sex determination pathway (see section 3.3), which has elevated levels of LD in both males and females. We found the top heterozygosity values across the chromosome in significant SNPs located around the putative sex determination gene specific to males, with the highest heterozygosity observed at *AMH* in males, whereas female heterozygosity values did not exceed the genome wide average at these loci (Figure 2B).

3.3 Sex determination factor

Gene content examination in the flanking regions of the peak SNP(s) on chromosomes 13 and 21 yielded several candidate genes (Table 3). On chromosome 13, geneID 117741289 was located within the flanking (13.7 kbp from top SNP) region and had a gene prediction product Müllerian -inhibiting factor-like (aka anti-Müllerian hormone, *AMH*), which is involved in the teleost sex determination pathway and has been documented as a master sex determination gene in other fishes (Hattori et al., 2012; Pan et al., 2019, Peichel et al., 2020; Nakamoto et al., 2021). GeneID 117741803 (9.63 kpb from top SNP) was uncharacterized and 1729 bp in length, but BLASTN results for this gene showed similarity to *AMH* in several other teleost species, with close similarity to *O. elongatus* (702/793 bp alignment length; 89 % similarity, E = 0.) and *A. ocellatus* (687/792 bp alignment length; 87 % similarity, E = 0.). Upstream of the *AMH* locus, two significant SNPs were located near and within the nectin cell adhesion molecule 3a (*nectin3a*) gene, which is known to be involved in spermatogenesis in mice (Inagaki et al., 2006). No genes of known potential sex determination processes or sex related traits were detected within the 100 kbp region of the top SNP on chromosome 21, but the gene intraflagellar transport 88 (*IFT*88) which is essential to mammalian spermatogenesis (Agustin, Pazour, & Witman, 2015) was detected adjacent to the flanking region spanning 17624 to

54559 bp. As *AMH* has been previously identified as a sex determination gene (Hattori et al., 2012; Pan et al., 2019, Peichel et al., 2020; Nakamoto et al., 2021) and no other genes of known sex determination function were identified, we focused further investigation on the *AMH* locus on chromosome 13.

We detected three copies of *AMH* in the Lumpfish genome, hereafter referred to as *AMH1*, *AMH2*, and *AMH3*. *AMH1* and *AMH2* are paralogs found in a tandem inverted arrangement in the primary assembly, representing one haplotype from chromosome 13, and *AMH3* was found in the alternate assembly and represents a second haplotype of chromosome 13. Long PacBio sequence reads spanning *AMH1*, *AMH2*, and the region between them support tandem inverted duplication and provide evidence against this feature being an assembly error. We used tissue-specific Iso-Seq analysis to determine whether the *AMH* copies are expressed. We identified a total of 36,486 transcripts from brain tissue, 22,235 from liver, and 31,315 from testis. An *AMH* transcript was detected in testis tissue, but no *AMH* transcripts were identified in brain or liver tissues. The testis-specific *AMH* transcript had identical sequence similarity to genomic sequence for the *AMH3* copy from the alternate haplotype assembly, and differed from *AMH1* and *AMH2* in the primary assembly. These results indicate that *AMH3* has tissue-specific expression in adult male gonad tissue.

3.4 Teleost *AMH* phylogenetic relationships

Phylogenetic analysis revealed a monophyletic relationship between the AMH sequences of all species within the Cottioidei suborder: Lumpfish, *O. elongatus*, *A. ocellatus*, and *A. fimbria* (Figure 3). The *O. bonariensis* and *O. hatcheri AMH* and *AMHY* sequences showed more similarity within gene than within species, consistent with previous studies (Bej, Miyoshi, Hattori, Strüssmann, & Yamamoto, 2017). Lumpfish *AMH1* and *AMH2* (located on the same genomic haplotype) clustered together by similarity, suggesting that one arose from the other via a recent inverted duplication event. Lumpfish *AMH3* (from the alternate haplotype assembly) clustered as an outgroup to Lumpfish *AMH1* and *AMH2*.

4.0 Discussion

High-quality genome assemblies are becoming more attainable for non-model species, and their development opens novel avenues of genomic investigation to important research questions in evolutionary biology, conservation, and the genetic underpinnings of economically important traits. The chromosome-level genome Lumpfish assembly presented here, in combination with population genomic data, allowed us to identify a locus that putatively acts as the sex determining master switch and is consistent with a male heterogametic sex determination system. We detected SNPs significantly associated with phenotypic sex located near *AMH*, a gene known to be involved in several sex determination pathways across teleosts (Hattori et al., 2012; Pan et al., 2019, Peichel et al., 2020; Nakamoto et al., 2021). Investigation of the genome sequence of the *AMH* locus revealed multiple copies, including a tandem inverted duplication of *AMH*. Phylogenetic analysis revealed that Lumpfish *AMH* sequences clustered closely with *AMH* and *AMHY* sequences of all other Cottioidei species. Together, genome wide association, evolutionary conservation, and tissue-specific expression of an *AMH* copy suggest that AMH is a master sex determination factor in Lumpfish. We have support for an XY/XX sex determination system in Lumpfish, suggesting the presence of a male-specific sex determining factor, likely *AMH3* as this *AMH* copy was expressed in the testis tissue but not other tissues. Knowledge of the sex determining region in Lumpfish will aid the development of a panel of diagnostic sex SNPs for application to management programs and future aquaculture breeding programs. Our results provide evidence that changes to a single gene in the core teleost sexual determination pathway led to the evolution of XX/XY genetic sex determination in Lumpfish but has not yet facilitated chromosome-wide differentiation, providing insight into the early stages of sex chromosome differentiation.

4.1 Genome assembly and annotation

The genome size of Lumpfish was 572.89 Mb, comparable to that of other Cottioidei species that have available sequenced genomes: *A. ocellatus* (612.79 Mb, GCF_004355925.1), *A. fimbria* (784.19 Mb, GCA_000499045.2), and *O. elongatus* (645.89 Mb, GCA_016806645.1), and represents the first published genome of a North American Lumpfish. A draft genome of a Norwegian origin Lumpfish (genome size of 553 MB, scaffold N50 = 811 KB) is publicly available (Knutsen, Kirubakaran, Mommens, & Moen, 2018) but is less complete than the assembly described here. In general, the

genomes of Cottiodei species were smaller (~ 0.65 Gb on average) than the genomes of species within *Oreochromis*, *Oryzia*, *Danio*, and *Salmo* genera (~1.42 Gb on average), implying that genome compaction may have occurred in a common ancestor to the Cottiodei suborder. A comprehensive comparative analyses of related species will be essential in investigating the potential for genome compaction in this group once more genomes are available. Annotation identified 21,480 protein-coding genes in the genome, which will benefit the identification of candidate genes in regions significantly associated with phenotypic traits. Historically harvested for roe, Lumpfish are gaining momentum as essential delousing agents in salmon aquaculture, and therefore, the reference genome assembled in this study will provide a valuable resource for future studies involving species management and conservation, adaptive evolution, and enhancement of aquaculture potential.

4.2 AMH as the sex determining factor in Lumpfish

Several SNPs on chromosome 13 were significantly associated with sex, suggesting this chromosome contains a sex determining factor. The SNPs with the greatest association with sex were located near the *AMH* locus, a known sex determining gene in other teleosts based on genomic and expression evidence, including knockdown studies (Patagonian pejerrey, Hattori et al., 2012; Northern pike, Pan et al., 2019, threespine stickleback, Peichel et al., 2020; ayu or sweetfish, Nakamoto et al., 2021). While having a master sex determining gene is not limited to the teleosts (Kikuchi & Hamaguchi, 2013), it is the only group in which *AMH* has been found to be the master sex determining gene. Sex-specific LD and SNP heterozygosity was indicative of a proto-sex chromosome undergoing early stages of differentiation between regions linked to the sex determining factor. Consistent with an XX/XY system, heterozygosity in males was slightly elevated in regions of the chromosome near significant SNPs and a 3.2 Mbp portion of the chromosome adjacent to the *AMH* locus, while LD was elevated above the genome wide average in this adjacent region for both males and females. Recombination suppression between sex chromosomes is expected, either in the vicinity of a sex determination gene(s) or along the entire length of the chromosome (reviewed in Bachtrog et al, 2011; Charlesworth, 2017). However, recombination suppression is classically found on the sex chromosome of the heterogametic sex in highly diverged sex chromosomes (Charlesworth, 2017; Charlesworth, Charlesworth, & Marais, 2005), and Lumpfish exhibited no substantial differences

between sexes in LD on chromosome 13 as both sexes exhibited elevated LD near the *AMH* locus. This suggests either that Lumpfish have a relatively nascent sex chromosome, or that the evolutionary forces that are hypothesized to cause expansion of LD around the sex-determining factor (e.g., sexually antagonistic selection; Furman et al. 2020) are for some reason not applicable in Lumpfish. However, some lineages have shown no sex chromosome differentiation over millions of years with no change to the master sex determining gene, indicating that lack of structural differences between X and Y chromosomes does not necessarily equate to a novel sex chromosome (Pan et al., 2021).

Another region found to be associated with phenotypic sex was upstream of *AMH* on chromosome 13, and the statistically significant SNP AX-297903115 is in an intragenic region of *nectin3a*, which is known to be involved in spermatogenesis in mice (Inagaki et al., 2006). The association of this SNP locus with sex is particularly interesting, as its proximity to a gene with a role in spermatogenesis suggests this locus could be an example of sexually antagonistic selection that may facilitate the evolution of suppression of recombination on the proto-sex chromosome. The SNP located within *nectin3a* is in a chromosomal region of higher LD and heterozygosity and is close to another SNP also associated with sex (AX-297903099, 11.3 kpb upstream), and could represent accumulation of advantageous sex-specific genes on the nascent heterogametic sex chromosome. As a corollary, the presence of a sex-associated SNP on a chromosome that otherwise exhibits low sex association may lead to instability of the sex determination factor's genomic position (van Doorn & Kirkpatrick, 2007). However, it is important to note that it is unknown if the spermatogenesis functions of these genes are conserved in teleosts.

Sex determination pathways are generally conserved among vertebrates with several core genes responsible for gonadal differentiation (Smith et al., 1999; Zarkower, 2001; Schartl, 2004; Herpin & Schartl, 2015); whereas, the initial trigger gene that initiates the signalling pathway is greatly varied, especially among teleosts. *AMH* is member of the transforming growth factor beta (TGF-β) signalling pathway, which also includes *AMHR2* and *GSDF*, and these members of the TGF-β pathway have been recruited as sex determining genes in many teleost species (Kamiya et al., 2012; Myosho et al., 2012; Einfeldt et al., 2021). Y-specific duplications of *AMH* have been identified as the male inducing sex determining locus in the *O. hatcheri* (Hattori et al., 2012), *O. niloticus* (Li et al., 2015), *E. lucius* (Pan et al., 2019), and *O. elongatus* (Rondeau et al., 2016). These Y-specific

duplications of *AMH* arose through convergent evolution in teleosts as the sex determining locus (Li et al., 2015). However, not all TGF-β super family master sex determining genes have male-specific duplications, indicating the different mechanisms that this gene family initiates the sex determining pathway. Future research to determine which Lumpfish *AMH* genes are expressed in early development, before gonadal differentiation, is required to confirm which *AMH* copy is the initial sex determination trigger.

4.3 Phylogeny of Lumpfish *AMH*

Here, we provide evidence of an expressed anti-Müllerian factor homolog (*AMH3*) that is conserved with the *O. elongatus* sex determining gene *AMHY*. Phylogenetically based on morphological and genetic data, Lumpfish are more closely related to *O. elongatus* than the other Cottioidei species, *A. ocellatus* and *A. fimbria* (Smith & Busby, 2014; Smith, Everman, & Richardson, 2018). However, based on our analysis the *AMH* gene in Lumpfish is more closely related *AMH* in *A. ocellatus* and *A. fimbria.* Lumpfish and *O. elongatus* both exhibit an *AMH* duplication, but only in *O. elongatus* the *AMH* duplication appears to have evolved to adopt a sex determination role and is referred to as *AMHY*. In contrast, the duplicated *AMH* copies in Lumpfish (*AMH1* and *AMH2*) have a very high sequence similarity to each other suggesting a recent duplication event and show no evidence of being expressed, while *AMH3* is expressed in Lumpfish testis tissue and sits as an outgroup to Lumpfish *AMH1* and *AMH2*. This suggests that *AMH3* could be the functional copy of *AMH* in Lumpfish involved in sex determination, while *AMH1* and *AMH2* likely descended from an ancestral copy of *AMH*. The role of AMH as the sex determining factor does not appear to be conserved across other Cottioidei taxa. The putative sex determining locus of *A. fimbria* was identified as *GSDF* (XX/XY system) (Luckenbach, Fairgrieve, & Hayman, 2017; Rondeau et al., 2013), while *A. ocellatus* has no published information regarding their sex determination system. Since *AMH* is not the master sex determining gene in all four of these Cottioidei species, it is likely that *AMH* had been recruited in the common ancestor of Lumpfish and *O. elongatus*; however, only a small subset of Cottioidei species have genomes available making broad scale inferences difficult. This suggests that while *AMH* had evolved as the master sex determining gene in some Cottioidei species, other modifications to the underlying sex determination pathway have led to transitions in systems and underlying mechanisms

of sex determination across this lineage. This is consistent with the frequent turnover of sex chromosomes that is observed generally in teleosts. For example, in the *Oryzias* genus at least three different sex determination genes and seven sex-associated linkage groups have been identified (Matsuda & Sakaizumi, 2016). As genome assemblies of other Cottioidei species become more available, greater insight into the initial occurrence of male-specific *AMH* and turnover of sex determining factors within this lineage can be garnered.

4.4 Applications

Identification of the sex determination locus in Lumpfish has applications for both the conservation management of wild stocks and the Lumpfish aquaculture industry, which provides a biological control for sea lice. Having non-lethal methods of rapidly and accurately identifying sex is important for conservation management monitoring of sex ratios in a fishery dominated by sex-specific harvest and for aquaculture applications where industry usually retains more females than males within the breeding populations, especially for juveniles in which size and colour sexual dimorphism is less pronounced than mature adults (Daborn and Greory, 1983; Davenport, 1985; Davenport and Bradshaw, 1995; Goulet, Green, & Shears, 1986). Future resequencing efforts should focus on differentiating *AMH3* and *AMH1/2* for both phenotypically known males and females to determine which *AMH* copy is male-specific, and identify fully diagnostic SNPs for sex determination within the region of *AMH*. The chromosome-level reference genome that we present here will allow the development of genomic tools for quick sex assessment, including via resequencing.

4.5 Conclusions

In summary, using a chromosome-level genome assembly, genome wide SNP markers with phenotypic sex, and tissue-specific transcriptome data, we were able to identify a sex determining locus on a nascent sex chromosome in Lumpfish. Genome wide association identified sex associated SNPs on chromosome 13, and sequence comparisons with other teleost species identified *AMH* as a putative master sex determination locus. We found three *AMH* copies, two of which represent a tandemly inverted duplication and another (*AMH3*) that is expressed in testis tissue and is therefore likely the male-determining copy. Investigation of LD and heterozygosity along the chromosome

revealed limited suppression of recombination and accumulated divergence between the putative X and Y homologs (as indicated by patterns in XY males relative to XX females) suggesting the male heterogametic sex determination system may be recently evolved in Lumpfish. Phylogenetic analysis of Lumpfish *AMH* sequences showed sequence similarity with the Cottioidei suborder, suggesting the sex-determining *AMH* gene may have arisen in a common ancestor of this group. Knowledge gained from this study will be useful for the development of sex-specific markers to non-invasively identify sex, and our results add to a greater understanding of the evolution of sex determination in teleosts.

Acknowledgements

We thank Olivier Fedrigo for coordination of DNA extraction and sequencing for the genome assembly at VGP Rockefeller. Genome curation and gEVAL was facilitated by Wellcome Sanger Institute. We also thank the Department of Ocean Sciences Centre/Dr. Joe Brown Aquatic Research Building staff members Jessica Fry, Jennifer Monk, Kelsie Jeannot and Denise Tucker for expert technical assistance. Thanks are extended to the Canadian Coast Guard crews and to the scientists who participated in the ecosystemic research surveys in the northern Gulf of St. Lawrence allowing the collection of Lumpfish samples. SNP genotyping was provided by staff at the Centre for Integrative Genetics (CIGENE). Funding and support for this project was provided by Fisheries and Oceans Canada funding programs; the Aquaculture Collaborative Research and Development Program in partnership with Newfoundland Aquaculture Industry Association (NAIA), Grieg NL Seafarms Ltd., Cooke Aquaculture Inc., and Mowi Canada East, and the Program for Aquaculture Regulatory Research.

Author Contributions

I.R.B and P.B. conceived and initiated the study. J.G. was involved with collection of wild Lumpfish samples and D.B. was involved with rearing and sample donation of aquaculture Lumpfish. S.J.D and A.M.M. were involved with sample acquisition and laboratory preparation of Lumpfish tissues. T.K. assisted with project planning and assisted with analysis. A.L.E, B.L.L, and M.K.H performed the data analysis. M.K.H wrote the manuscript and all authors contributed to editing and revision.

Data accessibility

The Lumpfish chromosome-level genome is publicly available through GenBank, ENA, and UCSC Genome Browser: https://vgp.github.io/genomeark/Cyclopterus_lumpus/. SNP genotypes and phenotypic sexes of the 139 Lumpfish are available on Dryad (https://doi.org/10.5061/dryad.dfn2z351q). The transcriptomes for the three tissues can be found at Brain: https://s3.amazonaws.com/genomeark/species/Cyclopterus_lumpus/fCycLum1/transcriptomic_data/brain/pacbio/fCycLum1_hq_brain_transcripts.fastq Testis: https://s3.amazonaws.com/genomeark/species/Cyclopterus_lumpus/fCycLum1/transcriptomic_data/ovary/pacbio/fCycLum1_hq_ovary_transcripts.fastq Liver: https://s3.amazonaws.com/genomeark/species/Cyclopterus_lumpus/fCycLum1/transcriptomic_data/liver/pacbio/fCycLum1_hq_liver_transcripts.fastq

References

Agustin, J. T. S, Pazour, G. J., & Witman, G. B. (2015). Intraflagellar transport is essential for mammalian spermiogenesis but is absent in mature sperm. *Molecular Biology of the Cell, 26*(24), 4358-4372. https://doi.org/10.1091/mbc.E15-08-0578

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W., & Valenzuela, N. (2011). Are all sex chromosomes created equal? *Trends in Genetics, 27*(9), 350-357. https://doi.org/10.1016/j.tig.2011.05.005

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., & The Tree of Sex Consortium. (2014). Sex determination: Why so many ways of doing it? *PLOS Biology, 12*(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899

Baroiller, J. F., D'Cotta, H., & Saillant, E. (2009). Environmental effects on fish sex determination and differentiation. *Sexual Development, 3*, 118-135. https://doi.org/10.1159/000223077

Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A., & Yamamoto, Y. (2017). A duplicated, truncated *amh* gene is involved in male sex determination in an Old World silverside. *G3: Genes, Genomes, Genetics, 7*(8), 2489-2495. https://doi.org/10.1534/g3.117.042697

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Bourdages, H., Brassard, C., Desgagnés, M., Galbraith, P., Gauthier, J., Nozères, C., Scallon-Chouinard, P.-M,. & Senay, C. (2020). Preliminary results from the ecosystemic survey in August 2019 in the Estuary and northern Gulf of St. Lawrence. *DFO Canadian Science Advisory Secretariat Research Document* 2020/009, 1-93 p.

Capel, B. (2017). Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics, 18*(11), 675. https://doi.org/10.1038/nrg.2017.60

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Charlesworth, D. (2017). Evolution of recombination rates between sex chromosomes. *Philosophical Transactions of The Royal Society B Biological Sciences, 372*, 20160456. https://doi.org/10.1098/rstb.2016.0456

Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity, 95*(2), 118-128. https://doi.org/10.1038/sj.hdy.6800697

Chue, J., & Smith, C. A. (2011). Sex determination and sexual differentiation in the avian model. *The FEBS Journal, 278*(7), 1027-1034. https://doi.org/10.1111/j.1742-4658.2011.08032.x

Daborn, G. R., & Gregory, R. S. (1983). Occurrence, distribution, and feeding habits of juvenile lumpfish, *Cyclopterus lumpus* L. in the Bay of Fundy. *Canadian Journal of Zoology*, *61*(4), 797-801. https://doi.org/10.1139/z83-105

Dandine-Roulland, C., & Perdry, H. (2018). Genome-wide data manipulation, association analysis and heritability estimates in R with Gaston 1.5. *Human Heredity, 83*, 1-29. https://doi.org/10.1159/000488519

Davenport, J. (1985). *Synopsis of biological data on the lumpsucker, Cyclopterus lumpus (Linnaeus, 1758).* (FAO Fisheries Synopsis No. 147). Rome: Food & Agriculture Organization.

Davenport, J., & Bradshaw, C. (1995). Observations on skin colour changes in juvenile lumpsuckers. *Journal of fish biology*, *47*(1), 143-154. https://doi.org/10.1111/j.1095-8649.1995.tb01880.x

Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture, 208*(3-4), 191-364. https://doi.org/10.1016/S0044-8486(02)00057-1

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

Fréchet, A., Gauthier, J., Schwab, P., Beaudoin, M., Collier, F., Way, M. (2011). Assessment of lumpfish (*Cyclopterus lumpus*) in the Gulf of St. Lawrence (3Pn, 4RST) in 2010. *DFO Canadian Science Advisory Secretariat Research Document 2012/119*, 1–33.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., Pato, M. T., Petryshen, T. L., Kolonel, L. N., Lander, E. S., Sklar, P., Henderson, B., Hirschhorn, J. N., & Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics, 36*, 388-393. https://doi.org/10.1038/ng1333

Furman, B. L., Metzger, D. C., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., Shu, J. J., & Mank, J. E. (2020). Sex chromosome evolution: So many exceptions to the rules. *Genome Biology and Evolution*, *12*(6), 750-763.

Goudet, J. (2005). hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes, 5*(1), 184-186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Goulet, D., Green, J. M., & Shears, T. H. (1986). Courtship, spawning, and parental care behavior of the lumpfish, *Cyclopterus lumpus* L., in Newfoundland. *Canadian Journal of Zoology*, *64*(6), 1320-1325. https://doi.org/10.1139/z86-196

Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., Fernandino, J. I., Somoza, G. M., Yokota, M., & Strüssmann, C. A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences, 109*(8), 2955-2959. https://doi.org/10.1073/pnas.1018392109

Herpin, A., & Schartl, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports, 16*, 1260-1274. https://doi.org/10.15252/embr.201540667

Heule, C., Salzburger, W., & Böhne, A. (2014). Genetics of sexual development: an evolutionary playground for fish. *Genetics, 196*(3), 579-591. https://doi.org/10.1534/genetics.114.161158

Imsland, A. K., Reynolds, P., Eliassen, G., Hangstad, T. A., Foss, A., Vikingstad, E., & Elvegård, T. A. (2014). The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeoptheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar* L.) *Aquaculture, 424-425*, 18-23. https://doi.org/10.1016/j.aquaculture.2013.12.033

Imsland, A. K., Reynolds, P., Eliassen, G., Hangstad, T. A., Nytrø, A. V., Foss, A., Vikingstad, E., & Elvegård, T. A. (2015). Feeding preferences of lumpfish (*Cyclopterus lumpus* L.) maintained in open net-pens with Atlantic salmon (*Salmo salar* L.). *Aquaculture, 436*, 47-51. https://doi.org/10.1016/j.aquaculture.2014.10.048

Inagaki, M., Irie, K., Ishizaki, H., Tanaka-Okamoto, M., Miyoshi, J., & Takai, Y. (2006). Role of cell adhesion molecule nectin-3 in spermatid development. *Genes to Cells, 11*(9), 1125-1132. https://doi.org/10.1111/j.1365-2443.2006.01006.x

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

Kennedy, J., Jónsson, S. Þ., Kasper, J. M., & Ólafsson, H. G. (2015). Movements of female lumpfish (*Cyclopterus lumpus*) around Iceland. *ICES Journal of Marine Science, 73*, 1160-1169. https://doi.org/10.1093/icesjms/fsu170

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research, 21*, 487–93.

Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution. *Developmental Dynamics, 242*(4), 339-353. https://doi.org/10.1002/dvdy.23927

Knutsen, T. M., Kirubakaran, G. T., Mommens, M., & Moen, T. (2018) Lumpfish (*Cyclopterus lumpus*) draft genome assembly. Trondheim: Aquagen AS. https://doi.org/10.6084/m9.figshare.7301546.v4

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. *Nature, 351*(6322), 117-121. https://doi.org/10.1038/351117a0

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li, M. F., & Clyburne, S. (1977). New cell line from the marine lumpfish, *Cyclopterus lumpus*. *Journal of the Fisheries Research Board of Canada, 34*(1), 134-139. https://doi.org/10.1139/f77-017

Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., & Nagahama, Y. (2015). A tandem duplicate of anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLoS Genetics, 11*(11), e1005678. https://doi.org/10.1371/journal.pgen.1005678

Lubieniecki, K. P., Lin, S. L., Cabana, E. I., Li, J., Lai, Y. Y. Y., & Davidson, W. S. (2015). Genomic instability of the sex-determing locus in Atlantic salmon (*Salmo salar*). *G3: Genes, Genomes, Genetics, 5*(11), 2513-2522. https://doi.org/10.1534/g3.115.020115

Luckenbach, J. A., Fairgrieve, W. T., & Hayman, E. S. (2017). Establishment of monosex female production of sablefish (*Anoplopoma fimbria*) through direct and indirect sex control. *Aquaculture, 479,* 285-296. https://doi.org/10.1016/j.aquaculture.2017.05.037

Luu, K., Bazin, E., & Blum, M.G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources, 17*(1), 67-77. https://doi.org/10.1111/1755-0998.12592

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867-2873. https://doi.org/10.1093/bioinformatics/btq559

Mank, J. E., Promislow, D. E. L., & Avise, J. C. (2006). Evolution of alterative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Scoiety, 87*(1), 83-93. https://doi.org/10.1111/j.1095-8312.2006.00558.x

Martin-Robichaud, D. J., Peterson, R. H., Benfey, T. J, & Crim, L. W. (1994). Direct feminization of lumpfish (feminization of lumpfish (*Cyclopterus lumpus* L.) using 17β-oestradiol-enriched *Artermia* as food. *Aquaculture, 123*(1-2), 137-151. https://doi.org/10.1016/0044-8486(94)90126-0

Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes, *Molecular Biology and Evolution, 38*(10), 4647–4654.

Matsuda, M., & Sakaizumi, M. (2016). Evolution of the sex-determing gene in the teleostean genus *Oryzias*. *General and Comparative Endocrinology, 239,* 80-88. https://doi.org/10.1016/j.ygcen.2015.10.004

Milet, J., Courtin, D., Garcia, A., & Perdry, H. (2020). Mixed logistic regression in genome-wide association studies. *BMC Bioinformatics, 21,* 536. https://doi.org/10.1186/s12859-020-03862-2

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamagucji, S., & Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determing gene in medaka, *Oryzias luzonensis*. *Genetics, 191*(1), 163-170. https://doi.org/10.1534/genetics.111.137497

Myosho, T., Takehana, Y., Hamaguchi, S., & Sakaizumo, M. (2015). Turnover of sex chromosomes in *Celebensis* group medaka fishes. *G3: Genes, Genomes, Genetics, 5*(12), 2685-2691. https://doi.org/10.1534/g3.115.021543

Nakamoto, M., Uchino, T., Koshimizu, E., Kuchiishi, Y., Sekiguchi, R., Wang, L., … & Sakamoto, T. (2021). A Y-linked anti-Mullerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLOS Genetics, 17*(8), e1009705. https://doi.org/10.1371/journal.pgen.1009705

Palmer, D. H., Rogers, T. F., Dean, R., Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology, 28*(21), 4709-4724. https://doi.org/10.1111/mec.15245

Pan, Q., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., ... & Guiguen, Y. (2021). The rise and fall of the ancient northern pike master sex-determining gene. *Elife, 10*, e62858. https://doi.org/10.7554/eLife.62858

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

in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genetics*, *15*(8), e1008013. https://doi.org/10.1371/journal.pgen.1008013

Peichel, C.L., McCann, S.R., Ross, J.A., Naftaly, A.F.S., Urton, F.R., Cech, J.N., … & White, M.A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, *21*(177). https://doi.org/10.1186/s13059-020-02097-x

Pennell, M. W., Mank, J. E., & Peichel, C. L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Molecular Ecology, 27*(19), 3950-3963. https://doi.org/10.1111/mec.14540

Powell, A., Treasurer, J. W., Pooley, C. L., Keay, A. J., Lloyd, R., Imsland, A. K., & Garcia de Leaniz, C. (2018). Use of lumpfish for sea-lice control in salmon farming: challenges and opportunities. *Reviews in Aquaculture, 10*(3), 683-702. https://doi.org/10.1111/raq.12194

Quinn, A. E., Sarre, S. D., Ezaz, T., Marshall Graves, J. A., & Georges, A. (2011). Evolutionary transitions between mechanisms of sex determination in vertebrates. *Biology Letters*, *7*(3), 443-448. https://doi.org/10.1098/rsbl.2010.1126

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rambaut, A. (2018). FigTree version 1.4.4. Computer program and documentation distributed by the author. http://tree.bio.ed.ac.uk/software/figtree/

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... & Cantin, L. J. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature, 592*, 737-746. https://doi.org/10.1038/s41586-021-03451-0

Rondeau, E. B., Laurie, C. V., Johnson, S. C., & Koop, B. F. (2016). A PCR assay detects a male-specific duplicated copy of Anti-Müllerian hormone (amh) in the lingcod (*Ophiodon elongatus*). *BMC Research Notes, 9*, 230. https://doi.org/10.1186/s13104-016-2030-6

Rondeau, E. B., Messmer, A. M., Sanderson, D. S., Jantzen, S. G., von Schalburg, K. R., Minkley, D. R., Leong, J. S., Macdonald, G. M., Davidsen, A. E., Parker, W. A., Mazzola, R. S. A., Campbell, B., & Koop, B. F. (2013). Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny, linkage map and indentification of a putative sex gene. *BMC Genomics, 14,* 452. https://doi.org/10.1186/1471-2164-14-452

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Sarre, S. D., Ezaz, T., & Georges, A. (2011). Transitions between sex-determining systems in reptiles and amphibians. *Annual Review of Genomics and Human Genetics, 12*, 391-406. https://doi.org/10.1146/annurev-genom-082410-101518

Schartl, M. (2004). Sex chromosome evolution in non-mammalian vertebrates. *Current Opinion in Genetics and Development, 14*(6), 634-641. https://doi.org/10.1016/j.gde.2004.09.005

Smith, C. A., McClive, P. J., Western, P. S., Reed, K. J., & Sinclair, A. H. (1999). Conservation of a sex-determining gene. *Nature, 402*, 601-602. https://doi.org/10.1038/45130

Smith, W. L., & Busby, M. S. (2014). Phylogeny and taxonomy of sculpins, sandfishes, and snailfishes (Perciformes: Cottoidei) with comments on the phylogenetic significance of their early-life history specializations. *Molecular Phylogenetics and Evolution, 79,* 332-352. https://doi.org/10.1016/j.ympev.2014.06.028

826 Smith, W. L., Everman, E., & Richardson, C. (2018). Phylogeny and taxonomy of flatheads,
827 scorpionfishes, sea robins, and stonefishes (Perceomorpha: Scorpaeniformes) and the evolution of the
828 lachrymal saber. *Copeia, 106*(1), 94-119. https://doi.org/10.1643/CG-17-669
829
830 Souvorov, A., Kapustin, Y., Kiryutin, B., Chetvernin, V., Tatusova, T., & Lipman, D. (2010).
831 Gnomon–NCBI eukaryotic gene prediction tool. *National Center for Biotechnology Information*,
832 http://www.ncbi.nlm.nih.gov/core/assets/genome/files/Gnomon-description.pdf
833
834 Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for
835 macOS. *Molecular Biology and Evolution, 37*(4), 1237-1239. https://doi.org/10.1093/molbev/msz312
836
837 The Tree of Sex Consortium. (2014). Tree of Sex: A database of sexual systems. *Scientific Data*, *1*,
838 140015. https://doi.org/10.1038/sdata.2014.15
839 Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of
840 progressive multiple sequence alignment through sequence weighting, position-specific gap penalties
841 and weight matrix choice. *Nucleic Acids Research, 22*(22), 4673-4680.
842 https://doi.org/10.1093/nar/22.22.4673
843
844 Traut, W., & Willhoeft, U. (1990). A jumping sex determining factor in the fly *Megaselia scalaris*.
845 *Chromosoma, 99*(6), 407-412. https://doi.org/10.1007/BF01726692
846
847 Van Doorn, G. S., & Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual
848 conflict. *Nature*, *449*(7164), 909-912.
849
850 Voight, B. F., & Pritchard, J. K. (2005). Confounding from cryptic relatedness in case-control
851 association studies. *PLOS Genetics, 1*(3), e32. https://doi.org/10.1371/journal.pgen.0010032
852

Volff, J.-N., Nanda, I., Schmid, M., & Schartl, M. (2007). Governing sex determination in fish: Regulatory putsches and ephemeral dictators. *Sexual Development, 1*(2), 85-99. https://doi.org/10.1159/000100030

Voskoboinikova, O. S., Kudryavtseva, O. Y., Orlov, A. M., Orlova, S. Y., Nazarkin, M. V., Chernova, N. V., & Maznikova, O. A. (2020). Relationships and evolution of lumpsuckers of the family Cyclopteridae (Cottoidei). *Journal of Ichthyology*, *60*, 154-181.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., ... & Flicek, P. (2020) Ensembl 2020. *Nucleic Acids Research, 48*(D1), D682–D688. https://doi.org/10.1093/nar/gkz966

Zarkower, D. (2001). Establishing sexual dimorphism: conservation amidst diversity? *Nature Reviews Genetics, 2,* 175-185. https://doi.org/10.1038/35056032

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology, 7*(1-2), 203-214. https://doi.org/10.1089/10665270050081478

Tables

Table 1. Fin clip sample origin and phenotypic sex composition. Ocean Science Centre samples are samples from a Lumpfish breeding program and Gulf of St. Lawrence are wild caught samples.

		Sex	
Site	Year	Male	Female
Ocean Science Centre	2017	40	12
Gulf of St. Lawrence	2018	29	15
Gulf of St. Lawrence	2019	27	16
Total		96	43

Table 2. Genome wide significant SNPs associated with the phenotypic sex of Lumpfish after false discovery rate correction for multiple testing. Effect size of the SNP (odds ratio) and standard deviation (SD) represent the effect of the SNP on the sex phenotype.

SNP	Chromosome	Position (bp)	*p* value	n	A_1	A_2	FreqA_2	Effect size (SD)
AX-298030667	13	13,583,153	2.07081×10^{-9}	139	A	G	0.608	-4.942 (0.825)

AX-297991767	13	13,636,087	2.71429×10^{-9}	139	G	A	0.673	-2.315 (0.389)
AX-297994078	13	13,645,411	1.96401×10^{-8}	139	C	A	0.687	-2.200 (0.392)
AX-298211598	13	12,669,197	3.74394×10^{-7}	138	T	C	0.500	1.719 (0.338)
AX-297903115	13	8,582,071	2.76546×10^{-6}	139	A	G	0.748	-2.089 (0.446)
AX-298053566	21	174,166	2.83182×10^{-6}	139	C	A	0.626	1.427 (0.305)
AX-297876558	13	15,209,092	3.32226×10^{-6}	139	T	C	0.644	-1.635 (0.352)
AX-297903099	13	8,570,769	5.13×10^{-6}	139	A	G	0.748	-2.021 (0.443)

882 n represents the number of fish genotyped for each SNP

883 FreqA_2 represents the allele frequency of the second allele (A_2)

Table 3. Genes located within a 100 kbp region of the most significant SNP on each chromosome.

Chromosome	GeneID	Gene symbol	Start	End	Distance to most significant SNP (kpb)	Gene description
13	117741609	LOC117741609	13495680	13502874	80.3	sodium channel subunit beta-4-like
13	117741608	scn2b	13515336	13522945	60.2	sodium channel, voltage-gated, type II, beta
13	117741803	LOC117741803	13592786	13594515	9.63	uncharacterized LOC117741803*
13	117741289	LOC117741289	13596840	13600947	13.7	AMH: Müllerian -inhibiting factor-like
21	117750319	cryl1	55320	84102	90.1	crystallin, lambda 1
21	117750956	cx30.3	86386	89204	85.0	connexin 30.3
21	117750303	gja3	97736	99130	75.0	gap junction protein, alpha 3
21	117750172	zmym2	112985	148121	26.0	zinc finger, MYM-type 2
21	117750173	mphosph8	153290	200110	0	M-phase phosphoprotein 8
21	117750852	ttf2	201029	212517	26.9	transcription termination factor, RNA polymerase II
21	117750264	LOC117750264	213491	225731	39.3	centromere protein J-like

21	117750456	rnf17	226007	250227	51.8	ring finger protein 17

* This gene is undescribed but has sequence similarity to AMH in other teleost species

Figures

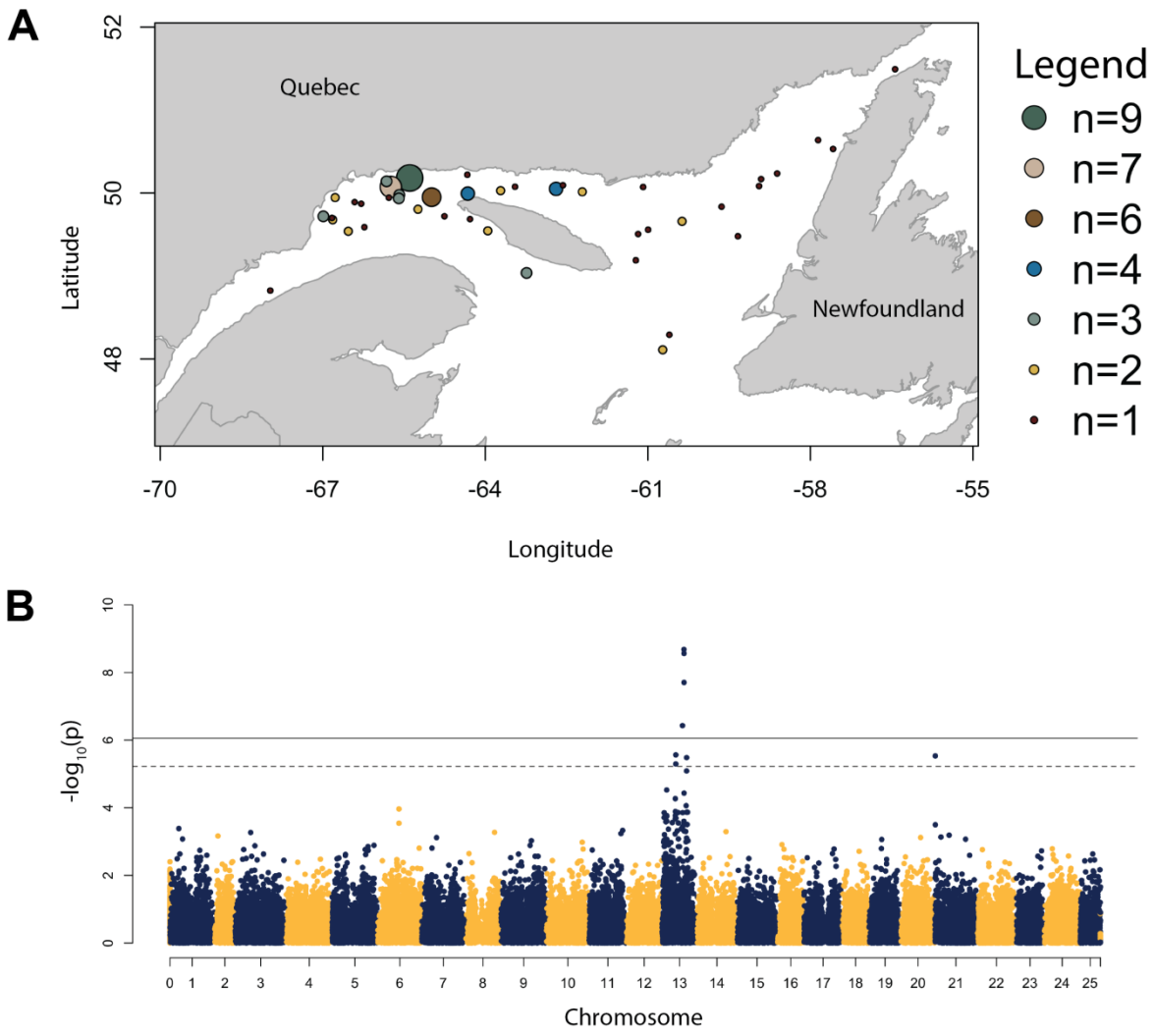

Figure 1. (A) Unique sampling locations of wild Lumpfish from Fisheries and Oceans Canada trawl surveys in the Gulf of St. Lawrence in late summer of 2018 and 2019. Point size and colour correspond to the number of Lumpfish sampled at each coordinate. (B) Manhattan plot of SNP markers associated with sex in Lumpfish, *Cyclopterus lumpus*, determined via genome wide association analysis and SNP genome position. The solid horizontal line represents the Bonferroni corrected significance threshold while the dashed line represents the false discovery rate significance threshold.

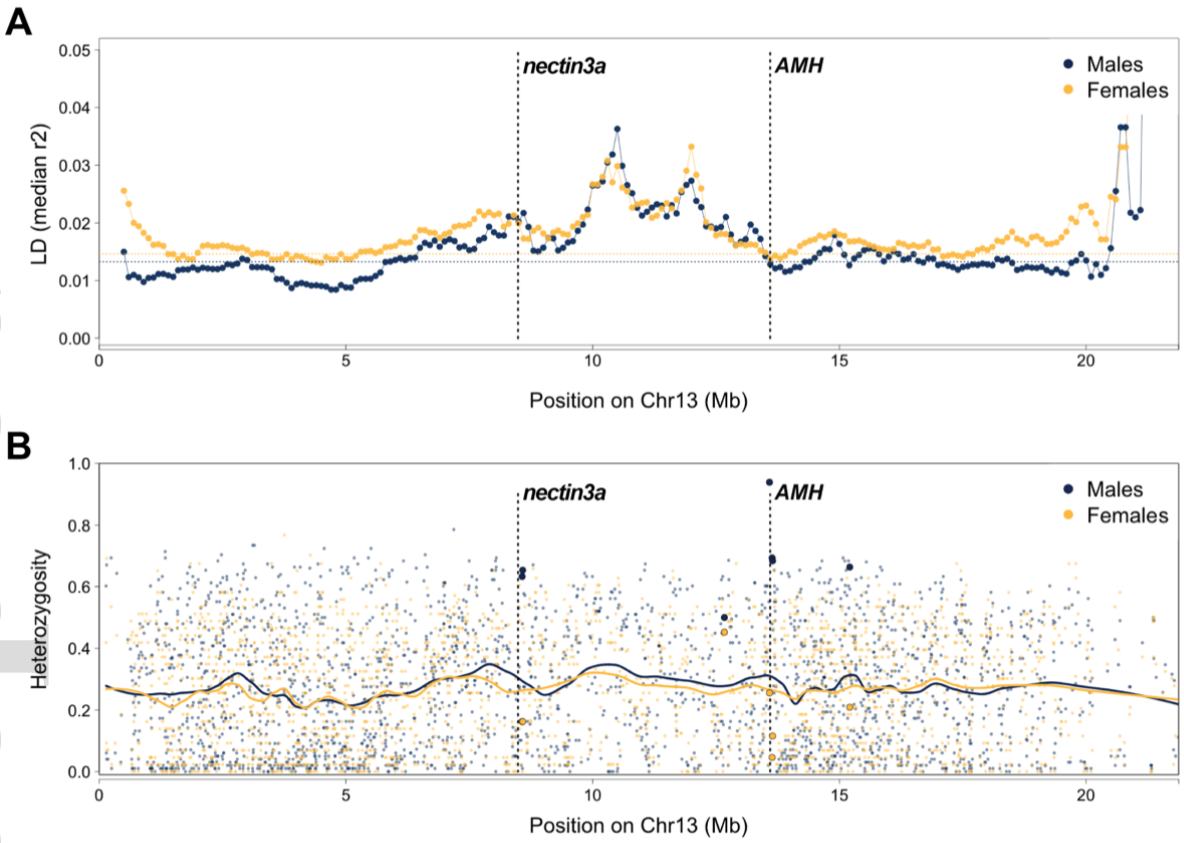

Figure 2. Sex-specific statistics for Lumpfish males (blue) and females (yellow) along chromosome 13, the putative sex chromosome, with the location of anti- Müllerian hormone (AMH), the putative sex determining factor, and *nectin3a*, shown with vertical black dashed lines. (A) Linkage

disequilibrium (LD) across chromosome 13. LD was calculated in sliding windows of 1 megabase pairs size. (B) Heterozygosity for SNPs (circles) located on chromosome 13. The enlarged SNPs represent the significant SNPs on chromosome 13.

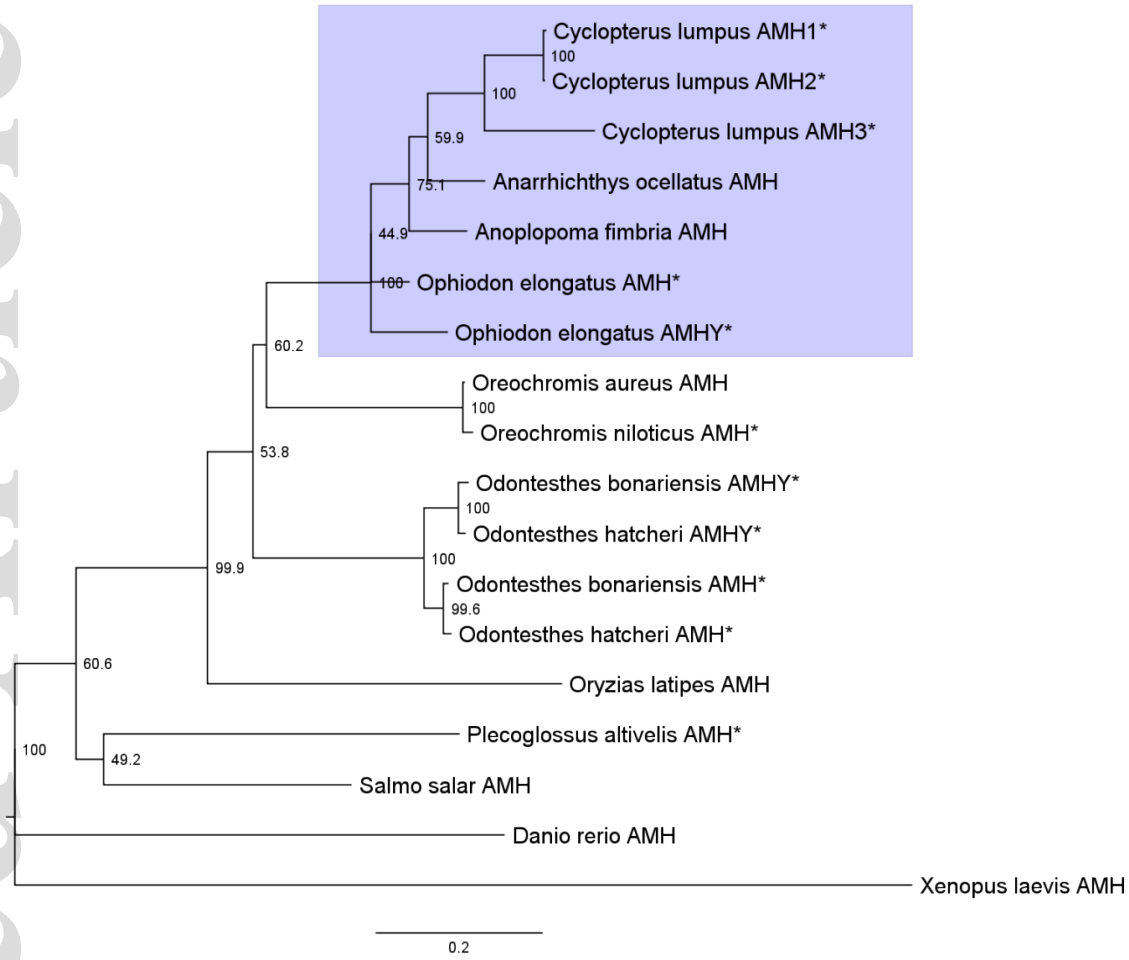

Figure 3. Phylogenetic relationship of AMH and AMHY gene sequences constructed using the Neighbor-Joining method after multiple alignment with ClustalW. Numbers indicate bootstrap values based on 1000 replicates. The purple box represents AMH/AMHY sequences from species within the Cottioidei suborder. Asterisks represent species in which AMH is the master sex determining gene. Scale bar represents 0.04 nucleotide substitutions per site.