***Phil. Trans. R. Soc. B.* article template**

PHILOSOPHICAL TRANSACTIONS B

Phil. Trans. R. Soc. B.
doi:10.1098/not yet assigned

A 180 My-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes

Heiner Kuhl[1], Yann Guiguen[2], Christin Höhne[1], Eva Kreuz[1], Kang Du[3, 14], Christophe Klopp[4], Céline Lopez-Roques[5], Elena Santidrian Yebra-Pimentel[6], Mitica Ciorpac[7,8], Jörn Gessner[1], Daniela Holostenco[7], Wibke Kleiner[1], Klaus Kohlmann[1], Dunja K. Lamatsch[9], Dmitry Prokopov[10], Anastasia Bestin[11], Emmanuel Bonpunt[12], Bastien Debeuf[13], Pierrick Haffray[11], Romain Morvezen[11], Pierre Patrice[11], Radu Suciu[7], Ron Dirks[6], Sven Wuertz[1], Werner Kloas[1], Manfred Schartl[3, 14], Matthias Stöck[1, *]

[1] *Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Department of Ecophysiology and Aquaculture, Müggelseedamm 301 & 310, D-12587 Berlin, Germany, e-mails: kuhl@igb-berlin.de , hoehne@igb-berlin.de , kreuz@igb-berlin.de , sturgeon@igb-berlin.de , werner.kloas@igb-berlin.de , wibke.kleiner@igb-berlin.de , kohlmann@igb-berlin.de , wuertz@igb-berlin.de*
[2] *INRAE, LPGP, 35000, Rennes, France., e-mail: yann.guiguen@inrae.fr*
[3] *Developmental Biochemistry, Biocenter, University of Wuerzburg, Wuerzburg, Germany, email: dukang1117@outlook.com , phch1@biozentrum.uni-wuerzburg.de*
[4] *SIGENAE, Plate-forme Bio-informatique Genotoul, Mathématiques et Informatique Appliquées de Toulouse, INRAe, Castanet-Tolosan, France, email: christophe.klopp@inrae.fr*
[5] *INRA, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France, e-mail: celine.lopez-roques@inrae.fr*
[6] *Future Genomics Technologies B.V., Sylviusweg 74, 2333 BD Leiden, The Netherlands, e-mails: dirks@futuregenomics.tech , santidrian@futuregenomics.tech*
[7] *Danube Delta National Institute for Research and Development, Tulcea 820112, Romania, e-mail: dana.holostenco@gmail.com , radu.suciu34@gmail.com*
[8] *Genetic Improvement Laboratory, Research Station for Cattle Breeding Dancu - Iasi (SCDCB Dancu), Academy of Agricultural and Forestry Sciences "Gheorghe Ionescu-Sisesti", Romania (Mail address: Iasi-Ungheni Street, No. 9, Holboca, Iaşi county, 707252, Romania), e-mail: ciorpac.mitica@gmail.com*
[9] *University of Innsbruck, Research Department for Limnology, Mondsee, Austria, e-mail: dunja.lamatsch@uibk.ac.at*
[10] *Institute of Molecular and Cellular Biology, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia, e-mail: dprokopov@mcb.nsc.ru*
[11] *SYSAAF, Station INRAE-LPGP, Campus de Beaulieu, 35042, Rennes cedex, France, e-mails: anastasia.bestin@inrae.fr, pierrick.haffray@inrae.fr, Romain.Morvezen@inrae.fr, Pierre.Patrice@inrae.fr*
[12] *L'Esturgeonnière, Allée de Mios, 33470 Le Teich, France, e-mail: emmanuelb@caviar-perlita.com*
[13] *SCEA Sturgeon, 29 rue du Carillon, 17240 Saint Fort sur Gironde, France, e-mail: b.debeuf@kaviar.com*
[14] *The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, TX, USA.*

Keywords: Acipenseridae, sturgeon, sex chromosomes, female-specific, polyploidy, evolution.

*Author for correspondence (matthias.stoeck@igb-berlin.de).
†Present address: Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Department of Ecophysiology and Aquaculture, Müggelseedamm 301, D-12587 Berlin, Germany.

Summary

Several hypotheses explain the prevalence of undifferentiated sex chromosomes in poikilothermic vertebrates. Turnovers change the master sex determination gene, the sex chromosome or the sex determination system (e.g. XY to WZ). Jumping master genes stay main triggers but translocate to other chromosomes. Occasional recombination (e.g. in sex-reversed females) prevents sex chromosome degeneration. Recent research has uncovered conserved heteromorphic or even homomorphic sex chromosomes in several clades of non-avian and non-mammalian vertebrates. Sex determination in sturgeons (Acipenseridae) has been a long-standing basic biological question, linked to economical demands by the caviar-producing aquaculture. Here, we report the discovery of a sex-specific sequence from sterlet (*Acipenser ruthenus*). Using chromosome-scale assemblies and pool-sequencing, we first identified a ~16 kb female-specific region. We developed a PCR-genotyping test, yielding female-specific products in six species, spanning the entire phylogeny with the most divergent extant lineages (*A. sturio, A. oxyrinchus* vs. *A. ruthenus, Huso huso*), stemming from an ancient tetraploidization. Similar results were obtained in two octoploid species (*A. gueldenstaedtii, A. baerii*). Conservation of a female-specific sequence for a long period, representing 180 My of sturgeon evolution, and across at least one polyploidization event, raises many interesting biological questions. We discuss a conserved undifferentiated sex chromosome system with a ZZ/ZW-mode of sex determination and potential alternatives.

Introduction

Sexual reproduction is an old feature of life. In vertebrates, sexual development is determined by environmental triggers (Environmental Sex determination, ESD) or genotypic sex determination (GSD) (e.g.[1]) or a combination thereof. In an existing GSD-system, the paradigmatic evolutionary model of sex chromosome evolution suggests the rise of a new sex chromosome from an autosome by mutation or translocation of a gene, which becomes a trigger of the sex-determining regulatory networks. Heterozygotes encode one sex and homozygotes the other [2-4]. According to this classical view, mutations and/or linkage to genes with sexually antagonistic effects may be favoured near this gene, i.e. mainly in unpaired parts of Y and W chromosomes, profiting from linkage disequilibrium. To preserve epistasis, recombination between such mutations and the sex-determining locus would then be suppressed in the heterogametic sex [5-6]. Such nonrecombining sex chromosomes are expected to degenerate through the progressive accumulation of deleterious mutations [7-8], except for small pseudo-autosomal regions. Recently, as an alternative or additional cause of degeneration, instability and divergence of *cis*-regulatory sequences in non-recombining genome regions, which become selectively haploidised to mask deleterious mutations on coding sequences, have been suggested [9].

In contrast to the conservation of differentiated sex chromosomes in most mammals and birds, in the majority of poikilothermic vertebrates differentiated sex chromosomes are rarely observed [10-12] (but see [13]). Several hypotheses have been offered to explain the prevalence of undifferentiated sex chromosomes. The high-turnover model proposes that the emergence of new master sex-determining genes on new chromosomes would prevent sex chromosome degeneration [14-17]. The "jumping sex locus"-hypothesis [18] implies that the sex determination gene is maintained but changes its position to different chromosomes, preventing evolutionary decay of the original sex chromosome [19]. The 'fountain of youth' hypothesis [20] suggests that sex chromosomes may escape from degeneration by occasional recombination Because sex-specific recombination depends on phenotypic rather than genotypic sex, homomorphic X and Y chromosomes might recombine in sex-reversed females. These rare events should generate bursts of new Y haplotypes, which will be quickly sorted out by natural or sexual selection [20]. Seemingly young sex chromosomes may thus carry old-established sex-determining genes, challenging the opinion that sex chromosomes unavoidably decay [21-22].

Beyond mammals and birds, conserved sex chromosomes have recently been discovered in several amniote (specifically reptile) clades [23-26], all of which feature differentiated sex chromosomes. Evolutionarily very old, conserved and homomorphic ZZ/ZW sex chromosomes are known in some ratite birds (Ratidae), dating back >130 My [27-28]. Similarly, skinks (Scincidae) share homologous, mostly poorly differentiated XX/XY sex chromosomes across a wide phylogenetic spectrum for at least 85 million years [29]. Recent findings in the teleost family Esocidae report undifferentiated sex chromosomes of similar evolutionary age (65-90 My) in teleosts [30]. The Salmonidae family, which experienced a whole genome duplication ca. 90 Mya [31], presumably harbours a conserved sex determination gene, perhaps as old as 50 My, on different chromosomes in different species [18-19], which encodes a homomorphic XX/XY sex chromosome system. Here we report the discovery of a sex-specific region in sturgeon (Acipenseridae), which is preserved in many extant species with a common ancestor about 180 Mya.

Sturgeons and paddlefish comprise 27 living species [32], which branched off 330 Mya [33], from the root of the more than 31,000 living teleosts. The evolutionary history of vertebrates is characterized by several rounds of whole genome duplications (WGD). The 1R and 2R WGD happened early on, before the jawed vertebrates appeared, but the sturgeon branch diverged before the teleost 3R WGD [34-35]. Later on, sturgeon experienced several clade-specific polyploidizations, resulting in chromosome numbers from ~120 up to ~380 [36]. Based on a chromosome-scale genome assembly of *Acipenser ruthenus*, Du *et al.* [33] have recently shown a WGS, early in the sturgeon radiation. Although subsequent rediploidization of the genome caused the loss of entire chromosomes or large fragments (i.e. segmental deduplication), structural and functional tetraploidy was maintained over 180 million years, and a decrease of redundancy in the tetraploid genome follows mostly random processes. Sturgeons possess undifferentiated microscopically indistinguishable sex chromosomes [33,37]. Although the sex ratio of offspring from experimental gynogenesis led to the

assumption that all sturgeons possess female heterogametic (ZZ/ZW) sex chromosome systems [38], a sex-specific marker, confirming female heterogamety and allowing genetic studies, has not been found [33].

Materials and Methods

Animal sampling and establishment of genetic families

Acipenser ruthenus: For pool-sequencing (pool-seq) 61 (31 female, 30 male) adult sterlets from the broodstock at the Leibniz-IGB (Berlin, Germany), representing the Danube River population [33], were sexed by gonadal biopsies, evidence from reproduction and a few autopsies. In addition, 52 sterlets (table 1) were sampled from three aquaculture populations. These came from two other *A. ruthenus* brood stocks, originating from the Danube (Wöllershof, Bavaria, Germany, 20-25 cm, ca. 1 year old; and Beucha, Saxony, Germany, 10-12 cm, 6 months old), as well as *A. ruthenus* interpopulation hybrids from the Volga and Danube Rivers (12-20 cm, 8-10 months). They were euthanized and gonadal sex determined histologically, and finclips used for genotyping. In spring 2020, a genetic family was derived from the IGB-broodstock; at the age of 4 months, histological sections of gonads from 14 offspring were prepared, and finclips from them and their parents also used for genetics.

Huso huso: Using minimal invasive biopsy samples from the gonads of 28 subadult and adult beluga sturgeon from two fish farms (Horia and Peceneaga, Romania, fish between 10 and 13 years old) were inspected by the unaided eye and/or microscopically after standard hematoxylin/eosin staining (described below). Finclips were sampled for genetic analyses (table 1). To generate genetic families, on April 6, 2018, a beluga sturgeon male (PIT-tag 000004691858; 200 cm, ~100 kg) was captured in the Danube delta region at river-Km 126 (distance from the Black Sea) in Romania. On April 13, 2018, a mature *H. huso* female (PIT 642099000568452; TL 260 cm, ~150 kg, near Km 100), and another male on April, 14, 2018 (PIT 968000004692054; TL 143 cm, ~22 kg, Km 126) were caught. All three were transferred to an aquaculture company (Horia, Romania). After stimulation of all the three wild-caught animals and one additional captive male with gonadotropin-releasing hormone (GnRH), ovulation started on April 16. Eggs were fertilized with sperm of each male separately. Fin clips of the four adult fish were taken, and the offspring of the individual matings were reared separately until the age of one year. All offspring fish were individually tagged with PIT-tags and then reared in common tanks. For this study, fish of one genetic family were sampled at the age of 15-24 months for gonadal histology and finclips. Wild-caught adults and, after two years, hundreds of their tagged offspring were released into the Danube to minimize the impact on this highly threatened species.

A. oxyrinchus: Tissue samples of 27 adult Baltic sturgeon (table 1) from the broodstock in Mecklenburg-Pomerania (Born, Germany), kept for re-introduction into Oder River and Baltic Sea, were taken from tagged males and females, sexed using gonadal biopsies and evidence from reproduction. In 2017, genetic families were derived from breeding pairs during the yearly reproduction. Finclips from the parental animals were kept in ethanol. A subset of juveniles was raised in separate tanks at the Leibniz-IGB (Berlin, Germany) for three years.

A. sturio: Tissue samples of 36 adult European sturgeon (table 1) from a broodstock of immature fish at the IGB, raised for conservation aquaculture to re-introduce the species into the Elbe River were initially sexed using ultrasonic evidence. Gonadal biopsies were taken if genotyping did not match the ultrasonic evidence. Few specimens came from autopsy of deceased fish.

A. baerii: Finclips of 56 adult Siberian sturgeon (table 1) used as breeders with known phenotypic sex were sampled in an aquaculture breeding company (l'Esturgeonnière, France) and kept in 100% ethanol until DNA extraction.

A. gueldenstaedtii: Finclips of 27 adult Russian sturgeons (table 1) were sampled in an aquaculture breeding company (Ecloserie de Guyenne, SCEA Sturgeon, France) and kept in 100% ethanol until DNA extraction. All fish were sexed based on evidence from ultrasound, reproduction and autopsy.

DNA processing for genotyping

DNA from finclips or other tissue samples was extracted using the DNeasy Tissue Kit (Qiagen) or the BioSprint robotic workstation with the 96 DNA Plant Kit (Qiagen, Germany), according to the manufacturer's protocols.

Histological sexing of juvenile and subadult fish as well as offspring of genetic families

Offspring from genetic families of *H. huso, A. ruthenus,* and *A. oxyrinchus* was sacrificed by an overdose of buffered tricaine methanesulfonate (MS222; 0.3 g/L, PHARMAQ), individual length and wet weight were recorded, finclips (stored in 100% ethanol) and gonadal samples were taken immediately. The right gonad strand was fixed in phosphate-buffered formaldehyde solution (ROTI®Histofix, 4%, Carl Roth) for at least 24 h at room temperature, then rinsed three times for 24 h with 70% ethanol, and stored at room temperature. Gonad strands were dehydrated in an increasing series of ethanol (75%, 90%, 95%, 100%), rinsed in xylol (Carl Roth), and transferred into Paraplast® (Leica), using the Shandon Excelsior ES Tissue Processor (Thermo Fisher Scientific). Gonads were embedded in a spiral-like arrangement, to ideally cut the whole organ into 5 µm sections across its entire length (microtomes: Leica 2065 or Microm HM 340E, Thermo Fisher Scientific). Sections were mounted on glass slides, and stained using standard hematoxylin/eosin (0.1%, Carl Roth). Histological evaluation was made at various magnifications using light microscope Nikon Eclipse Ni-U with Nikon DS-Fi3 camera, and the corresponding software Nikon DS-L4 to archive images.

For Review Only

Pool-sequencing

The genomic DNA of 31 female and 30 male *A. ruthenus* specimens was extracted from 90% ethanol-preserved fin clips using a classical phenol/chloroform protocol, quantified using Qubit fluorimetry and analysed using the Fragment Analyzer (Advanced Analytical Technologies, Inc., Iowa, USA). DNA shorts read sequencing was performed at the GeT-PlaGe core facility, INRAe Toulouse (http://www.get.genotoul.fr). Two DNA pool-seq libraries were prepared according to manufacturer's protocols using the Illumina TruSeq Nano DNA HT Library Prep Kit (Illumina, California, USA). Briefly, from 200 ng of each sample, DNA was fragmented (550 bp) by sonication on a M220 Focused-ultrasonicator (COVARIS). Size selection was performed using SPB beads (kit beads) and 3'-ends of the blunt fragments were mono-adenylated. Then, adaptors and indexes were ligated and the construction amplified with Illumina-specific primers (8 cycles). Library-quality was assessed using Fragment Analyzer and libraries were quantified by qPCR using the Kapa Library Quantification Kit (Roche). Sequencing was performed on a NovaSeq S4 lane (Illumina, California, USA), using a paired-end read length of 2x150 bp with the Illumina NovaSeq Reagent Kits.

Analysis of sex specific variants based on a male reference genome

Pool-seq data were mapped to the male *A. ruthenus* reference genome (*Acipenser_ruthenus*.ARUT1.2.dna.toplevel.fa, which is nearly identical to the current reference genome at NCBI GCF_010645085.1; electronic supplementary material, table S1) by Minimap2 (with parameters –t 12 –x sr –a; [39]) and converted to sorted bam-files by samtools [40]. Variants (SNPs, MNPs, INDELs) were called from both bam files using the Platypus variant caller [41] with integrated duplicate fragment reads removal (--filterDuplicates=1) and reassembly of reads (--assembleAll=1). The resulting vcf-file was screened for sex-specific variants using awk-scripting and different variant read coverage cutoffs (i.e. maximum reads in males (0, 1 …), minimum reads in females (10, 13 …) or *vice versa*). Variants were clustered according to their distance in the genome (maximal distance of variants in a cluster: 4000 bp), counted using bedtools (merge, annotate; [42]) and sorted by variant count. Regions with highest variant densities and neighbouring regions were inspected for genes related to sex determination.

Reconstruction of male and female specific genomic regions from unmapped pool-seq data

Unmapped reads potentially representing regions not present in the reference genome were extracted from bam-files and assembled using the IDBA assembler [43]. The resulting male/female specific contigs were added to the reference genome and pool-seq data was mapped again (as described above) to this extended reference genome. Finally, sequencing coverage by female and male pools was compared in bedgraph files derived from the bam-files (bedtools genomecov –bga –split –ibam … ; bedtools unionbedgraph). These were screened for putative haploid coverage in one sex and no or low, spurious coverage in the other sex to identify female/male specific regions.

Another high-quality assembly of a female *A. ruthenus* has recently been published at NCBI by the Vertebrate Genomes Project (GCA_902713425.1). We repeated our analyses from above using this reference with an emphasis on detection of coverage differences. Using bedgraph-files we compared female and male pool-sequencing coverage at the bp-level along the genome. We used different filtering approaches to call regions which had no or low coverage in one sex and approximately haploid coverage in the other sex. The most successful filtering applied awk-scripting to filter regions in male and female genomes where 0 coverage in male pool and a coverage of larger than 0.3 in the female pool was observed (here 1.0 was the normalized diploid coverage). These regions were combined in larger blocks, if they were as close as 30 bp to each other (bedtools merge). Finally, we filtered for blocks that were larger than 160 bp and derived at a clear signal present only in female data and female genome.

Primer design and PCR conditions

Using the male *vs*. female *A. ruthenus* genome sequence as well as preliminary Illumina-data of a female *A. oxyrinchus* genome, we designed a number of primers, which were primarily tested in gradient-PCRs on phenotypically sexed *A. ruthenus*. While several primers showed sex-specific amplification in this species, one (*AllWSex2*) could be optimized to work across multiple species (primer sequences and PCR-conditions: electronic supplementary material, texts S1-S3).

Phylogenomic reconstruction

To build a phylogenetic tree for sturgeons, we download transcriptome assemblies of *Acipenser baerii*, *A. oxyrinchus*, *A. schrencki*, *A. sinensis* and *A. transmontanus* from the Public Sturgeon Transcripts Database (http://publicsturgeon.sigenae.org/home.html) and of *Acipenser gueldenstaedtii* from NCBI (https://www.ncbi.nlm.nih.gov/Traces/wgs/?page=1&view=tsa&search=acipenser). The transcriptome of *Huso huso* was *de novo* assembled using Trinity [44] from mixed organs. Annotated genomes of *A. ruthenus* and *Latimeria chalumnae* were also downloaded from NCBI (electronic supplementary material, table S2). To identify orthologs between two species, we used the "Reciprocal best hit (RBH)" method, and *A. ruthenus* as the hub to retrieve orthologous genes across all species. Protein sequences of the transcriptome contigs were predicted using GeneWise [45] by aligning each contig to its orthologous protein in *A. ruthenus*. Contigs with less than 60% of the protein aligned were discarded. In total, we collected 1017 genes with orthology across all species. For each orthologous gene, the protein sequences were aligned across all species using MAFFT [46], and trimmed them, using trimAl [47]. All 1017 alignments were then combined in a concatenated alignment, consisting of 383506 aligned amino acid sites. Finally, this alignment was transferred to RAxML [48] to construct a Maximum Likelihood-phylogeny.

For Review Only

Results

Analyses of sterlet (A. ruthenus) female and male pool-sequencing

Pool-sequencing resulted in 553,117,774 reads (83.5 Gbp) and 486,762,084 reads (73.5 Gb) for the female and the male *A. ruthenus* pools, respectively. Pool-sequencing Illumina reads are available in the Sequence Read Archive (SRA: SRX9341540; SRX9341541), under BioProject reference PRJNA670610. After mapping to the reference genomes, the peak of the unimodal coverage distribution was 42x for the female pool and 37x for the male pool. There was no significant difference in the percentage of mapped reads for both pools (96.9% reads mapped).

We initially used two bioinformatic strategies to uncover female-specific pool-seq data in *A. ruthenus*. We first analysed sex-specific variants based on a male reference genome. The number of female or male specific variants was low, but their ratio already hinted at a ZW sex determination system (details: electronic supplementary material, text S4). In the top candidate region, which harboured 27 female-specific variants at HiC_scaffold_4:45,149,338-45,165,057 (electronic supplementary material, figure S1, table S1, text S5), we found a female-specific 15 bp-indel, which we used to design PCR-primers. These amplified a product concordant with phenotypic sex of some *A. ruthenus*, but not all (data not shown). We also analysed male- or female-specific pool-seq reads, not matching the male reference genome, and independently assembled these unmapped reads from the male/female pools. This resulted in 1,067 additional female-specific contigs spanning 658,732 bp. However, BlastX-searches using these contigs did not detect any sequences related to known sex determination genes.

Identification of a female-specific locus in A. ruthenus by coverage analysis using pool-seq data and a female reference genome

Due to a possible high repeat content of a W-locus, short-read assemblies might not be able to resolve it in large contigs but in rather fragmented small contigs, hindering further analysis. Recently, a long-read female haplotype reference of *A. ruthenus* has become available (GCA_902713425.1), which allowed remapping our pool-seq data and focusing on regions with different coverage between males and females. With stringent filtering parameters (no coverage in males; a minimum of 30% diploid coverage in females and minimum size of regions of 160 bp), we found 11 neighbouring matches (summing up to 2705 bp) in the female genome, while we did only find a single spurious match (185 bp) in the male genome. *Vice versa*, filtering for zero coverage in females and 30% of diploid coverage in males, also resulted in no hits (female genome) or two spurious hits (175 and 181 bp; male genome)

The filtered regions in the female genome fell into the region CACTIG010000179.1:61,229,779-61,285,849 (including 20 kb up- and down-stream). This region corresponds to the region HiC_scaffold_4:43,938,061-43,987,561 in the male assembly as shown by the alignments of the 20 Kbp up- and down-stream sequences. Figure 1 illustrates the female and the male pool-seq coverage of the identified locus. The female specific locus is located 1.162 Mb upstream of the identified cluster of 27 female-specific variants, mentioned above. Furthermore, a single contig (559 bp) of the 1,067 female-specific contigs assembled from the female pool-seq data, did match the identified locus. Thus, three independent analyses of the pool-seq data, specifically sequence variants, assembled female-specific reads and differential sequence coverage of the female genome assembly, support HiC_scaffold_4 as a sex chromosomal sequence, albeit with different specificity and resolution.

Establishment of a W-linked PCR across the sturgeon species tree

Our PCR-marker *AllWSex2* revealed female-specific products for all unambiguously phenotypically sexed females under identical PCR-conditions in six sturgeon species (table 1), spanning the entire phylogenetic tree and the most divergent extant lineages (*A. sturio, A. oxyrinchus* vs. *A. ruthenus, H. huso*), all of which are tetraploid (~120 chromosomes). The marker showed also sex-linkage in two octoploid species (*A. gueldenstaedtii* and *A. baerii*; table 1, figure 2 and figure 3).

Table 1. Numbers of individuals tested with the sex-linked PCR marker *AllWSex2* (p-values for association with sex were calculated using a Pearson's Chi-squared test with Yates' continuity correction; gel pictures in electronic supplementary material, figure S3 to figure S10).

Heredity of the sex-linked marker in three sturgeon species

In three genetic families, consisting of mother, father and their histologically sexed offspring, of *A. oxyrinchus*, *A. ruthenus* and *H. huso*, representing 180 My of sturgeon evolution, the marker *AllWSex2* was exclusively present in females, demonstrating mother to daughter transmission (gel pictures and histology in electronic supplementary material, figure S8 to figure S16).

Discussion

We have discovered a female-specific (W) region, consistent with a potential ZZ/ZW system in sturgeon. This finding is in accordance with previous assumptions or indirect evidence that sturgeons exhibit a female-heterogametic sex determination system [37,49-50]. This genomic region is conserved in at least six sturgeon species, spanning the entire phylogenetic tree, represented by the most divergent species *A. sturio, A. oxyrinchus A. ruthenus,* and *H. huso*, all of which

are formally ancient tetraploid (~120 chromosomes) [33]. In the *A. ruthenus* genome, the sequence is residing on chromosome 4 [33]. Of note, this genomic region is also sex-linked in two octoploid species (*A. gueldenstaedtii* with 258±4 chromosomes, and *A. baerii* with 249±5 chromosomes), demonstrating maintenance of this region and apparently a female heterogametic sex determination system through at least one round of polyploidization. In the future, conservation and evolution of the region in the remaining extant sturgeons should be examined to evaluate the degree to which this ancient female-specific genomic region is also conserved in other species. In particular, identification of the master sex determining gene(s) or other regulatory elements are required. At the current state of research, we also cannot rule out a situation like in Salmonidae, where a "jumping sex locus", perhaps as old as 90 My [31], resides on different chromosomes in different species, all of which share a male heterogametic (XX/XY) system [18-19].

Our data reveal the oldest known vertebrate sex determining system with undifferentiated sex chromosomes. Conservation of a female-specific sequence for 180 My of sturgeon evolution, and across at least one additional polyploidization event, raises many interesting biological and evolutionary questions. We hypothesize that this region might be part of the sex locus. Given the slow protein evolution of the sturgeon genome, evolving distinctly slower than in teleosts, including basal species such as arowana and arapaima, but also slower than gar, coelacanth or elephant shark [33], we speculate that the relatively conserved karyotypes of the ancient tetraploid species examined here (*A. sturio, A. oxyrinchus*, *A. ruthenus, H. huso*) may have even maintained the same undifferentiated sex chromosome for 180 My. It will be fascinating to examine the mechanisms, which stabilized the sex determination system and whether or not this involves undifferentiated homologues sex chromosomes. Results from previous RADseq [33] and pool-sequencing of the same *A. ruthenus* (sampled from aquaculture populations, and with few if any family-effects expected), revealing 99.98 % of the chromosome as largely undifferentiated, let us assume that the sturgeon sex chromosome contains a large pseudo-autosomal region (PAR; [51-52]; electronic supplementary material, figure S1). This and the large chromosome number (~120 to ~260 for the 6 species) make it challenging to apply SNPs or microsatellites to examine whether *AllWSex2* resides on the same or different linkage groups and therefore is beyond the framework of this study, despite we have generated genetic families from three species.

The disruption of sex determination in gonochoristic animals after genome duplication is considered the main cause for the relatively rare occurrence of polyploid animals compared to plants [53-54]. In vertebrates, in which autotetraploidy may be rarer than (hybrid-origin) allopolyploidy [55], polyploidy occurs especially frequently in amphibians [56], with few of which having their mostly undifferentiated sex chromosome characterized. In clawed frogs (*Xenopus*), alloploidy ranges from diploid to do-decaploid (12n) and subgenome evolution in allopolyploids has only recently been studied [57-58]. In *X. laevis*, the female-determining gene *DM-W* is a paralog of *DMRT1* [59-60], and arose after (and perhaps in response to) tetraploidization [61-63]. However, it is also found in some related *Xenopus* [61-62] but not in the entire radiation [64-65]. The clade-wide conservation of a male heterogametic sex determination gene and system (but not chromosome) in the 50 My-old, ancient tetraploid family Salmonidae (see above) curiously also evolved in a system with a common ancestral polyploidization event, as suggested here for Acipenseridae, with apparently conserved female heterogamety.

Our finding is too recent to allow comprehensive interpretation. Whole genome approaches are underway to address several of the most pressing questions, such as conservation of the region or the entire sex chromosomes in sturgeons. Conservation of the latter would even stronger challenge the classical paradigm of sex chromosome evolution [66-68] than the 180 My-long conservation of this small sex-linked sequence.

Reasons for the challenges to identify a female-specific genomic region in sturgeon

The search for a sex chromosome or sex-linked markers in sturgeons has not been successful for a long time [37,49-50]. We discuss four major challenges, why this has been the case.

First, in *A. ruthenus*, the identified W-specific sequence is very short and resides on a large undifferentiated chromosome. It comprises only ~16 kb, corresponding to ~0.001% of the genome or ~0.02% of the corresponding chromosome (~106 Mb). It is impossible to detect such a small structural difference by karyotype imaging methods unless the W sequence is known and FISH probes can be designed.

Secondly, the high repeat content of the W-specific locus (~43%; www.repeatmasker.org) complicates its assembly by short-read data alone. Thus, when reconstructing female-specific genomic regions from unmapped pool-seq data, we filtered female-specific reads and independently assembled them. Unfortunately, the resulting small contigs were of limited utility for female-specific primer design. Looking for an explanation, we mapped these contigs back to the female genome assembly and found that only 0.09% of the sequence (a single contig) matched the W region, which we have identified by a sequencing coverage analysis.

Thirdly, short-read-based methods that call and analyse variants between female and male sequencing data suffered from the low number of sex-specific variants in sturgeons (few hundreds to thousands depending on filter parameters), especially when reduced representation methods like RADseq are used [33]. Even our whole genome pool-sequencing approach was not able to define the W-specific region by SNP markers. This approach only uncovered a nearby region (distance 1.16 Mb, ~1% of chromosome length); a result, which could point to reduced recombination around the W-specific sequence (W-linked SNPs). Finally, analysis of sequencing coverage enabled identification of the W-specific locus in *A. ruthenus*, but detection of haploid/diploid coverage differences were still complicated due to high variability (electronic supplementary material, figure S2), causing noisy signals for haploid genomic regions. Completely deleted regions in one sex compared to the other provided significant signals, but rely on the availability of a high-quality genome assembly of the heterogametic sex (figure 1).

Potential application of the W-specific marker for conservation measures, aquaculture production and ethical issues

Beyond the basic biological interest in this ancient sex determination mechanism, molecular sexing of sturgeons is of great relevance in studies on wild sturgeon populations, their conservation and conservation aquaculture as well as caviar-producing aquaculture. Currently, ultrasonic diagnosis or biopsies are used for sexing. For ultrasonic sexing to become reliable, sturgeons have to reach progressed stages of maturation, six to ten years in some species, while biopsies are invasive and stress or even harm the fish [69]. Our molecular marker's reliability outperforms methods like early biopsies or ultrasound-sexing. Future application of cotton skin swabs will make molecular sexing much less stressful for the fish. It will also aid conservation by reducing time and efforts to rear fish intended as future broodstock in *ex situ*-programs, reducing costs and improving the selection of candidates for living gene banks, not only based on rare alleles but also on sex. In this case, fish not selected for broodstock development can be released in the frame of recovery programs. In addition, in the caviar aquaculture, females, which are commercially more attractive, can be selected early on, and differentiated rearing for caviar (females) and meat (males) production becomes feasible. Releases of commercially reared fish are no option in conservation programs since they show restricted fitness due to intensive aquaculture [70]. In any case, we strongly emphasize that the early selection of females must not lead to littering male sturgeons, since their meat is a valuable source of protein [71].

Permits and Licences

Permits to catch and propagate wild *H. huso* were provided by the Romanian Ministry of Agriculture and Rural Development (No. 7/19.03.2018). The license to keep and propagate sturgeon at the IGB is no. ZH 114 (LaGeSo, 18/02/2019).

Acknowledgments

This work was funded through COFASP/ERANET (STURGEoNOMICS) by the German Federal Ministry of Food and Agriculture through the Federal Office for Agriculture and Food (grant nos. 2816ERA04G, 2816ERA05G) to M. Stöck, S.W., J.G. and M. Schartl; by the French National Research Agency (grant nos. ANR-16-COFA-0002) to YG, CK and CR, and by the Romanian Executive Agency for Higher Education, Research Development and Innovation Funding (UEFISCDI, contract 3/2017) to RS, MC and DH. This research was also in part supported by the Russian Science Foundation (RSF) to DP (grant number 18-44-04-007), by the Deutsche Forschungsgemeinschaft to MSch (DFG SCHA 408/14-1), and by a European Maritime and Fisheries Fund to YG (EMFF/FEAMP project SIBERSEX). Research on *A. gueldenstaedtii* was supported by a European Maritime and Fisheries Fund to YG (EMFF/FEAMP project S'STURGEON). We thank Marilena Maereanu, Katarina Tosic, and Marian Paraschiv for help with *H. huso* catching, propagation and raising of animals; Asja Vogt, Marcus Ebert, Jan Hallermann, Kevin Bäumler, Udo Wolf for help with *A. oxyrinchus* or *A. ruthenus* propagation and raising of animals; Elodie Dupin-de-Beyssat, Camille Eché, Antje Tillack, Sophia Lambert, Alina Heide, and Petra Kersten for help in the lab, Lukáš Kratochvíl for references, Jörg Plötner and Robert Schreiber for access to the Qiagen robotic workstation, and two anonymous reviewers for their constructive comments on a previous version of this paper..

For Review Only

References

1. Beukeboom L, Perrin P. 2014 The evolution of sex determination. New York, NY: Oxford University Press.
2. Ohno S. 1967 Sex chromosomes and sex-linked genes. New York, NY: Springer.
3. Rice WR. 1996 Evolution of the Y sex chromosome in animals. *Biosci.* **46**, 331–343. (doi:10.2307/1312947)
4. Charlesworth B, Charlesworth D. 2000 The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572. (doi:10.1098/rstb.2000.0717)
5. Bull JJ. 1983 Evolution of sex determining mechanisms. London: Benjamin/Cummings.
6. Rice WR. 1987 The accumulation of sexually antagonistic genes as selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evol.* **41**, 911–914. (doi:10.2307/2408899)
7. Charlesworth B, Charlesworth D. 2000 The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572. (doi: 10.1098/rstb.2000.0717)
8. Charlesworth D. (in press): The timing of genetic degeneration of sex chromosomes. *Phil. Trans. R. Soc. B.* (this theme issue). (doi:10.1098/rstb.2015.0615)
9. Lenormand T, Fyon F, Sun E, Roze D. 2020 Sex chromosome degeneration by regulatory evolution. *Curr. Biol.* **30**, 1–6. (doi:10.1016/j.cub.2020.05.052)
10. Kikuchi K, Hamaguchi S. 2013. Novel sex-determining genes in fish and sex chromosome evolution. *Dev. Dyn.* **242**, 339–353. (doi:10.1002/dvdy.23927)
11. Eggert C. 2004 Sex determination: the amphibian model. *Reprod. Nutr. Dev.* **44**, 539–549. (doi:10.1051/rnd:2004062)
12. Schartl M, Schmid M, Nanda I. 2016 Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma* **125**, 553-571. (doi:10.1007/s00412-015-0569-y)
13. Schartl M. 2004. Sex chromosome evolution in non-mammalian vertebrates. *Curr. Opin. Gen. Develop.* **14**, 634–641. (doi:10.1016/j.gde.2004.09.005)

14. Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D. 2015 Restriction site-associated DNA sequencing. (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Mol. Biol. Evol.* **32**, 1296–1309. (doi:10.1093/molbev/msv023)
15. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-A, *et al.* 2014 Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899. (doi: 10.1371/journal.pbio.1001899)
16. Sarre SD, Ezaz T, Georges A. 2011 Transitions between sex-determining systems in reptiles and amphibians. *Annu. Rev. Genomics Hum. Genet*. **12**, 391-406. (doi:10.1146/annurev-genom-082410-101518)
17. Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, *et al*. 2018 A rapid rate of sex-chromosome turnover and nonrandom transitions in true frogs. *Nat. Comm*. **9**, 4088. (doi: 10.1038/s41467-018-06517-2)
18. Woram RA, Gharbi K, Sakamoto T, Hoyheim B, Holm LE, Naish K, McGowan C *et al.* 2003 Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Research* **13**, 272–280. (doi:10.1101/gr.578503)
19. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, *et al*. 2013 The sexually dimorphic on the Y-chromosome gene. (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* **6**, 486–496. (doi:10.1111/eva.12032)
20. Perrin N. 2009 Sex reversal: a fountain of youth for sex chromosomes? *Evol.* **63**, 3043–3049. (doi:10.1111/j.1558-5646.2009.00837.x)
21. Stöck M, Horn A, Grossen C, Lindtke D, Sermier R, Betto-Colliard C, *et al.* 2011 Ever-young sex-chromosomes in European tree frogs. *PLOS Biol.* **9**, e1001062. (doi:10.1371/journal.pbio.1001062)
22. Rodrigues N, Studer T, Dufresnes C, Perrin P. 2018 Sex-Chromosome recombination in common frogs brings water to the fountain-of-youth. *Mol. Biol. Evol*. **35**, 942–948. (doi:10.1093/molbev/msy008)
23. Rovatsos M, Vukić J, Altmanová M, Johnson Pokorná M, Moravec J, Kratochvíl L. 2016 Conservation of sex chromosomes in lacertid lizards. *Mol. Ecol.* **25**, 3120–3126. (doi: 10.1111/mec.13635)
24. Iannucci, A., Altmanová, M., Ciofi, C. *et al*. 2019 Conserved sex chromosomes and karyotype evolution in monitor lizards (Varanidae). *Heredity* **123**, 215–227. (doi:10.1038/s41437-018-0179-6)
25. Rovatsos M, Pokorná M, Altmanová M, Kratochvíl L. 2014 Cretaceous park of sex determination: sex chromosomes are conserved across iguanas. *Biol. Lett.* **10**, 20131093. (doi: 10.1098/rsbl.2013.1093)
26. Rovatsos M, Vukic J, Lymberakis P, Kratochvíl L. 2015 Evolutionary stability of sex chromosomes in snakes. *Proc. R. Soc. B***282**, 20151992. (doi:10.1098/rspb.2015.1992)
27. Vicoso B, Kaiser VB, Bachtrog D. 2013 Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc. Natl. Acad. Sci. USA **110**, 6453–6458. (doi:10.1073/pnas.1217027110)
28. Yazdi HP, Ellegren H. 2014 Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in Ratites. *Mol. Biol. Evol*. **31**, 1444–1453. (doi: 10.1093/molbev/msu101)
29. Kostmann A, Kratochvil L, Rovatsos M 2020. Poorly differentiated XX/XY sex chromosomes are widely shared across skink radiation. *BioRxiv*. (doi:10.1101/2020.08.29.273524)
30. Pan Q, Feron R, Jouanno E, Darras H, Herpin, A. 2020 The rise and fall of the ancient northern pike master sex determining gene. *BioRxiv*. (doi:10.1101/2020.05.31.125336)
31. Macqueen DJ, Johnston IA. 2014 A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. R. Soc. B* **281**, 20132881. (doi:10.1098/rspb.2013.2881)
32. Hochleithner M, Gessner J. 2001. The sturgeon and paddlefishes of the world— Biology and Aquaculture. *Aquatech. Publ*. ***106***, Aquatech.: Kitzbuehel, 165 pp.
33. Du K, Stöck M, Kneitz S, Klopp C, Woltering J, Adolfi M, *et al*. 2020 The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol*. **4**, 841–852. (doi:10.1038/s41559-020-1166-x)
34. Meyer A, Van de Peer Y. 2005 From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**, 937-945. (doi:10.1002/bies.20293)
35. Braasch I, Postlethwait JH. 2012 Polyploidy in fish and the teleost genome duplication. pp 341-383, in: Soltis P, Soltis D. eds. Polyploidy and genome evolution. Springer: Berlin, Heidelberg. (doi:10.1007/978-3-642-31442-1_17)
36. Havelka M, Hulák, M, Bailie D, Prodöhl P, Flajšhans M. 2013 Extensive genome duplications in sturgeons: new evidence from microsatellite data. *J. Appl. Ichthyol.* **29**, 704-708. (doi:10.1111/jai.12224)
37. Keyvanshokooh S, Gharaei, A. 2010 A review of sex determination and searches for sex-specific markers in sturgeon. *Aquac. Res*. **41**, e1–e7. (doi:10.1111/j.1365-2109.2009.02463.x)
38. Fopp-Bayat D, Kolman R, Woznicki P. 2007 Induction of meiotic gynogenesis in sterlet (*Acipenser ruthenus*) using UV-irradiated bester sperm. *Aquac*. **264**, 54-58. (doi: 10.1016/j.aquaculture.2006.12.006)
39. Li H. 2018 Minimap2: pairwise alignment for nucleotide sequences. *Bioinf.* **34**, 3094–3100. (doi:10.1093/bioinformatics/bty191)
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* 2009. 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools. *Bioinf.* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
41. Rimmer A, Phan H, Mathieson I. *et al*. 2014 Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Gen.* **46**, 912–918. (doi:10.1038/ng.3036)
42. Quinlan AR, Hall IM. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinf.* **26**, 841–842. (doi:10.1093/bioinformatics/btq033)

For Review Only

43. Peng Y, Leung HC, Yiu SM, Chin FY. 2012 IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinf*. **28**, 1420-1428. (doi:10.1093/bioinformatics/bts174)
44. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD *et al*. 2013 *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Prot.* **8**, 1494. (doi:10.1038/nprot.2013.084)
45. Birney E, Clamp M, Durbin R. 2004 GeneWise and genomewise. *Genome Res.* **14**, 988-995. (doi:10.1101/gr.18655049)
46. Nakamura T, Yamada KD, Tomii K, Katoh, K. 2018 Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinf.* **34**, 2490-2492. (doi:10.1093/bioinformatics/bty121)
47. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinf.* **25**, 1972-1973 (doi:10.1093/bioinformatics/btp348)
48. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinf*. 30, 1312-1313. (doi:10.1093/ bioinformatics/btu033)
49. Havelka M, Arai K. 2018 Hybridization and polyploidization in sturgeon. pp. 669–687 In: Sex control in aquaculture (eds Wang H.-P. *et al*.; John Wiley & Sons Ltd.
50. Fajkowska M, Ostaszewska T, Rzepkowska M. 2019 Review: Molecular mechanisms of sex differentiation in sturgeons. *Rev. Aquac*. **12**, 1003–1027. (doi:10.1111/raq.12369)
51. Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, *et al.* 2011 About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet*. **27**, 358–367. (doi:10.1016/j.tig.2011.05.001)
52. Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice WR, Valenzuela N. 2011 Are all sex chromosomes created equal? *Trends Genet.* **27**, 350-357. (doi: 10.1016/j.tig.2011.05.005)
53. Muller H. 1925 Why polyploidy is rarer in animals than in plants. *Am. Nat.* **59**, 346–353 (1925). (doi: 10.1086/280047)
54. Orr HA. 1990 'Why polyploidy is rarer in animals than in plants' revisited. *Am. Nat*. **136**, 759–770. (doi: 10.1086/285130)
55. Mable B. 2004 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biol. J. Linn. Soc.* **82**, 453–466 (doi:10.1111/j.1095-8312.2004.00332.x)
56. Schmid M, Evans BJ, Bogart JP. 2015 Polyploidy in Amphibia. *Cytogen. Genome Res*. **145**, 315–330. (doi:10.1159/000431388)
57. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Shuji Takahashi S *et al.* 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **538**, 336–343. (doi: 10.1038/nature19840)
58. Furman BLS, Dang UJ, Evans BJ, Golding GB. 2018 Divergent subgenome evolution after allopolyploidization in African clawed frogs (*Xenopus*). *J. Evol. Biol*. **31**, 1945-1858. (doi:10.1111/jeb.13391)
59. Yoshimoto S, Ito M. 2011 A ZZ/ZW-type sex determination in *Xenopus laevis*. *FEBS J*. **278**, 1020-1026. (doi:10.1111/j.1742-4658.2011.08031.x)
60. Ito M. 2018 Sex determination and differentiation in frogs. pp 349-366. In: Kobayashi K., Kitano T., Iwao Y., Kondo M. (eds) Reproductive and developmental strategies. Diversity and commonality in animals. Springer: Tokyo (doi:10.1007/978-4-431-56609-0_17)
61. Bewick AJ, Anderson DW, Evans BJ. 2011 Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). *Evol*. **65**, 698–712. (doi:10.1111/j.1558-5646.2010.01163.x)
62. Furman BLS, Evans BJ. 2016 Sequential turnovers of sex chromosomes in African clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination. *G3 (Bethesda)* **6**, 3625–3633. (doi:10.1534/g3.116.033423)
63. Mawaribuchi S, Takahashi S, Wada M, Uno Y, Matsuda Y, Kondo M. *et al.* 2017 Sex chromosome differentiation and the W- and Z-specific loci in *Xenopus laevis*. *Dev. Biol*. **426**, 393–400. (doi: 10.1016/j.ydbio.2016.06.015)
64. Furman BLS, Evans BJ. 2018 Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. *Genome Biol. Evol.* **10**,742–755. (doi:10.1093/gbe/evy045)
65. Song X-Y, Furman BLS., Premachandra T, Knytl M, Cauret CMS *et al.* (in press) Sex chromosome degeneration, turnover, and sex-biased expression of sex-linked transcripts in African clawed frogs (*Xenopus*). *Phil. Trans. R. Soc. B*. (doi:10.1098/20200095)
66. Muller HJ. 1914 A gene for the fourth chromosome of *Drosophila*. *J. Exp. Zool*. *17*, 325 – 336. (doi:10.1002/jez.1400170303)
67. Muller HJ, Painter TS. 1932 The differentiation of the sex chromosomes of *Drosophila* into genetically active and inert regions. *Z. Indukt. Abstamm. Vererb.-lehre*. **62**, 316 – 365. (doi:10.1007/BF01948610)
68. Wright AE, Dean R, Zimmer F, Mank, JE. 2016 How to make a sex chromosome? *Nat. Commun*. **7**, 12087 (doi:10.1038/ncomms12087)
69. Chebanov M, Rosenthal H, Gessner J, Van Anrooy R, Doukakis P, Pourkazemi M, Williot P, 2011 Sturgeon hatchery practices and management for release - guidelines. FAO Fisher. Aquac. Techn. Paper 570. FAO, Ankara, 110 pp.
70. Gessner J, Friedrich T, Rosenthal H. 2019 Summary report on the major topics addressed at ISS 8 (8th International Symposium on Sturgeons). *J. Appl. Ichthyol*. **35**, 1-7. (doi:10.1111/jai.13866)
71. Bronzi P, Chebanov M, Michaels JT, Wei Q, Rosenthal H, Gessner J. 2019 Sturgeon meat and caviar production: Global update to 2017. *J. Appl. Ichthyol*. **35**, 257-266. (doi: 10.1111/jai.13870)

For Review Only

Tables

Species	Ploidy	Positive Females	Positive Males	*p*-value of association with sex	Type of phenotypic sexing
A. sturio	4n	18/18	0/18	1.456e-08	ultrasound, biopsy if contradiction to genotype, few autopsies
A. oxyrinchus	4n	12/12	0/15	1.536e-06	adult breeders with known reproduction, histology in family offspring
H. huso	4n	12/12	0/16	9.311e-07	adult breeders with known reproduction, biopsy, histology in family offspring
A. ruthenus	4n	25/25	0/27	4.11e-12	adult breeders with known reproduction, biopsy, histology in family offspring
A. baerii	8n	28/28	0/28	5.352e-13	adult breeders with known production of gametes
A. gueldenstaedtii	8n	15/15	0/12	1.536e-06	adult breeders with known reproduction, ultrasound, few autopsies

Table 1. Numbers of individuals tested with the sex-linked PCR marker *AllWSex2* (p-values for association with sex were calculated using a Pearson's Chi-squared test with Yates' continuity correction; gel pictures in electronic supplementary material, figure S3-S10). To avoid circular evidence, neither the 61 (30 females, 31 males) *A. ruthenus* used for pool-sequencing, nor the offspring of the three genetic families of *A. ruthenus, A. oxyrhinchus* and *H. huso* (supplementary materials, figures S8-S10) but only their parents are counted here.

For Review Only

Figures

Figure 1. Detection of a female-specific locus in *Acipenser ruthenus*, when mapping female/male pool-sequencing reads to a female (ZW) reference genome assembly. Genome-wide signal detection was highly specific, when screening for larger regions (here >= 160 bp) with zero and non-zero coverage between sexes. The red block and arrow show the position of the *AllWSex2*-PCR-marker (CACTIG010000179.1: 61246236 - 61246344), which distinguishes sex in six sturgeon species. The green block and arrow mark the location of the only contig matching the region (CACTIG010000179.1: 61261166 - 61261703), when using the female-specific pool-seq read assembly approach. For final submissions, figures should be uploaded as separate, high resolution, figure files.

Figure 2. PCR products using the *AllWSex2* marker in six species across the sturgeon phylogeny under identical PCR-conditions (electronic supplementary material, texts S1 - S3) for two females and two males of each species. PCR products migrated ca. 45 min at 70 V on a 2% agarose gel in 1x TAE-buffer. Symbols and abbreviations: ♀: female; ♂: male; B: Blank, i.e. no DNA template in PCR; L: ladder, 100 bp-size-marker (Quantitas, Biozym).

Figure 3. Phylogenomic tree of sturgeon species built by using RAxML with Maximum Likelihood (details: Materials and Methods). Numbers at nodes refer to bootstrap support values. The bar above indicates branch length of sturgeon lineages, while the bar below indicates branch length of the outgroup lineage coelacanth (*Latimeria chalumnae*). The approximate position of *A. sturio*, for which reliable transcriptomic data were unavailable, was added according to www.timetree.org. The root of Acipenseridae was dated at ~180 Mya, according to [33]. Red arrows mark those species, for we have obtained female-specific PCR-products, using primers *AllWSex2*_F/_R.

Electronic Supplementary material

Submitted in separate file: texts S1-S5, tables S1-S2, and figures S1-S16.

Additional Information

Ethics

The research using sturgeon in France, Romania and Germany were carried out in accordance with approved guidelines and obtained ethical approval by the institutional animal care and use committees; permits and licences as shown above.

Data Accessibility

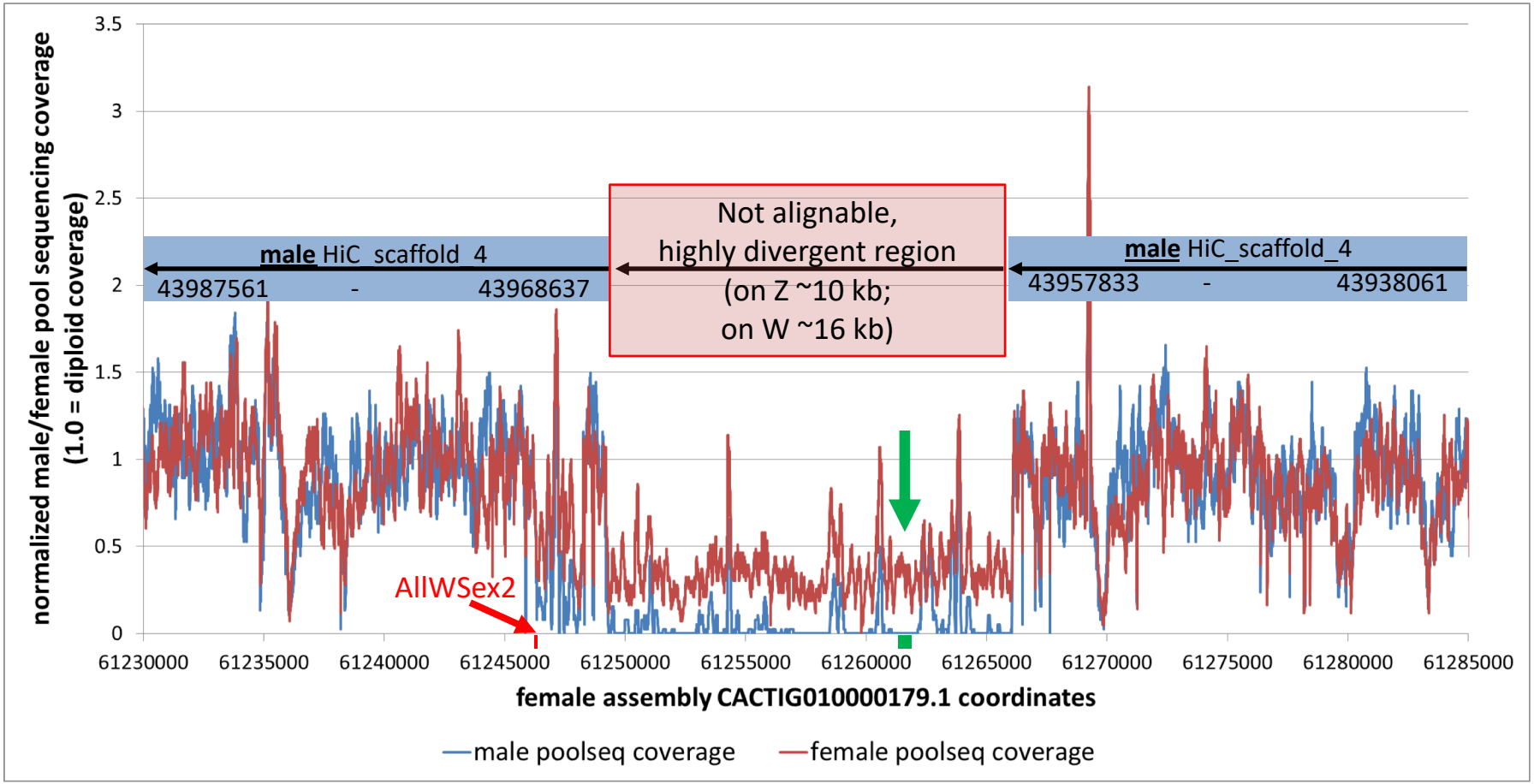
The pool-sequencing Illumina reads are available in the Sequence Read Archive (SRA: SRX9341540 and SRX9341541), under BioProject reference PRJNA670610; all other datasets supporting this article have been uploaded as part of the Supplementary Material.

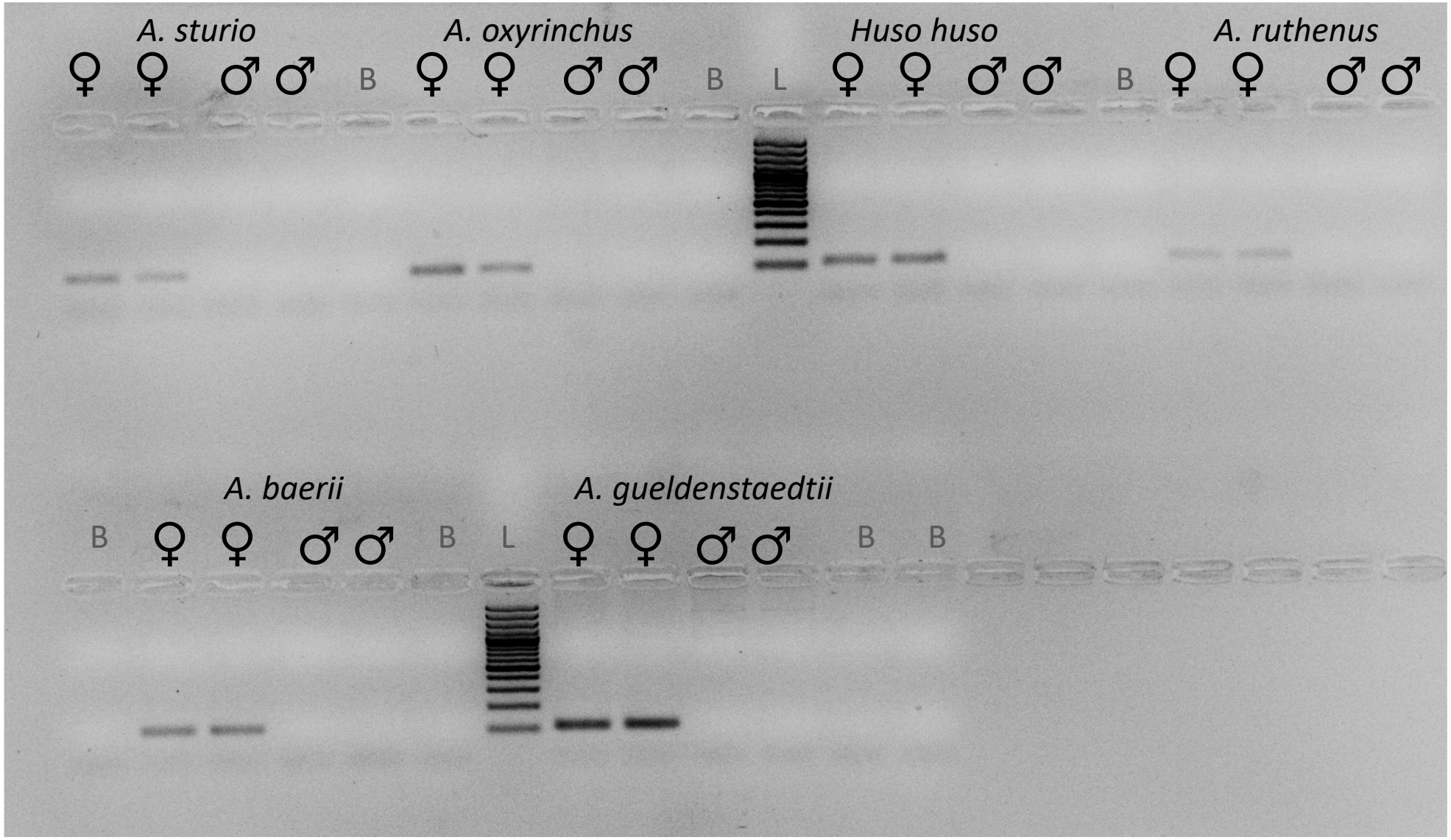
Authors' Contributions

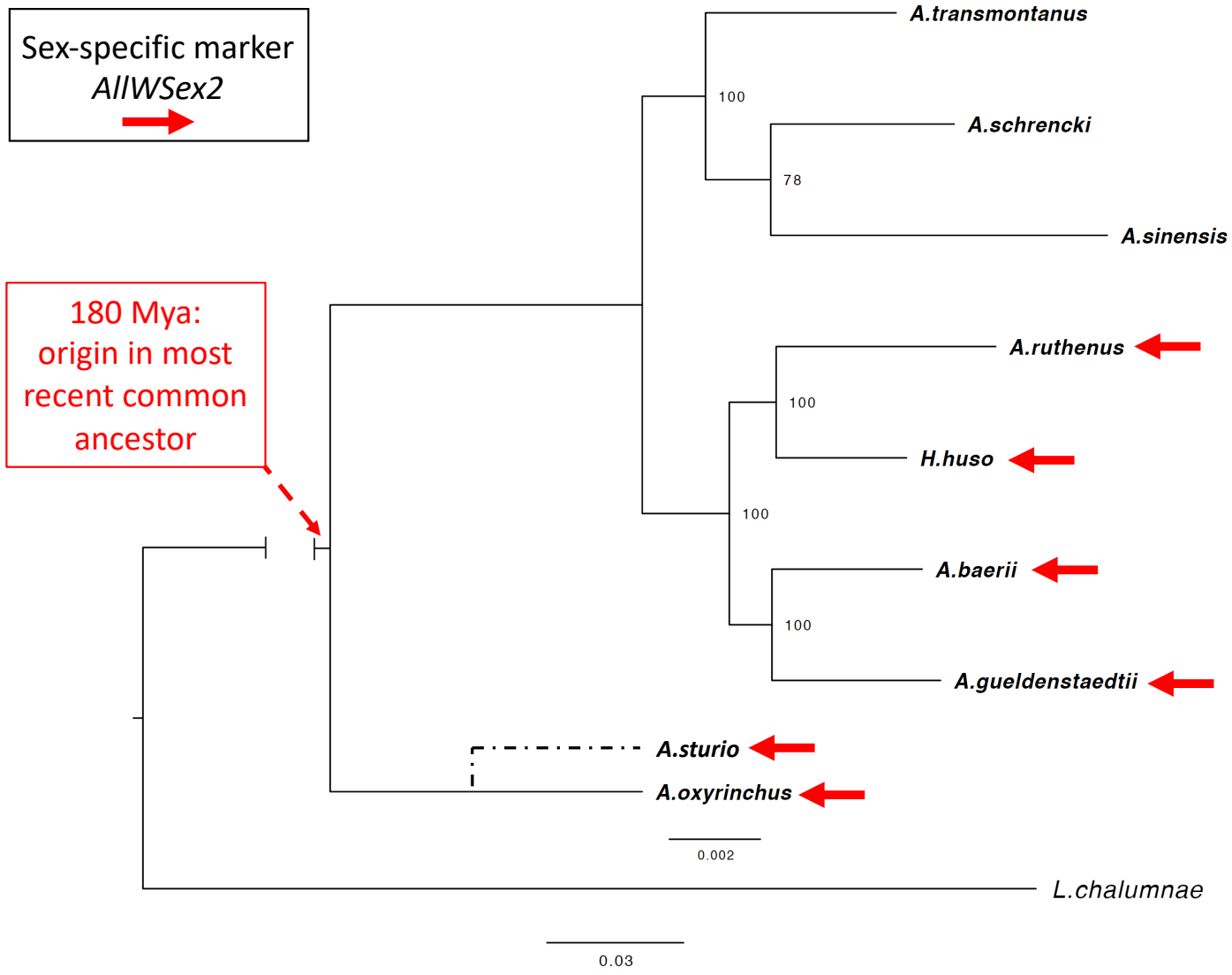
Designed the study: MSchartl, MStöck, YG, SW, JG; wrote manuscript draft: HK, MS, MSch, YG, CH, JG, CK; analysed data and performed molecular analyses: HK, MStöck, CH, EK, CK, CLR, KD, ESYP, RD, JG, DP, WKleiner, KK, DKL; provided biological samples: RS, DH, MC, JG, HK, MStöck, WKloas, AB, EB, BD, PH, RM, PP. All authors read, contributed to and approved the manuscript.

Competing Interests

We have no competing interests.


normalized male/female pool sequencing coverage
(1.0 = diploid coverage)
3.5
3
2.5
2
1.5
1
0.5
0
male HiC_scaffold_4
43987561 - 43968637
Not alignable,
highly divergent region
(on Z ~10 kb;
on W ~16 kb)
male HiC_scaffold_4
43957833 - 43938061
AllWSex2
61230000
61235000
61240000
61245000
61250000
61255000
61260000
61265000
61270000
61275000
61280000
61285000
female assembly CACTIG010000179.1 coordinates
male poolseq coverage
female poolseq coverage


A. sturio
♀ ♀ ♂ ♂ B
A. oxyrinchus
♀ ♀ ♂ ♂ B
Huso huso
L ♀ ♀ ♂ ♂ B
A. ruthenus
♀ ♀ ♂ ♂
A. baerii
B ♀ ♀ ♂ ♂ B
A. gueldenstaedtii
L ♀ ♀ ♂ ♂ B B


Sex-specific marker
AllWSex2
180 Mya:
origin in most
recent common
ancestor
A.transmontanus
100
A.schrencki
78
A.sinensis
A.ruthenus
100
H.huso
100
A.baerii
100
A.gueldenstaedtii
A.sturio
A.oxyrinchus
0.002
L.chalumnae
0.03

Phil. Trans. R. Soc. B.

Heiner Kuhl, Yann Guiguen, Christin Höhne, Eva Kreuz, Kang Du, Christophe Klopp, Céline Lopez-Roques, Elena Santidrian Yebra-Pimentel, Mitica Ciorpac, Jörn Gessner, Daniela Holostenco, Wibke Kleiner, Klaus Kohlmann, Dunja K. Lamatsch, Dmitry Prokopov, Anastasia Bestin, Emmanuel Bonpunt, Bastien Debeuf, Pierrick Haffray, Romain Morvezen, Pierre Patrice, Radu Suciu, Ron Dirks, Sven Wuertz, Werner Kloas, Manfred Schartl, Matthias Stöck

A 180 My-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes

Electronic Supplementary Material

(Texts S1-S5, tables S1-S2, and figures S1-S16: pp. 1-10)

Supplementary Texts

Text S1. *Primer sequences for marker* AllWSex2

AllWSex2_F	5‘_TGATCAACCTCTTCAGCAATGTC_3‘
AllWSex2_R	5‘_TGAGAGCCACTGTACTAACACA_3‘

Text S2. *Pipetting scheme for* AllWSex2-*PCR*

	Final concentration	Stock	Volume [µl]
ddH_2O			17.3
10x TopTaqbuffer (Qiagen)	1	10x	2.5
dNTP	0.2	[10mM]	0.5
AllWSex2_F	0.4	[10 pmol/µl]	1.25
AllWSex2_R	0.4	[10 pmol/µl]	1.25
TopTaq Polymerase (Qiagen)	2 U/rxn	[120 rxn/24µl]	0.2
DNA	20 ng/µl		2.0

Text S3. *Temperature scheme for* AllWSex2-*PCR*

STEP		T [°C]	TIME	REPETITION
1	Denaturation	95	2 min	
2	Denaturation	94	1 min	
3	Annealing	56	45 s	30-35 cycles
4	Elongation	72	45 s	
5	Elongation	72	5 s	
6	Storage	10	for ever	

Text S4. *Bioinformatics to obtain sex-specific variants based on a male reference genome of A. ruthenus*

Variant calling using all data resulted in a number of 15,682.485 genome-wide variant calls (avg. 8.5 per kbp). Stringent filtering for variants with supporting reads, only present in the female or male pools, resulted in 1,682 female-specific variants and 318 male specific variants (ratio female/male = 5.3). Using different cut-offs for variant filtering, we consistently found the female pool to harbour more sex-specific variants, which hints at a ZW sex determination system. However, the variants did not match known genes related to sex determination. We clustered sex-specific variants by genomic distance and ranked these clusters by variant count. The region with the most female specific variants was HiC_scaffold_4:45,149,338-45,165,057, which contained 27 variants. The second-best cluster consisted of just 6 variants, observed on HiC-scaffold_7:44,514,447-44,516,696 followed by several other hits on HiC_scaffold_4. Looking at male-specific variants, the best region only consisted of 3 variants (HiC_scaffold_4). In the top candidate region, we found a female-specific 15 bp-indel, which we used to design PCR primers, these primers could already sex some of our sexed *A. ruthenus* samples, but not all (details not shown).

Text S5. *Corresponding regions in the current A. ruthenus NCBI reference genome*

The current NCBI reference genome (GCF_010645085.1, as of Oct 2020) corresponds to the male genome assembly used for read mapping in our study with a few differences:

A) NCBI assemblies have different sequence identifiers
B) The NCBI assembly has a different gap-size, leading to a shift in coordinates

The changes of sequence identifiers and coordinates are shown in table S1 for the regions mentioned in this study.

Supplementary Tables

Table S1. Corresponding coordinates of two *A. ruthenus* genome assembly versions.

Acipenser_ruthenus.ARUT1.2.dna.toplevel.fa	NCBI: GCF_010645085.1 (Oct 2020)
HiC_scaffold_4: 43,938,061 - 43,987,561	NC_048326.1: 43,898,461 - 43,947,961
HiC_scaffold_4: 45,149,338 - 45,165,057	NC_048326.1: 45,108,538 - 45,124,257
HiC_scaffold_7: 44,514,447 - 44,516,696	NC_048329.1: 44,475,147 - 44,477,396

Table S2. Data used for phylogenetic tree computation.

Species	Resource type	Number of contigs/proteins	Resource
A. baerii	transcriptome	46028	http://publicsturgeon.sigenae.org/home.html
A. gueldenstaedtii	transcriptome	369441	https://www.ncbi.nlm.nih.gov/Traces/wgs/?page=1&view=tsa&search=acipenser
A. oxyrinchus	transcriptome	42114	http://publicsturgeon.sigenae.org/home.html
A. ruthenus	genome	36020	https://www.ncbi.nlm.nih.gov/genome/?term=txid7906[orgn]
A. schrencki	transcriptome	57124	http://publicsturgeon.sigenae.org/home.html
A. sinensis	transcriptome	63077	http://publicsturgeon.sigenae.org/home.html
A. transmontanus	transcriptome	55806	http://publicsturgeon.sigenae.org/home.html
H. huso	transcriptome	548535	Trinity assembly from mixed organ
L. chalumnae	genome	20932	https://www.ncbi.nlm.nih.gov/genome/?term=coelacanth

Supplementary Figures

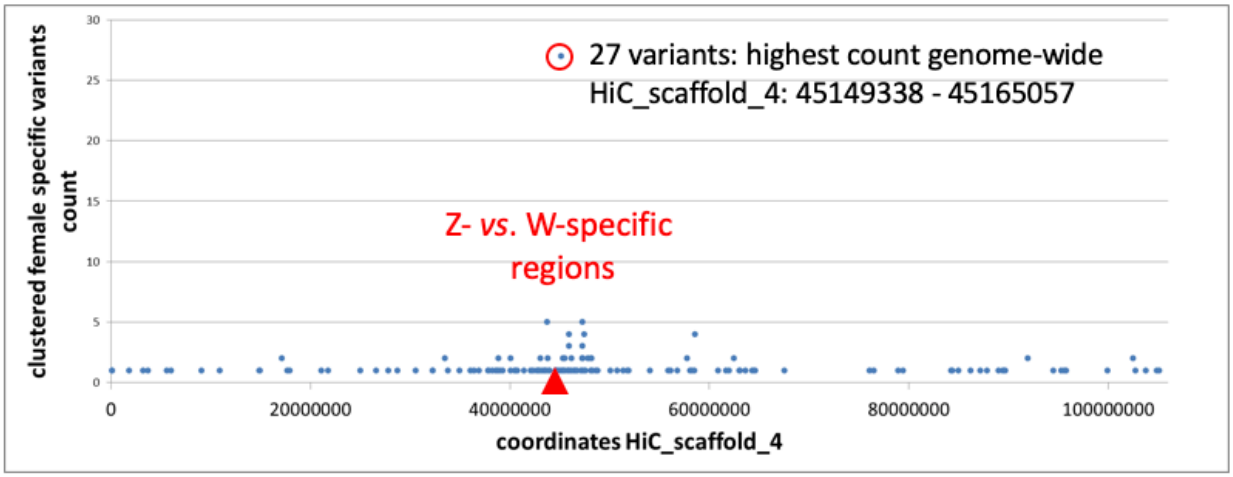


Figure S1. Counts of female-specific variants, clustered according to their distance on HiC_scaffold 4.

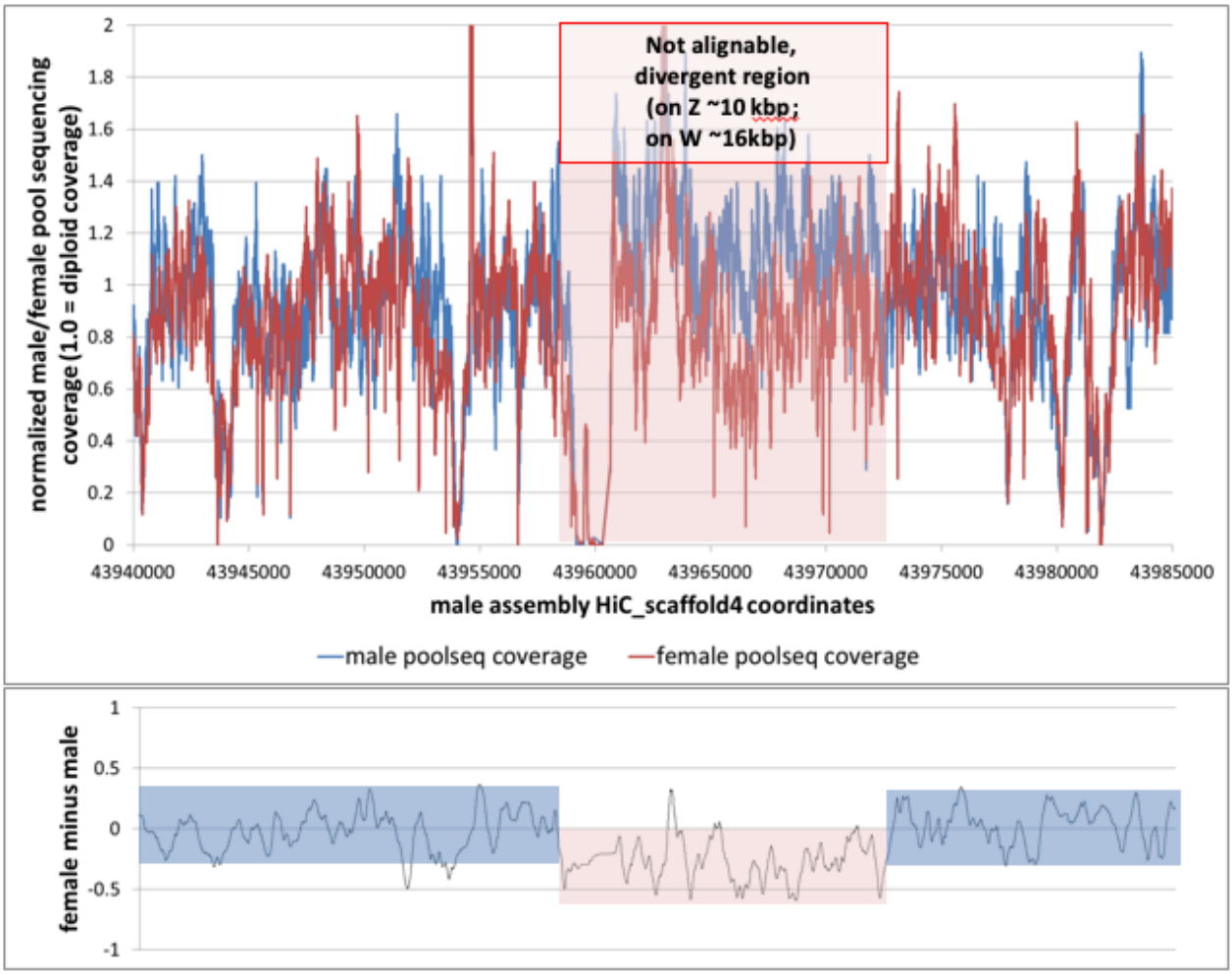


Figure S2. Mapping of pool-seq data of *A. ruthenus* to the male (ZZ) reference genome, female pool-seq data (ZW) shows a ~10 Kb region with reduced coverage. Due to overall coverage fluctuations, it is difficult to reliably detect small loci with haploid coverage differences in genome-wide approaches.

PCR-Results using the AllWSex2*-marker*

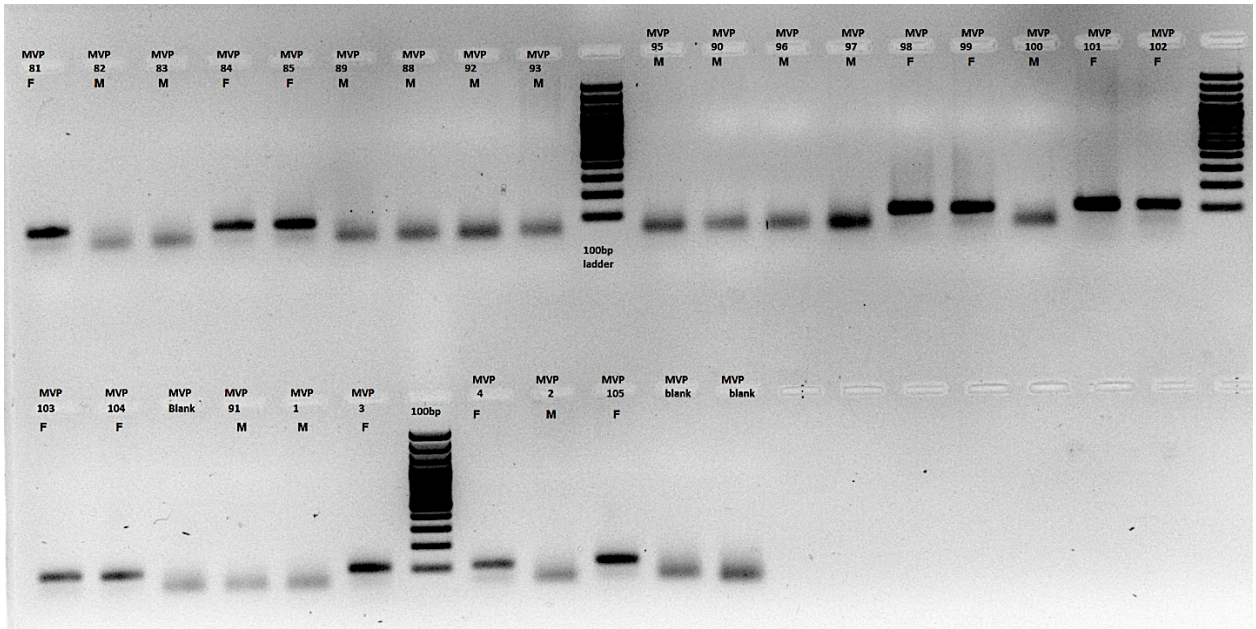


Figure S3. PCR-Result using primers AllWSex2F/R for 26 *A. oxyrinchus* (12 females F, 14 males M) with known phenotypic sex (MVP presents internal sample numbers)

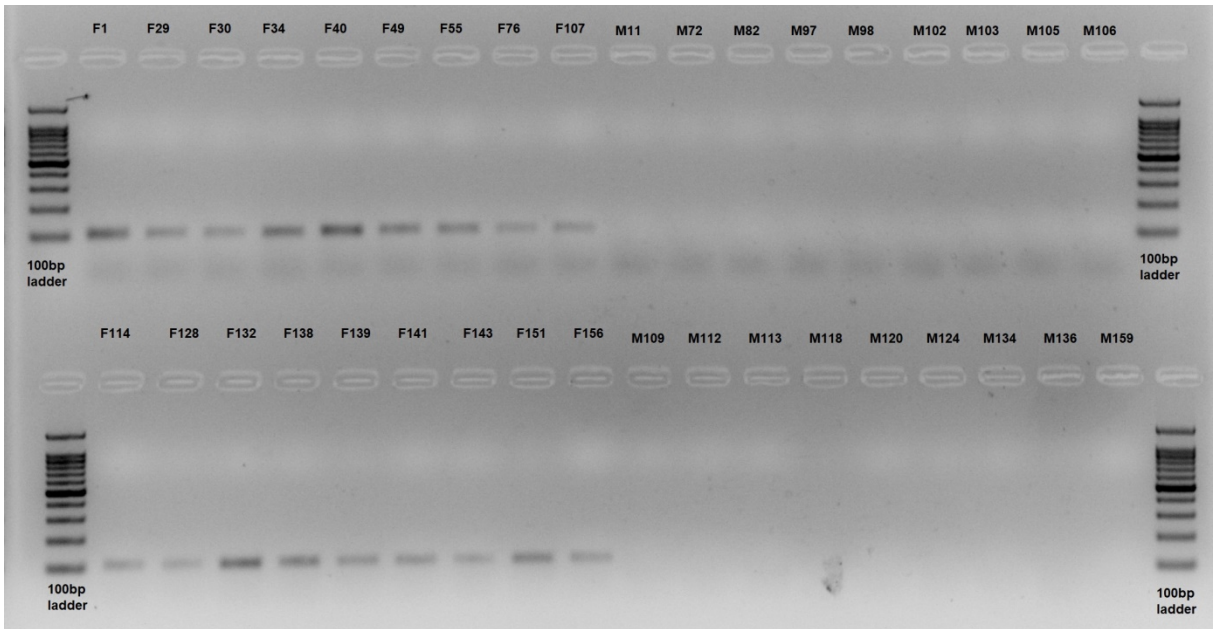


Figure S4. PCR-Result using primers AllWSex2F/R for 36 *A. sturio* (18 females, 18 males) with known phenotypic sex.

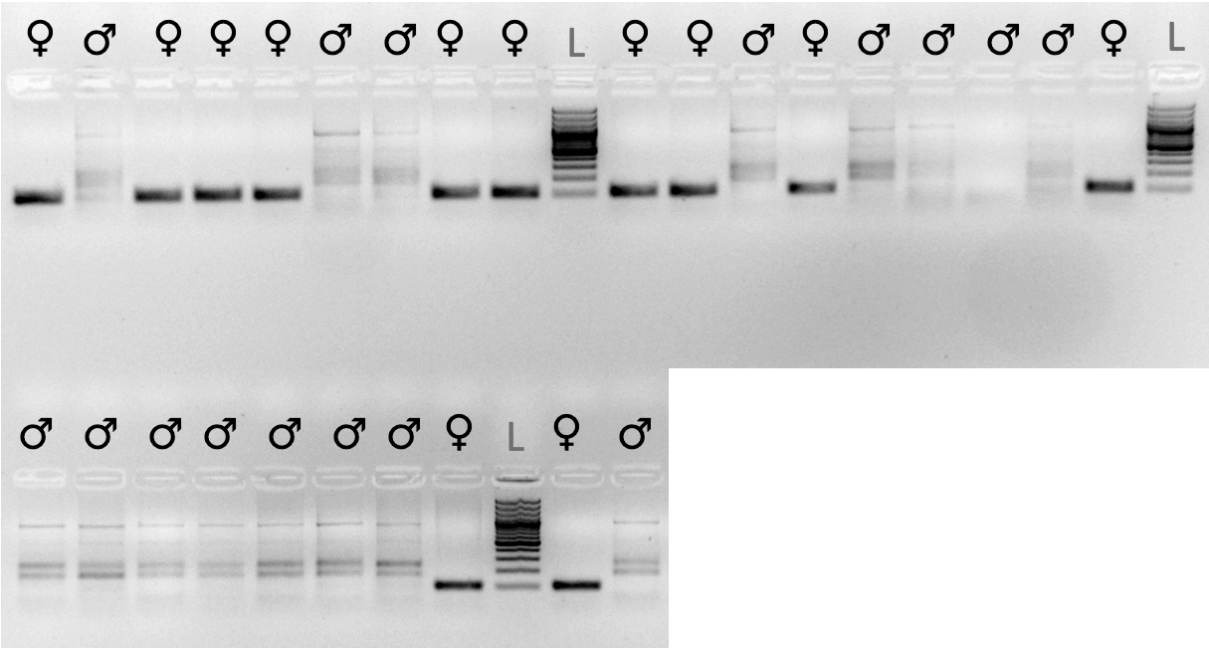


Figure S5. PCR-Result using primers AllWSex2F/R for 28 *Huso huso* (12 females, 16 males) with known phenotypic sex.

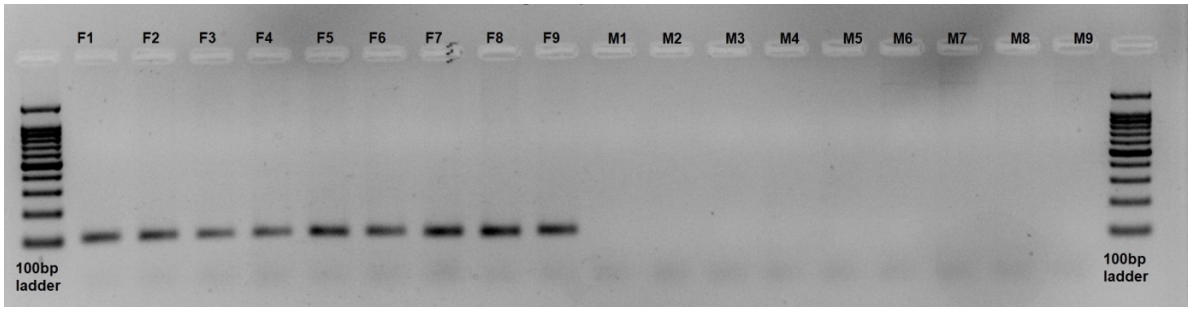


Figure S6. PCR-Result using primers AllWSex2F/R for 18 *A. gueldenstaedtii* (9 females, 9 males) with known phenotypic sex.

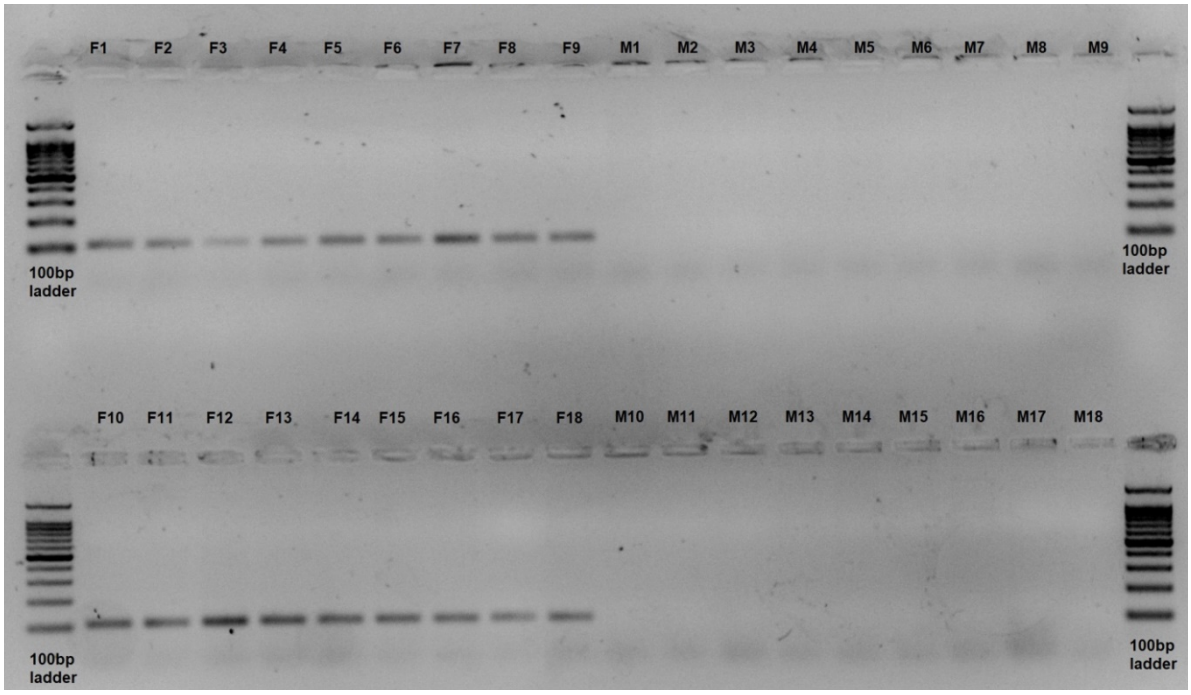


Figure S7. PCR-Result using primers AllWSex2F/R for 18 *A. baerii* (9 females, 9 males) with known phenotypic sex.

Heredity of marker AllWSex2

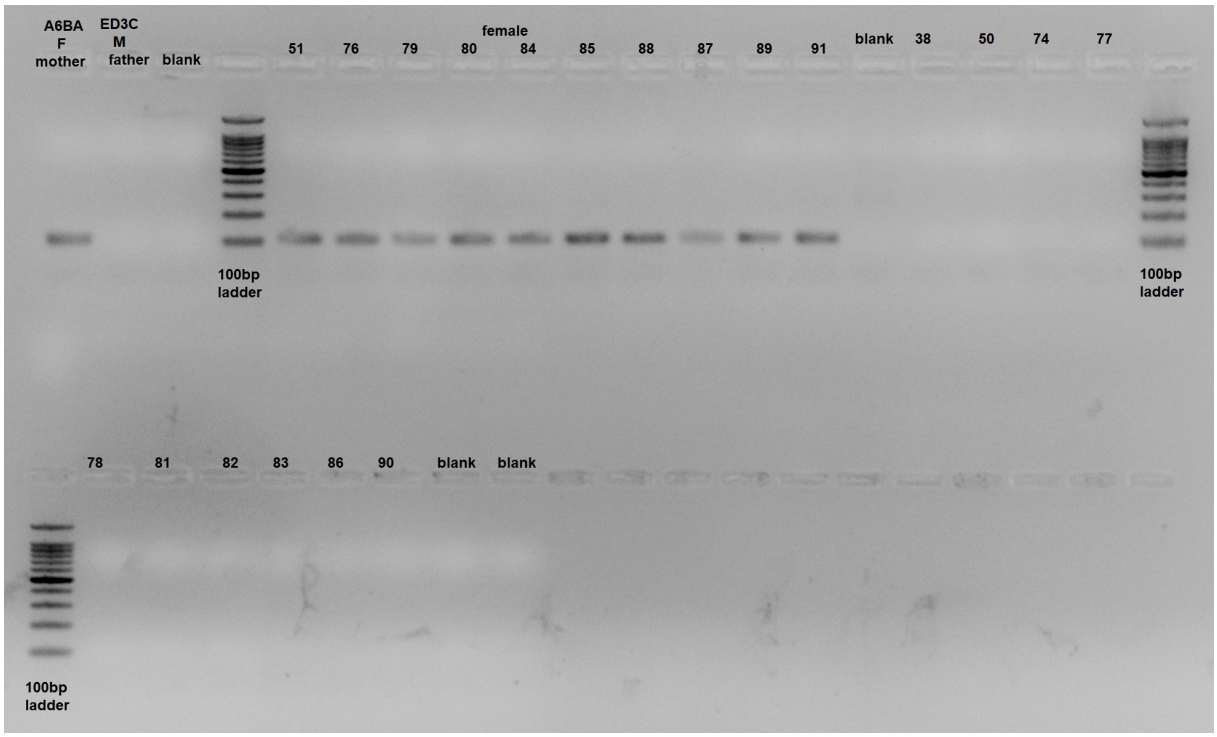


Figure S8. Inheritance of the PCR-marker *AllWSex2* in a genetic family of *A. oxyrinchus*, with mother (A6BA), father (ED3C) and 10 daughters, followed by 10 sons.

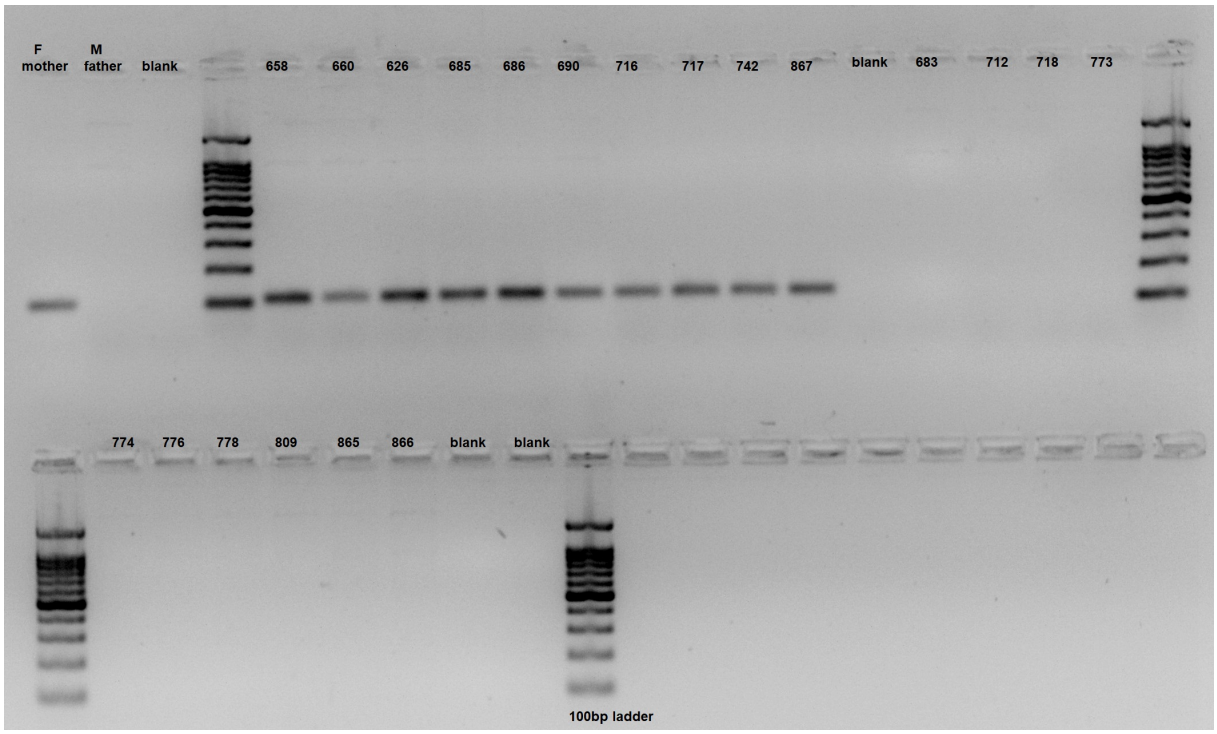


Figure S9. Inheritance of the PCR-marker *AllWSex2* in a genetic family of *H. huso*, with mother (F), father (M) and 10 daughters, followed by 10 sons.

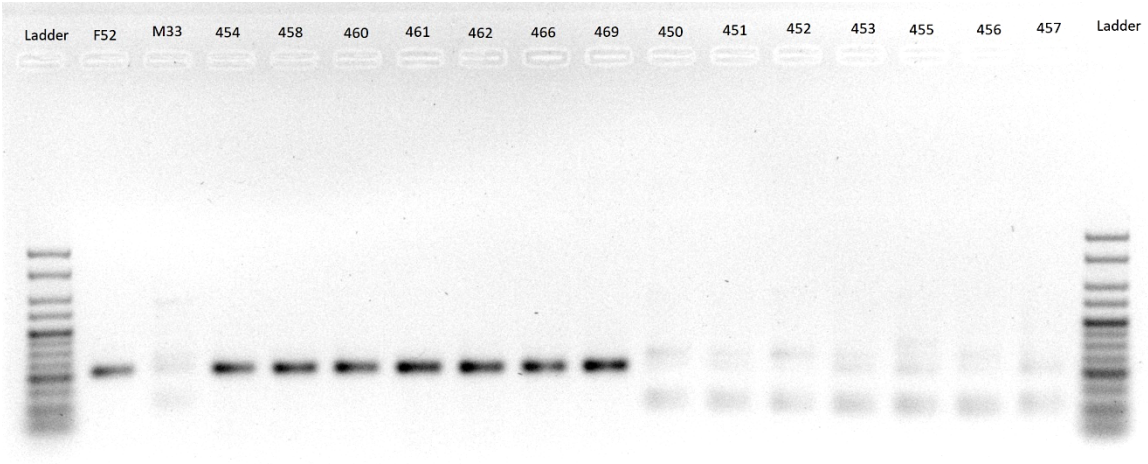


Figure S10. Inheritance of the PCR-marker *AllWSex2* in a genetic family of *A. ruthenus*, with mother (F52), father (M33) and 7 daughters, followed by 7 sons.

Examples of histological sections of gonads from offspring in three genetic families (*A. oxyrinchus, H. huso, A. ruthenus*)

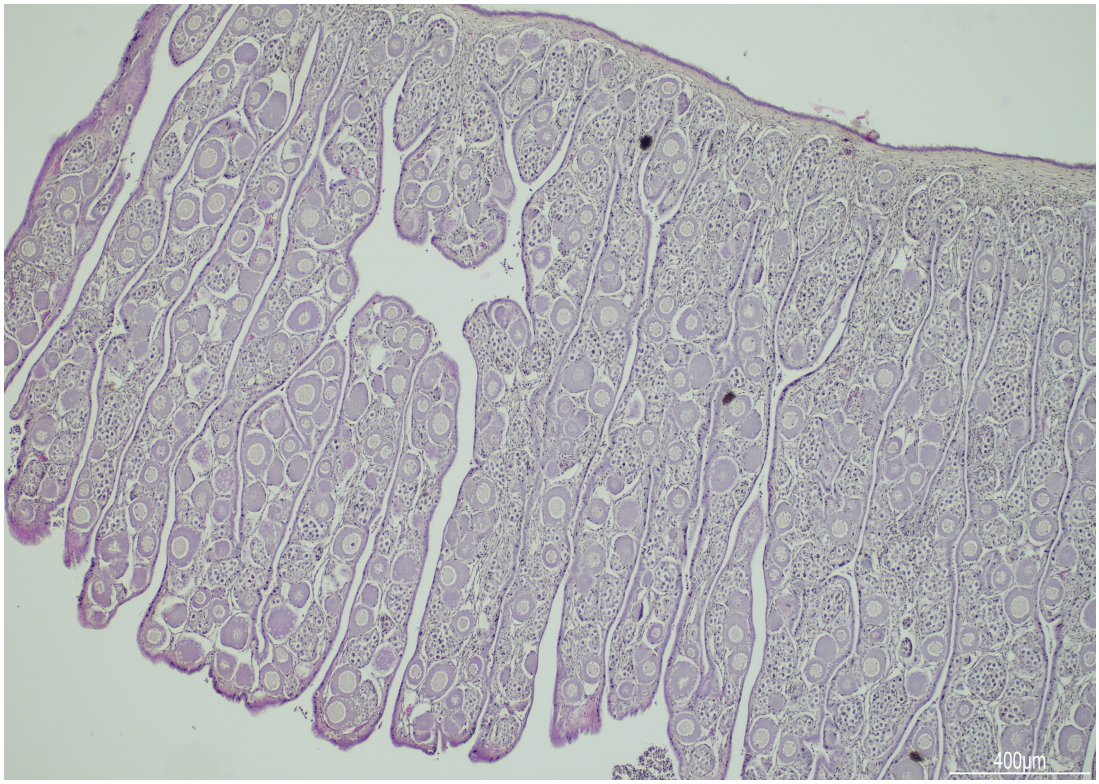


Figure S11. Histology of an *A. oxyrinchus* early female gonad (JG17Ao_F_79_0001.jpg).

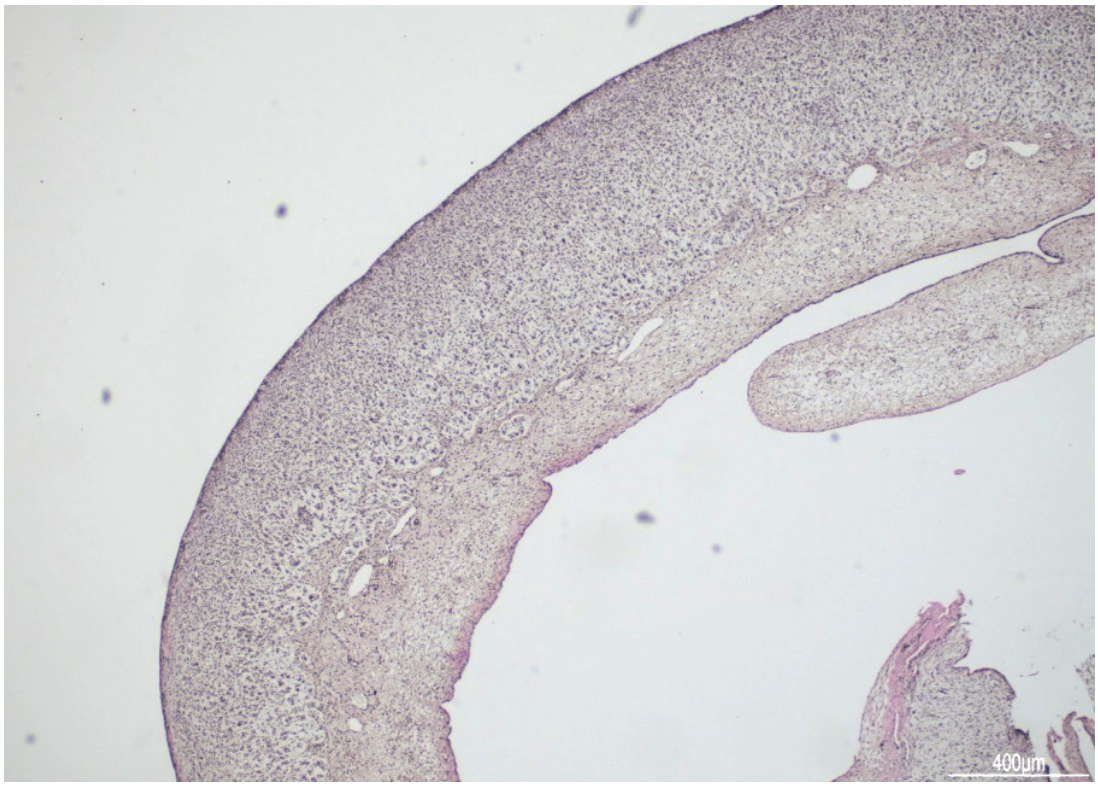


Figure S12. Histology of an *A. oxyrinchus* early male gonad (JG17Ao_M_74_0001.jpg).

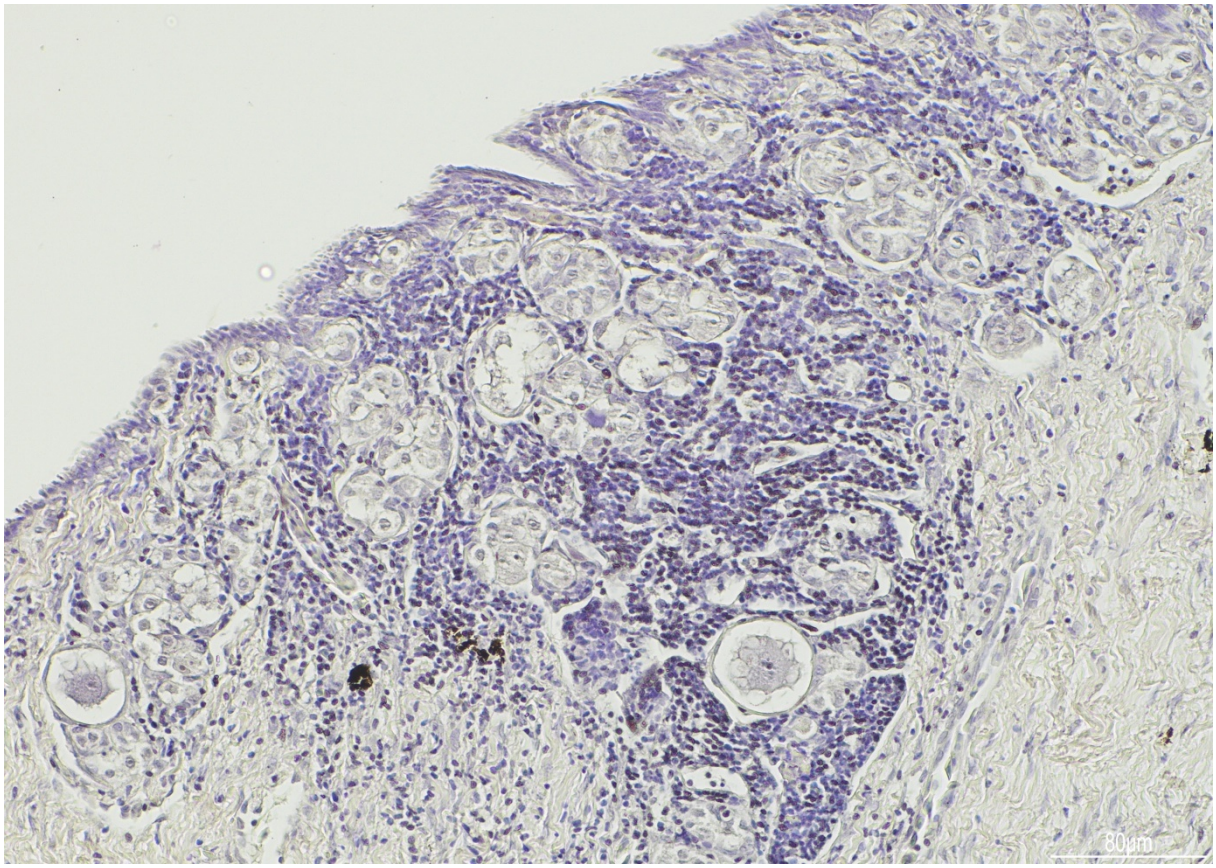


Figure S13. Histology of a *H. huso* early female gonad (Hh_F_H446R626_0006.jpg).

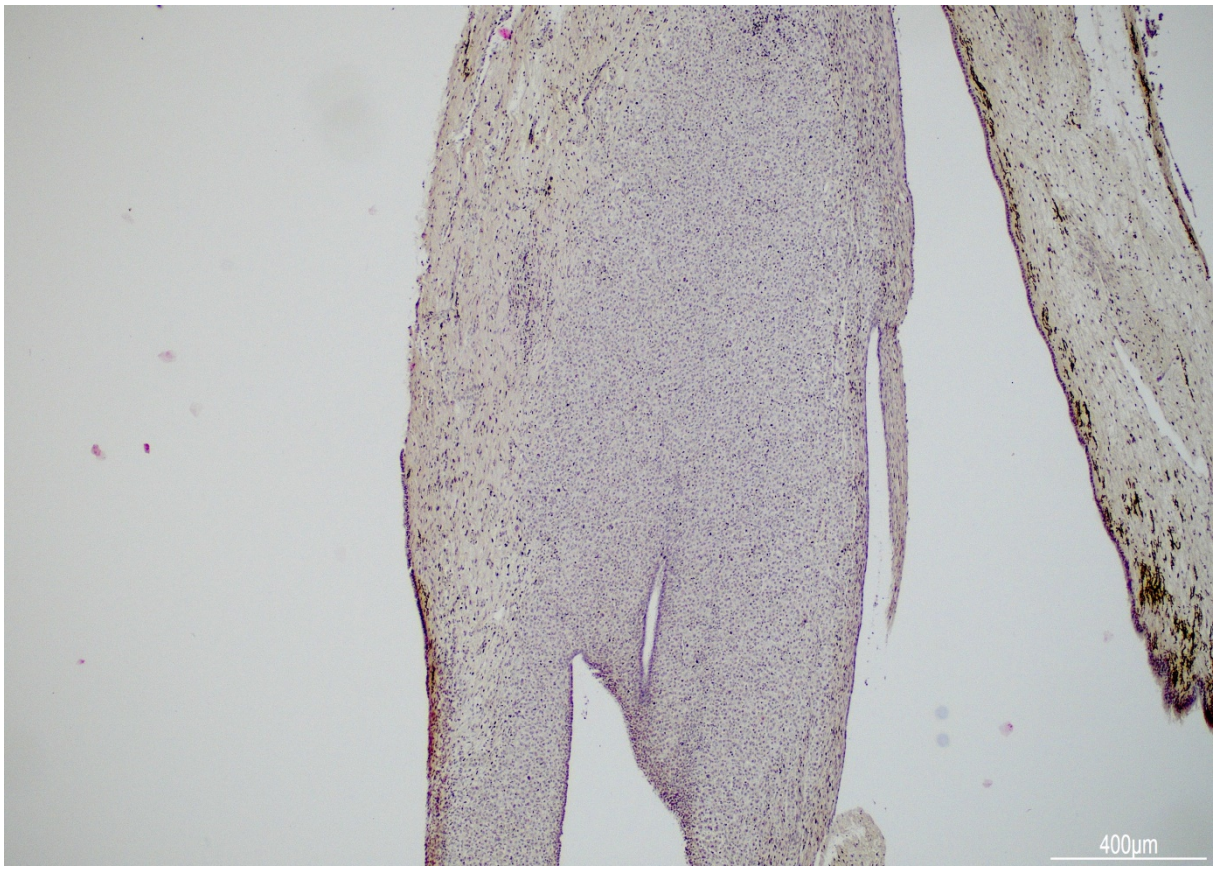


Figure S14. Histology of a *H. huso* early male gonad (Hh_M_H593R773_0001.jpg).

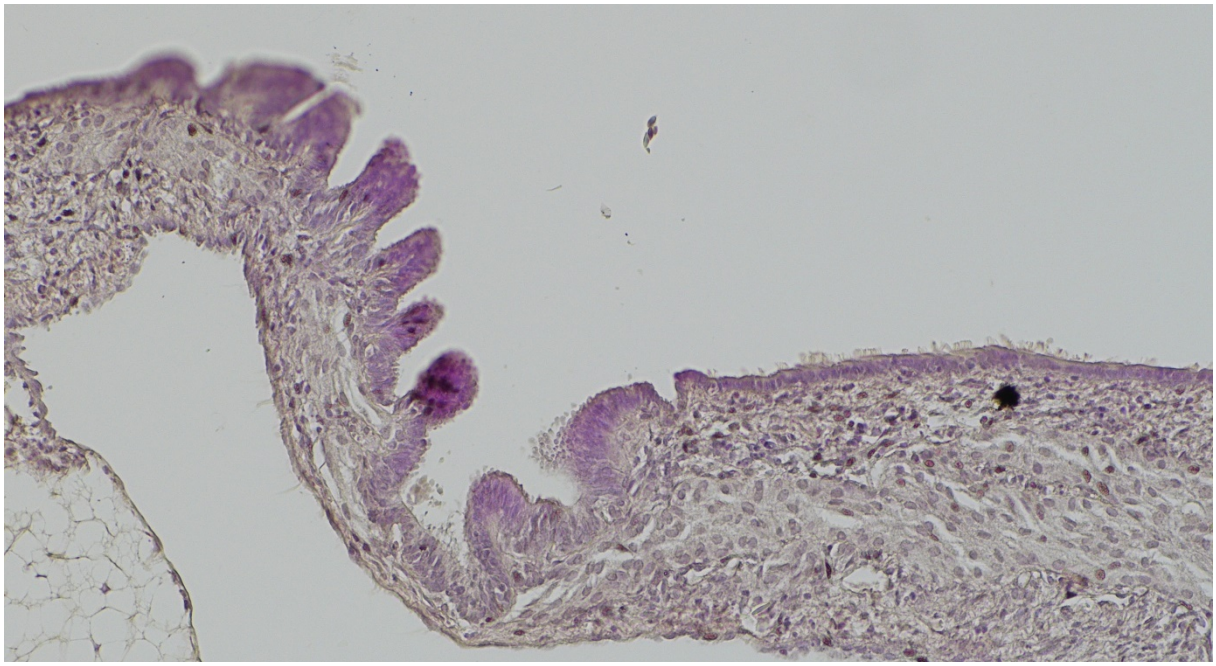


Figure S15. Histology of an *A. ruthenus* early female gonad (AruthX3_F_458_0018.jpg).

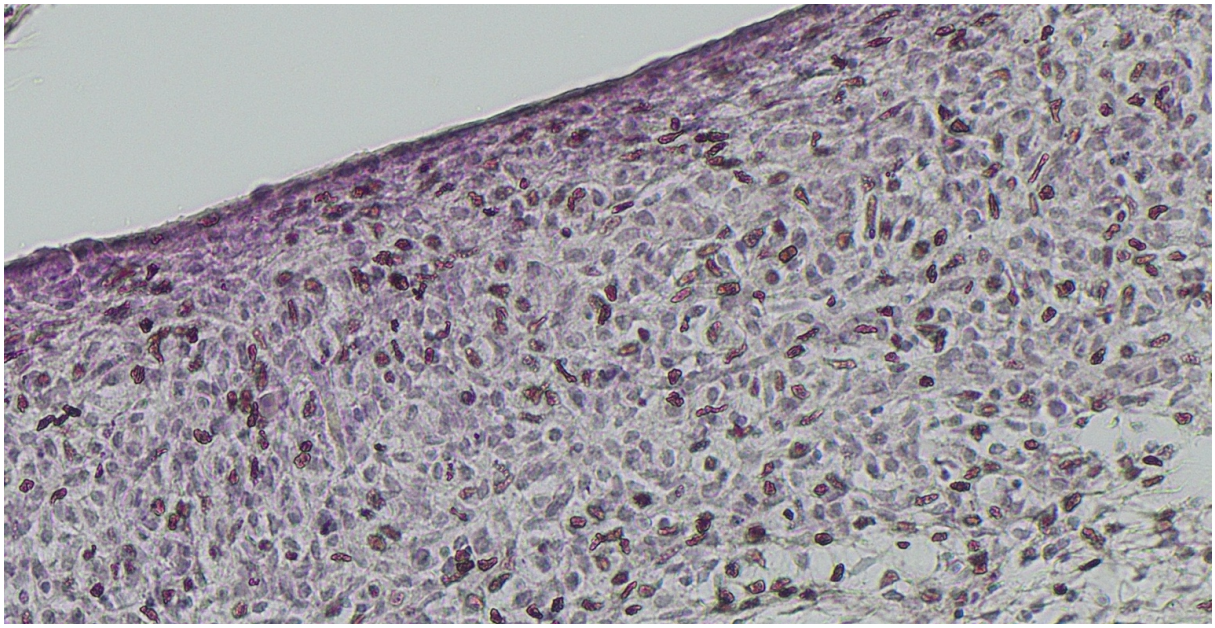


Figure S16. Histology of an *A. ruthenus* early male gonad (AruthX3_M_457.jpg).