

1 **Multi-genome comparisons reveal gain-and-loss evolution**
2 **of the *anti-Mullerian hormone receptor type 2* gene,**
3 **an old master sex determining gene, in Percidae**

4
5 **Running title:** Genomic evolution of sex determination in Percidae

6 Heiner Kuhl^{1*}, Peter T Euclide², Christophe Klopp³, Cedric Cabau⁴, Margot Zahm³, Céline Roques⁵,
7 Carole Iampietro⁵, Claire Kuchly⁵, Cécile Donnadieu⁵, Romain Feron^{6,7}, Hugues Parrinello⁸, Charles
8 Poncet⁹, Lydia Jaffrelo⁹, Carole Confolent⁹, Ming Wen^{10,11}, Amaury Herpin¹⁰, Elodie Jouanno¹⁰,
9 Anastasia Bestin¹², Pierrick Haffray¹², Romain Morvezen¹², Taina Rocha de Almeida¹³, Thomas Lecocq¹³,
10 Bérénice Scherlinger¹³, Dominique Chardard¹³, Daniel Żarski¹⁴, Wes Larson¹⁵, John H. Postlethwait¹⁶,
11 Serik Timirkhanov¹⁷, Werner Kloas¹, Sven Wuertz¹, Matthias Stöck¹, Yann Guiguen^{10*}

12 **Affiliations**

13 ¹ Leibniz-Institute of Freshwater Ecology and Inland Fisheries – IGB (Forschungsverbund Berlin),
14 Müggelseedamm 301/310, D-12587 Berlin, Germany.

15 ² Department of Forestry and Natural Resources | Illinois-Indiana Sea Grant, Purdue University, West
16 Lafayette, Indiana, USA

17 ³ Sigenae, Plateforme Bioinformatique, Genotoul, BioinfoMics, UR875 Biométrie et Intelligence
18 Artificielle, INRAE, Castanet-Tolosan, France

19 ⁴ Sigenae, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France

20 ⁵ INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

21 ⁶ Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland.

22 ⁷ Swiss Institute of Bioinformatics, Lausanne, Switzerland.

23 ⁸ Montpellier GenomiX (MGX), c/o Institut de Génomique Fonctionnelle, 141 rue de la Cardonille,
24 34094, Montpellier Cedex 05, France

25 ⁹ GDEC Gentyane, INRAE, Université Clermont Auvergne, Clermont-Ferrand, France

26 ¹⁰ INRAE, LPGP, 35000, Rennes, France.

27 ¹¹ State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan
28 Normal University, Changsha, China

29 ¹² SYSAAF, Station INRAE-LPGP, Campus de Beaulieu, 35042, Rennes cedex, France

30 ¹³ University of Lorraine, INRAE, UR AFPA, Nancy, France

31 ¹⁴ Department of Gamete and Embryo Biology, Institute of Animal Reproduction and Food Research,
32 Polish Academy of Sciences, ul. Tuwima 10, 10-748, Olsztyn, Poland

33 ¹⁵ National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska
34 Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau, AK, 99801,
35 USA

36 ¹⁶ Institute of Neuroscience, University of Oregon, Eugene, OR, 97403, USA.

37 ¹⁷ Sea Biology LLP, Almaty, Kazakhstan.

38 * corresponding authors:

39 Heiner Kuhl (Heiner.Kuhl@igb-berlin.de) and Yann Guiguen (yann.guiguen@inrae.fr)

40 **ABSTRACT**

41 The Percidae family comprises many fish species of major importance for aquaculture and fisheries.
42 Based on three new chromosome-scale assemblies in *Perca fluviatilis*, *Perca schrenkii* and *Sander*
43 *vitreus* along with additional percid fish reference genomes, we provide an evolutionary and
44 comparative genomic analysis of their sex-determination systems. We explored the fate of a
45 duplicated anti-Mullerian hormone receptor type-2 gene (*amhr2bY*), previously suggested to be the
46 master sex determining (MSD) gene in *P. flavescens*. Phylogenetically related and structurally similar
47 *amhr2* duplications (*amhr2b*) were found in *P. schrenkii* and *Sander lucioperca*, potentially dating this
48 duplication event to their last common ancestor around 19-27 Mya. In *P. fluviatilis* and *S. vitreus*, this
49 *amhr2b* duplicate has been lost while it was subject to amplification in *S. lucioperca*. Analyses of the
50 *amhr2b* locus in *P. schrenkii* suggest that this duplication could be also male-specific as it is in
51 *P. flavescens*. In *P. fluviatilis*, a relatively small (100 kb) non-recombinant sex-determining region (SDR)
52 was characterized on chromosome-18 using population-genomics approaches. This SDR is
53 characterized by many male-specific single-nucleotide variants (SNVs) and no large
54 duplication/insertion event, suggesting that *P. fluviatilis* has a male heterogametic sex determination
55 system (XX/XY), generated by allelic diversification. This SDR contains six annotated genes, including
56 three (*c18h1orf198*, *hsdl1*, *tbc1d32*) with higher expression in testis than ovary. Together, our results
57 provide a new example of the highly dynamic sex chromosome turnover in teleosts and provide new
58 genomic resources for Percidae, including sex-genotyping tools for all three known *Perca* species.

59 **Keywords:** sex-determination, genome, perches, pikeperches, sex-chromosomes

60

61 INTRODUCTION

62 The percid family (Percidae, Rafinesque) encompasses a large number (over 250) of diverse
63 ecologically and economically important fish species, assigned to 11 genera [1]. Two genera, *Perca* and
64 *Sander* are found across both Eurasia and North America, with separate species native to each
65 continent (Eurasia: *Perca fluviatilis* / *Sander lucioperca*; North America: *Perca flavescens* / *Sander*
66 *vitreus*). Percids are classically described as typical freshwater species of the Northern hemisphere,
67 even if some species can be regularly found in brackish waters (e.g. *Sander lucioperca*, *Perca fluviatilis*).
68 In the context of declining fisheries over the past few decades, but also due to their high value and
69 good market acceptance, four percid species - *Perca flavescens* (yellow perch) and *Sander vitreus*
70 (walleye) in North America and *P. fluviatilis* (European perch) and *S. lucioperca* (zander) in Eurasia are
71 particularly promising for aquaculture. Rearing these fish in recirculation aquaculture systems (RAS)
72 allows for a control of reproduction and a year-round production of stocking fish [2,3]. Although year-
73 round production represents an important competitive goal, current production targets premium
74 markets and an up-scaling of production faces several bottlenecks [4].

75 Among these bottlenecks is better control of the sex of developing individuals because in both *Perca*
76 and *Sander* genera, females grow faster than males [5,6]. Due to faster female growth (up to 25-50%
77 in *Perca*, 10% in *Sander*), all-female stocks are highly desirable. In *Perca fluviatilis*, sex determination
78 has been assumed to be male heterogametic (XX/XY) based on gynogenesis or hormonal treatment
79 experiments [7,8]. These methodologies also produced genetic female but phenotypic male individuals
80 (neomales) that can be used to produce all-female stocks by crossing normal XX females with these
81 chromosomally XX neomales. This approach would, however, greatly benefit from a reliable sexing
82 method allowing the identification of genetic sex early during development to select rare genetically
83 XX neomales as future breeders in aquaculture. In *P. flavescens*, an XX female/XY male heterogametic
84 genetic sex determination system has been also recently uncovered, with duplication / insertion of an
85 anti-Mullerian hormone receptor type 2 (*amhr2*) gene as a potential master sex determining gene [9].

86 Genes encoding many members of the transforming growth factor beta (TGF- β) gene family, including
87 anti-Mullerian hormone (*amh*) and anti-Mullerian hormone receptor type-2 (*amhr2*), have repeatedly
88 and independently evolved as master sex-determining (MSD) genes in vertebrates [10]. For instance,
89 *amh* has been characterized or suspected to be the MSD gene in pikes [11,12], Nile tilapia [13],
90 lumpfish [14], *Sebastes* rockfish [15], lingcod [16], and Patagonian pejerrey [17]. The cognate receptor
91 gene of *Amh*, *amhr2*, has also been found as a potential MSD gene in *Pangasiidae* [18], *Takifugu* [19],
92 *Ayu* [20], common seadragon and alligator pipefish [21], as well as in yellow perch [9]. The repeated
93 and independent recruitment of TGF- β receptors, including *Amhr2*, in teleost fish sex determination is
94 even more puzzling as many of these MSD genes, encoding a TGF β receptor, share a similar N-terminal

95 truncation [9,18,21], supporting their evolution towards a ligand-independent mechanism of action
96 [18]. Therefore, the extent of evolutionary conservation of the Y-linked *amhr2bY* gene found in yellow
97 perch in closely related species (genus *Perca* and *Sander*) is an important question with implications
98 for better sex-control in these aquaculture species and for understanding the evolution of sex linkage
99 and protein structure.

100 Regarding genomics of Percidae, two long-read reference quality genome assemblies have recently
101 been published for *P. flavescens* and *S. lucioperca* [9,22]. While for *P. fluviatilis* and *S. vitreus* only draft
102 genomes, generated from short-read sequencing, have been available [23]. Here, we provide three
103 new long-read chromosome-scale genome assemblies for *P. fluviatilis*, *P. schrenkii* and *Sander vitreus*
104 and thus complete genomic resources for the economically most important species of Percidae. These
105 data enabled us to develop PCR-assays for sexing of all three *Perca* species and shed light on gene gain-
106 and-loss in the evolution of an old MSD gene in Percidae.

107

108 MATERIAL AND METHODS

109 *Biological samples*

110 In *Perca fluviatilis*, high molecular weight (HMW) genomic DNA (gDNA) for genome sequencing was
111 extracted from a blood sample of a male called “Pf_M1” (BioSample ID SAMN12071746) from the
112 aquaculture facility of the University de Lorraine, Nancy, France. Blood (0.5 ml) was sampled and
113 directly stored in 25 ml of a TNES-Urea lysis buffer (TNES-Urea: 4 M urea; 10 mM Tris-HCl, pH 7.5; 125
114 mM NaCl; 10 mM EDTA; 1% SDS). HMW gDNA was extracted from the TNES-urea buffer using a slightly
115 modified phenol/chloroform protocol as described [12]. For the chromosome contact map (Hi-C), 1.5
116 ml of blood was taken from the same animal and slowly (1 K/min) cryopreserved with 15 % dimethyl
117 sulfoxide (DMSO) in a Mr. Frosty Freezing Container (ThermoFisher) at -80°C. Additional fin clip
118 samples for RAD-Sequencing (RAD-Seq), Pool-Sequencing (Pool-Seq) or sex-genotyping assays were
119 collected and stored in 90% ethanol, either at the Lucas Perche aquaculture facility (Le Moulin de Cany,
120 57170 Hampont, France), at Kortowskie Lake in Poland, or at Mueggelsee Lake in Germany.

121 Samples of *Perca schrenkii* were obtained for genome sequencing and sex genotyping from male and
122 female wild catches at lake Alakol, Kazakhstan (46.328 N, 81.374 E). Different organs and tissues (brain,
123 liver, muscle, ovary, testis) were sampled for genome and transcriptome sequencing (Biosample ID
124 SAMN15143703) and stored in RNAlater. HMW gDNA for genome sequencing was extracted from
125 brain tissue of the male *P. schrenkii* individual, using the MagAttract HMW DNA Kit (Qiagen, Germany).

126 Total RNA for transcriptome sequencing was isolated using a standard Trizol protocol, in combination
127 with the RNAeasy Mini Kit (Qiagen, Germany).

128 For genome sequencing of *Sander vitreus* a fin clip of a male was sampled by Ohio Department of
129 Natural Resources (Ohio, DNR) in spring 2017 and stored in 96% ethanol. The *S. vitreus* sample called
130 "19-12246" originated from Maumee River, Ohio [41.554 N; -83.6605W]. DNA was extracted using the
131 DNeasy Tissue Kit (Qiagen). Short DNA fragments were removed/reduced by size-selective, magnetic-
132 bead purification using 0.35x of sample volume AMPure beads (Beckmann-Coulter) and two washing
133 steps with 70% ethanol.

134 ***Sequencing***

135 Genomic sequencing of *P. fluviatilis* was carried out using a combination of 2x250 bp Illumina short-
136 reads, Oxford Nanopore long reads and a chromosome contact map (Hi-C). For long-read sequencing,
137 DNA was sheared to 20 kb using the megaruptor system (Diagenode). ONT (Oxford nanopore
138 technologies) library preparation and sequencing was performed using 5 µg of sheared DNA and
139 ligation sequencing kits SQK-LSK108 or SQK-LSK109, according to the manufacturer's instructions. The
140 libraries were loaded at a concentration of 0.005 to 0.1 pmol and sequenced for 48 h on 11 GridION
141 R9.4 or R9.4.1 flowcells. Short read wgs (whole genome shotgun) sequencing for consensus polishing
142 of noisy long read assemblies was carried out by shearing the HMW DNA to approximately 500 bp
143 fragments and using the Illumina Truseq X kit, according to the manufacturer's instructions. The library
144 was sequenced using a read length of 250 bp in paired-end mode (HiSeq 3000, Illumina, California,
145 USA). Hi-C library generation for chromosome assembly was carried out according to a protocol
146 adapted from Rao *et al.* 2014 [24]. The blood sample was spun down, and the cell pellet was
147 resuspended and fixed in 1% formaldehyde. Five million cells were processed for the Hi-C library. After
148 overnight-digestion with *Hind*III (NEB), DNA-ends were labeled with Biotin-14-DCTP (Invitrogen), using
149 Klenow fragment (NEB) and re-ligated. A total of 1.4 µg of DNA was sheared to an average size of 550
150 bp (Covaris). Biotinylated DNA-fragments were pulled down using M280 Streptavidin Dynabeads
151 (Invitrogen) and ligated to PE adaptors (Illumina). The Hi-C library was amplified using PE primers
152 (Illumina) with 10 PCR amplification cycles. The library was sequenced using a HiSeq3000 (Illumina,
153 California, USA), generating 150 bp paired-end reads.

154 Genomic sequencing of *P. schrenkii* and *S. vitreus* was carried out using Oxford Nanopore long reads
155 on a MinION nanopore sequencer (Oxford Nanopore Technologies, UK) in combination with the MinIT
156 system. Several libraries were constructed using the tagmentation-based SQK-RAD004 kit with varying
157 amounts of input DNA (0.4 to 1.2 µg) from a male individual or using the ligation approach of the SQK-
158 LSK109 kit (input DNA 2 µg). Libraries were sequenced on R9.4.1 flowcells with variable run times and

159 exonuclease washes by the EXP-WSH003 kit to remove pore blocks and improve the data yield. Short-
160 read wgs-sequencing of *P. schrenkii* was conducted at BGI (BGI Genomics Co., Ltd.). A *P. schrenkii* male
161 and a female wgs library (300 bp fragment length) were constructed and paired end reads of 150 bp
162 length were generated on an Illumina Hiseq4000 system. Public short-read wgs data of *S. vitreus* were
163 obtained from the NCBI Sequence Read Archive (SRA) using the accession SRR9711286. Transcriptome
164 sequencing of six *P. schrenkii* samples (female brain, male brain, male muscle, female liver, ovary and
165 testis) was conducted at BGI. Transcriptome-sequencing libraries were constructed from total RNA,
166 applying enrichment of mRNA with oligo(dT) hybridization, mRNA fragmentation, random hexamer
167 cDNA synthesis, size selection and PCR amplification. Sequencing of 150 bp paired-end reads was
168 performed by an Illumina HiSeq X Ten system.

169

170 ***Genome assembly of Perca fluviatilis***

171 Residual adaptor sequences in ONT GridION long reads were trimmed and split by Porechop (v0.2.1)
172 [25]. Reads longer than 9999 bp were assembled by SmartDeNovo (May-2017) [26] using default
173 parameters. Long reads were remapped to the SmartDeNovo contigs by Minimap2 (v2.7) [27] and
174 Racon (v1.3.12) [28] was used to polish the consensus sequence. In a second round of polishing,
175 Illumina short-reads were mapped by BWA mem (v0.7.12-r1039) [29] to the contigs, which were
176 subsequently polished by Pilon (v1.223) [30]. The chromosome-scale assembly was performed by
177 mapping Hi-C data to the assembled contigs, using the Juicer pipeline (v1.5.6) [31] and subsequent
178 scaffolding by 3D-DNA (v180114) [32]. Juicebox (v1.8.8) [33] was used to manually review and curate
179 the chromosome-level scaffolds. A final gap-closing step, applying long reads and LR_gapcloser (v1.1,
180 default parameters) [34], further increased contig length. After gap-closing, a final consensus sequence
181 polishing step was performed by mapping short reads to the scaffolds, sequence variants (1/1
182 genotypes were considered as corrected errors) were detected with Freebayes (v0.9.7) [35] and
183 written to a vcf-file. The final fasta file was then generated by vcf-consensus from Vcftools (v0.1.15,
184 default parameters).

185 ***Genome assembly of Perca schrenkii and Sander vitreus***

186 Illumina short reads were trimmed using Trimmomatic (v0.35) [36]. Short reads were assembled using
187 a custom compiled high kmer version of idba-ud (v1.1.1) [37] with kmer size up to 252. The resulting
188 contigs were mapped against available Percidae genomes (*P. flavescens*, *P. fluviatilis* and *S. lucioperca*)
189 by Minimap2 and analysis of overall mapped sequence length resulted in *P. schrenkii* aligned best with
190 *P. flavescens* and *S. vitreus* aligned best with *S. lucioperca*. According to the benchmarks published in
191 [38], the publicly available chromosome-level assembly of *P. flavescens* (RefSeq: GCF_004354835.1)

192 could be used to aid the chromosome assembly of *P. schrenkii* as follows: ONT MinION long reads
193 (male sample) were trimmed and split using Porechop (v0.2.1) [25]. The inhouse developed CSA
194 method (v2.6) [38], was used to assemble the *P. schrenkii* genome from long-read data and short-read
195 contigs and to infer chromosomal scaffolds using the *P. flavescens* reference genome. CSA parameters
196 were optimized to account for relatively low long-read sequencing coverage and hybrid assembly of
197 long reads and short-read contigs:

198 CSA2.6.pl -r longreads.fa.gz -g *P. flavescens*.fa -k 19 -s 2 -e 2 -l „, -i shortreadcontigs.fa -L3000 -A“

199 Similarly, we assembled *S. vitreus*, using the *S. lucioperca* contigs and *P. flavescens* chromosomes as
200 references for chromosomal assembly. Here, the short-read contigs were treated as long-reads:

201 CSA2.6c.pl -r longreads+contigs.fa.gz -g *sanLuc.CTG*.fa.gz, *PFLA_1.0_genomic*.fna.gz -k 19 -s 2 -e 2

202 The assemblies were manually curated, and the consensus sequences were polished using long reads
203 and flye (v2.6) [39], with options: --nanoraw --polish-target, followed by two rounds of polishing by
204 Pilon (v1.23) [30], using the short-read data, which had been mapped by Minimap2 (v2.17-r941) [27],
205 to the genome assemblies.

206 ***Genome annotation***

207 *De novo* repeat annotation was performed using RepeatModeler (version open-1.0.8) and Repeat
208 Masker (version open-4.0.7). The *P. fluviatilis* genome has been assigned to the RefSeq assembly
209 section of NCBI and has been annotated by GNOMON
210 (www.ncbi.nlm.nih.gov/genome/annotation_euk/process), which included evidence from
211 Actinopterygii proteins (n=154,659) and *P. fluviatilis* RNAseq reads (n = 3,537,868,978)
212 (www.ncbi.nlm.nih.gov/genome/annotation_euk/Perca_fluviatilis/100). To annotate our *P. schrenkii*
213 and *S. vitreus* assemblies, we used the high-quality GNOMON annotations from their closest relatives
214 *P. flavescens* (www.ncbi.nlm.nih.gov/genome/annotation_euk/Perca_flavescens/100) and
215 *S. lucioperca* (www.ncbi.nlm.nih.gov/genome/annotation_euk/Sander_lucioperca/101), respectively.
216 We performed high-throughput comparative protein coding gene annotation by spliced alignment of
217 GNOMON mRNAs and proteins by Spaln (v2.06f, [40]) to our assemblies and combined the resulting
218 CDS- and UTR-matches into complete gene models by custom scripts. All annotations were
219 benchmarked using BUSCO [41] with the Actinopterygii_odb9 database and obtained highly similar
220 values as the reference annotations used for the comparative annotation approach.

221 ***Genome browsers and data availability***

222 We provide UCSC genome browsers [42] for the five available *Perca* and *Sander* reference genomes
223 (this study: *P. fluviatilis*, *P. schrenkii*, *S. vitreus*; earlier studies: *P. flavescens* [9] and *S. lucioperca* [22]
224 at <http://genomes.ibg-berlin.de/Percidae/>. These genome browsers provide access to genomic
225 sequences and annotations (either public NCBI GNOMON annotations or annotations resulting from
226 our comparative approach). Blat [43] servers for each genome are available to align nucleotide or
227 protein sequences.

228 ***Phylogenomics and divergence time estimation***

229 We performed pair-wise whole-genome alignments of 36 teleost genome assemblies as in [44], using
230 Last-aligner and Last-split [45] for filtering 1-to-1 genome matches, Multiz [46] for multiple alignment
231 construction from pairwise alignments and filtered for non-coding sequences to calculate the species
232 tree using iqtree2 and raxml-ng [47,48]. We added the genomes of *P. schrenkii*, *S. vitreus* and
233 *Etheostoma spectabile* (GCF_008692095.1) to this dataset and re-analyzed the highly-supported
234 subclade containing Percidae species using several outgroups (*Lates*, *Oreochromis*, *Pampus* and
235 *Thunnus* sp.). We estimated divergence times using a large subset of our multiple alignment (10^6 nt
236 residues) and the approximate method of Mcmctree (Paml package version, [49]). We calibrated 5
237 nodes of the tree by left or right CI values, obtained from www.timetree.org and applied independent
238 rates or correlated rates clock models and the HKY85 evolutionary model. Approximately 10^8 samples
239 were calculated, of which we used the top 50% for divergence time estimation. Each calculation was
240 performed in two replicates, which were checked for convergence using linear regression. The final
241 tree was plotted using FigTree (v1.4.4, <http://tree.bio.ed.ac.uk/software/figtree>).

242 ***Perca fluviatilis RAD-Sequencing***

243 *Perca fluviatilis* gDNA samples from 35 males and 35 females were extracted with the NucleoSpin Kit
244 for Tissue (Macherey-Nagel, Duren, Germany), following the manufacturer's instructions. Then, gDNA
245 concentrations were quantified with a Qubit3 fluorometer (Invitrogen, Carlsbad, CA) using a Qubit
246 dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA). RAD libraries were constructed from each individual's
247 gDNA, using a previously described protocol with the single *Sbf1* restriction enzyme [50]. These
248 libraries were sequenced on an Illumina HiSeq 2500. Raw reads were demultiplexed using the
249 process_radtags.pl wrapper script of stacks, version 1.44, with default settings [51], and further
250 analyzed with the RADSex analysis pipeline [52] to identify sex-specific markers.

251 ***Perca fluviatilis Pool-Sequencing***

252 Sequencing of pooled samples (Pool-Seq) was carried out in *Perca fluviatilis* to increase the resolution
253 of RAD-Sequencing for the identification of sex-specific signatures characteristic of its sex-determining

254 region. The gDNA samples used for RAD-Sequencing were pooled in equimolar quantities according to
255 their sex. Pooled male and pooled female libraries were constructed using a Truseq nano kit (Illumina,
256 ref. FC-121-4001) following the manufacturer's instructions. Each library was sequenced in an Illumina
257 HiSeq2500 with 2x 250 reads. Pool-Seq reads were analyzed as previously described [9,11,53–55] with
258 the PSASS pipeline (psass version 2.0.0: https://zenodo.org/record/2615936#.XtyIS3s6_AI) that
259 computes the position and density of single nucleotide variations (SNVs), heterozygous in one sex but
260 homozygous in the other sex (sex-specific SNVs), and the read depths for the male and female pools
261 along the genome to look for sex coverage differences. Psass was run with default parameters except
262 –window-size, which was set to 5,000, and –output-resolution, which was set to 1,000.

263 **PCR-based sex diagnostics**

264 A *Perca schrenkii* PCR-based sex-diagnostic test was designed based on multiple alignments of the
265 different *amhr2* genes in *P. fluviatilis* (one autosomal gene only), *P. flavescens* (two genes), and *Perca*
266 *schrenkii* (two genes) to target a conserved region for all *Perca amhr2* genes, allowing the design of
267 PCR-primers that amplify both the autosomal *amhr2a* and the male-specific *amhr2bY* with different
268 and specific PCR-amplicon sizes. Selected PCR primer sequences were forward: 5'-
269 AGTTTATTGTGTTAGTTGGCT-3' and reverse: 5'-CAAATAAATCAGAGCAGCGCATC-3'. PCRs were
270 carried out with 1U Platinum Taq DNA Polymerase and its corresponding Buffer (Thermofisher)
271 supplemented with 0.8 mM dNTPs (0.2mM each), 1.5 mM MgCl₂ and 0.2 μM of each primer with the
272 following cycling conditions, 96°C for 3 min; 40 cycles of denaturation (96°C, 30 s), annealing (54°C, 30
273 s) and extension (72°C, 1 min); final extension (72°C, 5 min); storage at 4°C. PCR amplicons were
274 separated on 1.5% agarose gels (1.5% std. agarose, 1x TBE buffer, 5 V/cm, running time 40 min) and
275 the systematic amplification of the autosomal (*amhr2a*) amplicon was used as a positive PCR control.

276 *Perca fluviatilis* primers were designed to amplify a 27 bp-deletion variant in the third intron of the
277 *P. fluviatilis* *hsd1* gene, which was identified as a male specific (Y-specific) variation based on the pool-
278 seq analysis. Selected PCR-primer sequences were forward 5'-ACACTGATCACACATTCTGTCTCA-3' and
279 reverse 5'-TGTTAACATTGAGAATTTCGCCTT-3'. PCRs were carried out as described above with the
280 following cycling conditions: denaturation 96°C for 3 min; 40 cycles of denaturation (96°C, 30 s),
281 annealing (60°C, 30 s) and extension (72°C, 30 min); final extension (72°C, 5 min); storage at 4°C. PCR
282 amplicons were separated on 5% agarose gels (5% Biozym sieve 3:1 agarose, 1x TBE buffer, 5 V/cm,
283 1 h 40 min running time) and the amplicon derived from the amplification of the X-chromosome allele
284 was used as a positive PCR control. In addition to this classical PCR sex-genotyping method, we also
285 explored the sex-linkage of some sex-specific SNVs in *P. fluviatilis* using Kompetitive Allele-Specific
286 Polymerase chain reaction (KASPar) assays [56]. Seven sex-specific SNVs were selected at different
287 locations within the *P. fluviatilis* sex-determining region. Primers (Table 1) were designed using the

288 design service available on the 3CR Bioscience website (www.3crbio.com/free-assay-design). KASPar
289 genotyping assays were carried out with a single end-point measure on a Q-PCR Light Cycler 480
290 (Roche) using the Agencourt® DNAAdvance kit (Beckman), following the manufacturer's instructions.

291 ***RNAseq expression analyses***

292 Already available gonadal datasets of *P. fluviatilis* (age 9 month) RNAseq [57] (SRA accessions:
293 SRR14461526 and SRR14461527) were used to compare gene expression between ovary and testes
294 for the gene models annotated in the SD-region. Reads were mapped to our *P. fluviatilis* reference
295 genome using HISAT2 [58] and transcript assembly and FPKM-values were calculated using STRINGTIE
296 [59].

297

298 **RESULTS**

299 ***Genome assemblies of Perca fluviatilis, Perca schrenkii and Sander vitreus***

300 The genome of *P. fluviatilis* was sequenced to high coverage using Oxford Nanopore long-read
301 sequencing (estimated coverage: 67-fold / N50 read length: 11.9 kbp), Hi-C data was generated to
302 allow for chromosome-level assembly (coverage: 52-fold / alignable pairs 89.1% / Hi-C map see Suppl.
303 Fig. 1). The final assembly yielded a highly complete reference genome (99.0% of sequence assigned
304 to 24 chromosomes (N50 length: 39.6 Mbp) and highly continuous contigs (N50 length: 4.1 Mbp).
305 Compared to a previously published genome assembly of *P. fluviatilis*, obtained from “linked-short-
306 reads” (10X Genomics), these numbers represent a 316-fold improvement of contig continuity and a
307 6.3-fold increase of scaffold continuity. The better continuity resulted in an increased percentage of
308 predictable genes (BUSCO results below). Genome assembly statistics are also highly congruent with
309 the previously published reference quality *P. flavesiensis* genome, except for obvious size differences
310 as the *P. fluviatilis* assembly is about 8.1% larger than the *P. flavesiensis* assembly and represents the
311 largest genome known in the genus *Perca* (Table 2).

312 The genome of *P. schrenkii* was assembled by a hybrid assembly method, which was highly efficient
313 regarding long-read sequencing coverage and read length needed (here only 30-fold / N50 read length:
314 4.95 kb). *De novo* assembled contigs from short reads were combined with long reads and scaffolded
315 using our CSA-pipeline [38], with the *P. flavesiensis* as the closest reference genome (Fig. 1: divergence
316 time about 7.1 Mya) for genome comparison and inferring chromosomal-level sequences. Using this
317 approach, we were able to assemble the genome of *P. schrenkii* to similar quality as those obtained
318 for *P. flavesiensis* and *P. fluviatilis* (94.7% assigned to 24 chr. / contig N50 length: 3.2 Mbp; Table 2). The

319 genome assembly size of *P. schrenkii* was in between the other two *Perca sp.* genomes
320 (877 Mbp < 908 Mbp < 951 Mbp).

321 The genome of *Sander vitreus* was assembled from long reads (coverage: 12-fold / N50 read length:
322 10 kb) and short reads similar to procedures for *P. schrenkii* but we used two reference genomes for
323 chromosomal assembly. First, contigs of *S. lucioperca* (closest relative, divergence time 9.8 Mya) served
324 to order the *S. vitreus* contigs (N50: 6.2 Mb), which improved scaffold N50 significantly (result. N50:
325 16.8 Mb), then these nearly chromosome-scale scaffolds were ordered according to the *P. flavescens*
326 (div. time 19.1 Mya) Hi-C chromosomes (result. N50: 33.3 Mb / 96.5% assigned to 24 chromosomes).
327 This two-step approach resulted in more consistent results than just using the *S. lucioperca*
328 chromosomal scaffolds, which were generated by genetic linkage mapping. We observed similar
329 genome size differences as in *Perca sp.* between both *Sander* species. The *S. vitreus* genome assembly
330 (791 Mb) is smaller than the one of *S. lucioperca* (901 Mb), thus the North American *Sander* and *Perca*
331 species tend to have smaller genome sizes than their Eurasian relatives (Table 2).

332 *De novo* repeat analysis showed that 60.1% of the genome assembly size difference between *Sander*
333 *vitreus* and *S. lucioperca* could be explained by repeat expansion/reduction. Similarly, for *Perca*
334 *flavescens* and *P. fluviatilis* about 64.8% of the genome size differences were due to repeat
335 expansion/reduction. In both species pairs most repeat expansions/reductions were observed in
336 repeat elements classified as “unknown”. Regarding annotated repeat element classes, L2, DNA,
337 Helitron, CMC-EnSpm, hAT, Rex-Babar, hAT-Charlie and PiggyBac elements expanded the most in both
338 Eurasian species (together adding roughly 20 Mbp to the genomes of *S. lucioperca* and *P. fluviatilis*),
339 while a clear expansion of only a single repeat element, called RTE-BovB, was found in both North
340 American species (adding about 3 and 7 Mbp of sequence to *S. vitreus* and *P. flavescens*, respectively;
341 Suppl. table 1).

342 The genome of *P. fluviatilis* was annotated by NCBI/GNOMON, which included ample public RNAseq
343 data and protein homology evidence. For *P. schrenkii* and *S. vitreus*, we transferred the
344 NCBI/GNOMON annotations of *P. flavescens* and *S. lucioperca*, respectively. BUSCO analysis (Table 3)
345 revealed values larger than 95.9% for complete BUSCOs (category “C:”) for all annotations. The
346 comparative annotation approach resulted only in small losses (category “M:”) of a few BUSCO genes
347 in the range of 0.4% - 1.1%. In this regard, the *S. vitreus* assembly performed better than the *P. schrenkii*
348 assembly, possibly due to the higher N50 read length of the underlying long-read data.

349 ***Percomorpha phylogenomics and divergence time estimation***

350 To calculate the phylogenetic tree of 36 Percomorpha species and their divergence times, we aligned
351 whole genomes and extracted the non-coding alignments (Fig. 1). The use of non-coding sequences is

352 preferable to calculate difficult-to-resolve phylogenetic trees that occur after massive radiations [60–
353 62]. Our multiple alignment consisted of 6,594,104 nt residues (2,256,299 distinct patterns; 1,652,510
354 parsimony-informative; 1,136,496 singleton sites; 3,805,098 constant sites) and resulted in a highly-
355 supported tree (raxml-ng and IQtree2 topologies were identical; all IQtree2's SH-aLRT and ultrafast
356 bootstrapping (UFBS) values were 100). According to the divergence time analysis, most Percomorpha
357 orders emerged after the Cretaceous–Paleogene (K-Pg) boundary about 65.9 Mya ago (CI: 51.3 - 83.6).
358 The lineage leading to the Percidae (represented with species of *Perca*, *Sander*, and *Etheostoma*)
359 emerged about 58.9 Mya (CI: 45.8 - 74.2), and the extant Percidae species analyzed in this study
360 diverged from a last common ancestor (LCA) about 26.9 Mya (CI: 16.8 - 43.4). The *Perca* and *Sander*
361 genera split about 19.1 Mya (CI: 11.8 -31.1), and *S. vitreus* and *S. lucioperca* splitted at 9.8 Mya (CI: 5.7
362 - 18.9) similar to the divergence of *P. fluviatilis* from *P. flavescens* and *P. schrenkii* 10.9 Mya (CI: 7.1 -
363 18.6). The closest *Perca* species are *P. flavescens* and *P. schrenkii* which diverged about 7.1 Mya (CI:
364 4.2 - 11.4), although today both species are occurring in completely different global ranges.

365 ***The fate of amhr2 genes during evolution of Perca and Sander species***

366 In the genome of *P. flavescens*, two *amhr2* paralogs were previously described, i.e., an autosomal gene,
367 *amhr2a*, present in both males and females on chromosome 04 (Chr04), and a male-specific
368 duplication on the Y-chromosome (Chr09), *amhr2bY* [9]. A similar *amhr2* gene duplication was also
369 found in our male *P. schrenkii* assembly, and sequence homologies and conserved synteny analyses
370 show that these two *P. schrenkii* *amhr2* genes are orthologs of *P. flavescens* *amhr2a* and *amhr2bY*,
371 respectively (Fig. 2A and AB). Genotyping of one male and one female also suggests that the *amhr2bY*
372 gene could be male-specific in *P. schrenkii* (Suppl. Fig. 2, Table 4), as described in *P. flavescens* [9]. This
373 potential sex-linkage is also supported by a half coverage of reads in the genomic region containing
374 the *amhr2bY* locus in our male *P. schrenkii* genome assembly (Suppl. Fig. 3), in agreement with the
375 hemizygosity of a male-specific region on the Y in species with a XX/XY sex determination system.
376 Alignment of the *P. schrenkii* *amhr2bY* ortholog shows that its coding sequence (CDS) shares 98%
377 identity with the *P. flavescens* *amhr2bY* CDS and 95.7 % identity at the protein level the *P. flavescens*
378 *Amhr2bY*. As in *P. flavescens* [9], the *P. schrenkii* *amhr2bY* gene encodes a N-terminal-truncated type
379 II receptor protein that lacks most of the cysteine-rich extracellular part of the receptor, which is
380 crucially involved in ligand-binding specificity [63] (Fig. 3).

381 In contrast, in our *P. fluviatilis* male genome assembly, only one copy of *amhr2* could be identified and
382 sequence homologies and conserved synteny analysis (Fig. 2A and 2B) shows that this *amhr2* gene is
383 the ortholog of *P. flavescens* and *P. schrenkii* *amhr2a*. In addition, PCR with primers designed to amplify
384 both *amhr2a* and *amhr2bY* from *P. flavescens* and *P. schrenkii* did not show any sex-differences in *P.*
385 *fluviatilis*. Altogether, these results support the absence of an *amhr2b* gene in *P. fluviatilis*.

386 In the *Sander lucioperca* male genome assembly, four copies of *Amhr2* were detected. Using the same
387 primers as those used for the amplification of both *amhr2a* and *amhr2bY* in *P. flavescens* and
388 *P. schrenkii*, PCR genotyping on males and females of *S. lucioperca* produced complex amplification
389 patterns with multiple bands and no visible association with sex. In the publicly available genome
390 assembly of *Sander vitreus*, sequence homologies and /or conserved synteny analysis (Fig. 2A and 2B)
391 allowed the identification of a single autosomal *amhr2a* gene.

392 A phylogenetic analysis of the sequences with similarity to *amhr2* in *Perca* and *Sander* (Fig. 2A and 2B)
393 shows that the different *amhr2* genes likely originated from a gene duplication event that happened
394 in the branch leading to the last common ancestor of these species, dated around 19-27 Mya. Since
395 that time, the *amhr2bY* gene has been conserved in *P. schrenkii* and *P. flavescens*, lost in *P. fluviatilis*
396 and *S. vitreus*, and amplified in *S. lucioperca*.

397 ***Evolution of sex determination in P. fluviatilis***

398 Because *P. fluviatilis* sex determination does not rely on an *amhr2* duplication like what has been found
399 in *P. flavescens* [9] and *P. schrenkii* (this study), we used genome-wide approaches to better
400 characterize its sex-determination system. RAD-Seq analysis on 35 males and 35 females of *P.*
401 *fluviatilis*, carried out with a minimum read depth of one, allowed the characterization of a single
402 significant sex marker (Suppl. Fig. 4). This 94 bp RAD sequence matches (Blast Identities: 93/95%) a
403 portion of *P. fluviatilis* chromosome 18 (GENO_Pfluvi_1.0, Chr18: 27656212 - 27656305). This RAD-Seq
404 analysis suggests that Chr18 could be the *P. fluviatilis* sex chromosome, and that its sex-determining
405 region could be very small because we only detected a single significant sex-linked RAD-sequence. To
406 get a better characterization of the *P. fluviatilis* sex chromosome and sex-determining region, we then
407 used Pool-Sequencing (Pool-Seq) to re-analyze the same samples used for RAD-Seq by pooling together
408 DNA from the males in one pool and DNA from females in a second pool. Using these Pool-Seq
409 datasets, we identified a small 100 kb region on *P. fluviatilis* Chr18 with a high density of male-specific
410 SNVs (Fig. 4), confirming the RAD-Seq hypothesis that Chr18 is the sex chromosome in that species. No
411 male-specific duplication / insertion event was found in this sex-determining region on Chr18, which
412 contains six annotated genes (Fig. 4D). These genes encode a protein of unknown function
413 (*C18h1orf198*), three gap-junction proteins (*Cx32.2*, *Gja13.2* and *Cx32.7*), a protein annotated as
414 inactive hydroxysteroid dehydrogenase-like protein (*Hsd1*), and a protein known as protein broad-
415 minded or Tcb1 domain family member 32 protein (*Tbc1d32*). Three of these six annotated genes, i.e.,
416 *c18h1orf198*, *hsd1* and *tbc1d32* display a higher expression in testis than in the ovary (Suppl. Fig. 5).

417 To provide a better support for the sex-linkage of the male-specific variants found within this sex
418 determining region on Chr18, we designed different types of genotyping assays (classical PCR and

419 KASpar) that have been applied to different *P. fluviatilis* individuals which were phenotypically sexed
420 with confidence. A classical PCR-assay was first developed based on the detection of a Y-allele-specific
421 27 bp deletion in the third intron of the *P. fluviatilis hsd1* gene, (Suppl. Fig. 6, Table 4) and this assay
422 successfully identified all males of a Lake Mueggelsee from Germany (10 males and 9 females; p-value
423 of association with sex = 9.667e-05). In addition, KASpar allele-specific PCR-assays were developed
424 based on seven single nucleotide sex-specific variants, located at different positions within the Chr18
425 sex-determining region of *P. fluviatilis* (Fig. 3D). Tests of 48 males and 48 females showed that of seven
426 KASpar allele-specific PCR-assays, five resulted in a high proportion of correctly-genotyped individuals
427 with males being heterozygote and females being homozygote (>95%). Two of the targeted SNVs
428 (SNV1 and SNV3) that displayed 100% sex-linkage accuracy (Suppl. Fig. 7, Tables 4 and 5). Sex-linkage
429 of SNV1 was then checked on a wild-type population from Kortowskie Lake in Poland for which the
430 association of male phenotype and SNV1 heterozygosity was also complete (17 males and 20 females;
431 p-value = 8.83e-09, Table 4).

432

433 **DISCUSSION**

434 The Percidae family is represented by 239 species and 11 genera. The genera *Perca* and *Sander* are
435 especially important for aquaculture and fisheries. By providing new genome sequence assemblies for
436 *Perca fluviatilis*, *P. schrenckii* and *Sander vitreus*, we provide for the first-time access to all economically
437 important species of the *Percidae* at the DNA-level. Assembly statistics for these three new genome
438 sequence assemblies are reference grade with N50 continuity in the megabase range and
439 chromosomal length scaffolds, obtained either by Hi-C scaffolding or by conserved synteny analysis
440 involving the closest relatives. Importantly completeness on the gene-level is significantly higher than
441 in a short-read based draft genome for *P. fluviatilis*, published earlier [64], allowing much stronger
442 conclusions to be drawn regarding the presence or absence of possible sex-determining genes.

443 ***The origin of Percidae***

444 A phylogenomic approach using aligned non-coding sequences of 36 genomes resulted in a highly-
445 supported tree showing a rapid radiation of fish families. It has recently been shown that
446 phylogenomics based on non-coding sequences may be more reliable at resolving difficult-to-resolve
447 radiations in species trees (i.e. in Aves). According to our time calibration, many taxonomic orders
448 related to the Percidae emerged shortly after the Cretaceous-Paleogene (K-Pg) mass extinction event,
449 about 66 Mya, and gave rise to Percidae about 59 Mya. This is in contrast to many older studies, which
450 have, for example, dated the split of *Perca* ssp. and *Gasterosteus* ssp. back to the Cretaceous (73-165
451 Mya; 18 of 23 studies listed at www.timetree.org). Similar patterns in rapid radiations have been

452 observed in the avian tree of life and have likewise been attributed to the K-Pg mass extinction [61,62].
453 In context, it has been argued that the so-called “Lilliput effect”, which describes the selection in favor
454 of species with small body sizes and fast generation times after mass extinction events, can lead to an
455 increase in substitution rates and results in overestimations of node-ages for molecular clocks [65].

456 ***The evolution of sex determination in *Perca* and *Sander* species***

457 ***Evolution and turnover of Amhr2***

458 Sex determination systems with MSD evolved from duplications of the *amh* [10–17] or *amhr2* [9,18–
459 21] genes have now been characterized in many fish species, all with a male-heterogametic system
460 (XX/XY). In addition, the fact that *Amh* in monotremes [66], or *Amhr2* in some lizards [67] are Y-linked
461 also makes them strong MSD gene candidates in other vertebrate species. In Percidae, sex
462 determination has only been explored in some species of *Perca* [8,9], and *Amhr2* has been
463 characterized as a potential MSD gene in yellow perch, *P. flavescens* [9]. Our results suggest that this
464 duplication of *amhr2* in *P. flavescens* is also shared by *P. schrenkii* and *S. lucioperca*, implying an origin
465 of duplication in their last common ancestor, dated around 19-27 Mya. However, the fate of this
466 duplication seems to be complex - with multiple duplications/insertions on different chromosomes
467 with no clear sex-linkage in *S. lucioperca*, a secondary loss in *P. fluviatilis* and *S. vitreus*, contrasting
468 with a single potentially sex-linked duplication/insertion in *P. schrenkii* and in *P. flavescens*. This finding
469 suggests that the shared ancestral *amhr2b*-duplicated locus (Fig. 2A) might be a jumping locus that has
470 been moving around during its evolution as found for the *sdY* MSD jumping sex locus in salmonids
471 [68,69]. Additional evidence that these *amhr2b* genes originated a single ancestor also rely on the fact
472 that the *Amhr2b* proteins of *P. flavescens* [9], *P. schrenkii* and *S. lucioperca* share a similar gene
473 structure with an N-terminal truncation that results in the absence of the cysteine-rich extracellular
474 part of the receptor. This part of the receptor is known to be crucial for ligand binding [70]. A similar
475 N-terminal truncation of a duplicated *amhr2* was also described in catfishes from the Pangasidae
476 family, where this truncation had been hypothetically linked to a potentially new sex-determination
477 function that lacks ligand dependency [18]. In the genus *Sander*, the situation might be similar to that
478 in *Perca* regarding the changes of the sex-determination systems between species. In *S. lucioperca*,
479 *amhr2b* might still serve as the MSD-gene, but the several recent *amhr2b* duplications have
480 complicated our analysis so far. Similar to *P. fluviatilis*, *S. vitreus* has lost *amhr2b* and likely another
481 factor took over as a potential MSD-gene.

482 ***A new sex-specific locus in *P. fluviatilis****

483 The fact that *P. fluviatilis* sex does not rely on an *amhr2*-duplication like *P. flavescens* [9] and *P.*
484 *schrenkii* do, indicates that *P. fluviatilis* evolved a completely different and new MSD-gene. Our results

485 also show that this sex locus on *P. fluviatilis* Chr18 is very small compared to what is observed in many
486 fish species, with a potentially non-recombining size around 100 kb. This locus however is not the
487 smallest SD-locus described in teleost fish: in the pufferfish, *Takifugu rubripes*, the sex locus is limited
488 to a few SNPs that differentiate *amhr2* alleles on the X- and Y-chromosomes [19]. Because we did not
489 find any sign of a sex chromosome-specific duplication/insertion event in the *P. fluviatilis* SDR, this sex
490 locus seems to result from pure allelic diversification and is thus in contrast to *P. flavescens* [9] and
491 probably also *P. schrenkii* (this study). The *P. fluviatilis* sex specific-region on Chr18 contains six
492 annotated genes, which encode a protein of unknown function (*C18h1orf198*), three gap-junction
493 proteins (*Cx32.2*, *Gja13.2* and *Cx32.7*), a protein annotated as inactive hydroxysteroid dehydrogenase-
494 like protein (*Hsd1*) and a protein known as protein broad-minded or Tcb1 domain family member 32
495 protein (*Tbc1d32*). Of these genes, *hsd1* and *tbc1d32* are interesting potential MSD candidates in
496 *P. fluviatilis*, based on their potential functions and the fact that both display a higher testicular
497 expression compared to the ovary. The *Hsd1* protein is indeed annotated as “inactive” [71], but this
498 annotation only refers to its lack of enzymatic activity against substrates so far tested, leaving other
499 potential functional roles for a protein that is highly conserved in vertebrates [71]. In *Epinephelus*
500 *coioides*, *hsd1* has been shown to be differentially expressed during female-to-male sex-reversal, and
501 its expression profile clustered with *hsd17b1* [72], which plays a central role in converting sex steroids
502 and has recently been identified as a potential MSD gene in different species with a female
503 heterogametic (ZZ/ZW) sex determination system, like in oyster pompano, *Trachinotus anak* [73] and
504 different amberjack species [74,75]. The *Tbc1d32* protein has been shown to be required for a high
505 Sonic hedgehog (*Shh*) signaling in the mouse neural tube [76]. Given the role of *Shh* signaling
506 downstream of steroidogenic factor 1 (*nr5a1*) for the proper steroidogenic lineage fate [77] and the
507 importance of steroids in gonadal sex differentiation [78,79], *tbc1d32* would be also an interesting
508 candidate as a potential MSD gene in *Perca fluviatilis*.

509

510 CONCLUSIONS

511 Our study shows that *Percidae* exhibit a remarkable high variation in sex-determination mechanisms.
512 This variation is connected to deletion or amplification of *amhr2bY*, which if lost in certain species (like
513 *Perca fluviatilis* or *Sander vitreus*) should cause re-wiring of the sex determining pathways and result
514 in the rise of new SD-systems. The mechanisms behind a “jumping” *amhr2bY* expansion or loss and
515 which genes replace it as the MSD remain to be elucidated. The new *Percidae* reference sequence
516 assemblies presented here and the highly reliable sex markers developed for *P. fluviatilis* can now be
517 applied for sex genotyping in basic science as well in aquaculture.

518 **DATA AVAILABILITY**

519 All genome assemblies and raw sequence datasets have been submitted to NCBI/GENBANK under the
520 bioproject accessions: PRJNA549142, PRJNA637487, PRJNA808842 (*P. fluviatilis*, *P. schrenkii*, *S. vitreus*,
521 respectively).

522 **BENEFIT-SHARING STATEMENT**

523 A research collaboration was developed with scientists from the countries providing genetic samples
524 (PE and WL in USA, DZ in Poland and ST in Russia), all collaborators are included as co-authors, the
525 results of research have been shared with the provider communities, and the research addresses a
526 priority concern, in this case the conservation of organisms being studied. More broadly, our group is
527 committed to international scientific partnerships, as well as institutional capacity building.

528 **ACKNOWLEDGEMENTS**

529 We kindly acknowledge the NCBI/Genbank team for providing a GNOMON annotation of the *P.*
530 *fluviatilis* genome. This work was funded by the German Research Foundation (DFG) "Eigene Stelle"
531 grant KU 3596/1-1; project number: 324050651, by the Agence Nationale pour la Recherche (ANR) /
532 DFG PhyloSex project (ANR-13-ISV7-0005), the CRB-Anim "Centre de Ressources Biologiques pour les
533 animaux domestiques" project PERCH'SEX, the FEAMP "Fonds européen pour les affaires maritimes et
534 la pêche" project SEX'NPERCH, NIH (R35 GM139635), and NSF (2232891). The GeT-PlaGe and MGX
535 sequencing facilities were supported by France Génomique National infrastructure as part of the
536 "Investissement d'avenir" program managed by (ANR-10-INBS-09). We thank Dr. Tatjana Dujsebayeva,
537 Salamat Karlybai and his colleague for help during field work in Kazakhstan. We thank Eva Kreuz and
538 Wibke Kleiner for technical assistance. We thank Matt Faust (Ohio, DNR) for help with samples of *S.*
539 *vitreus*.

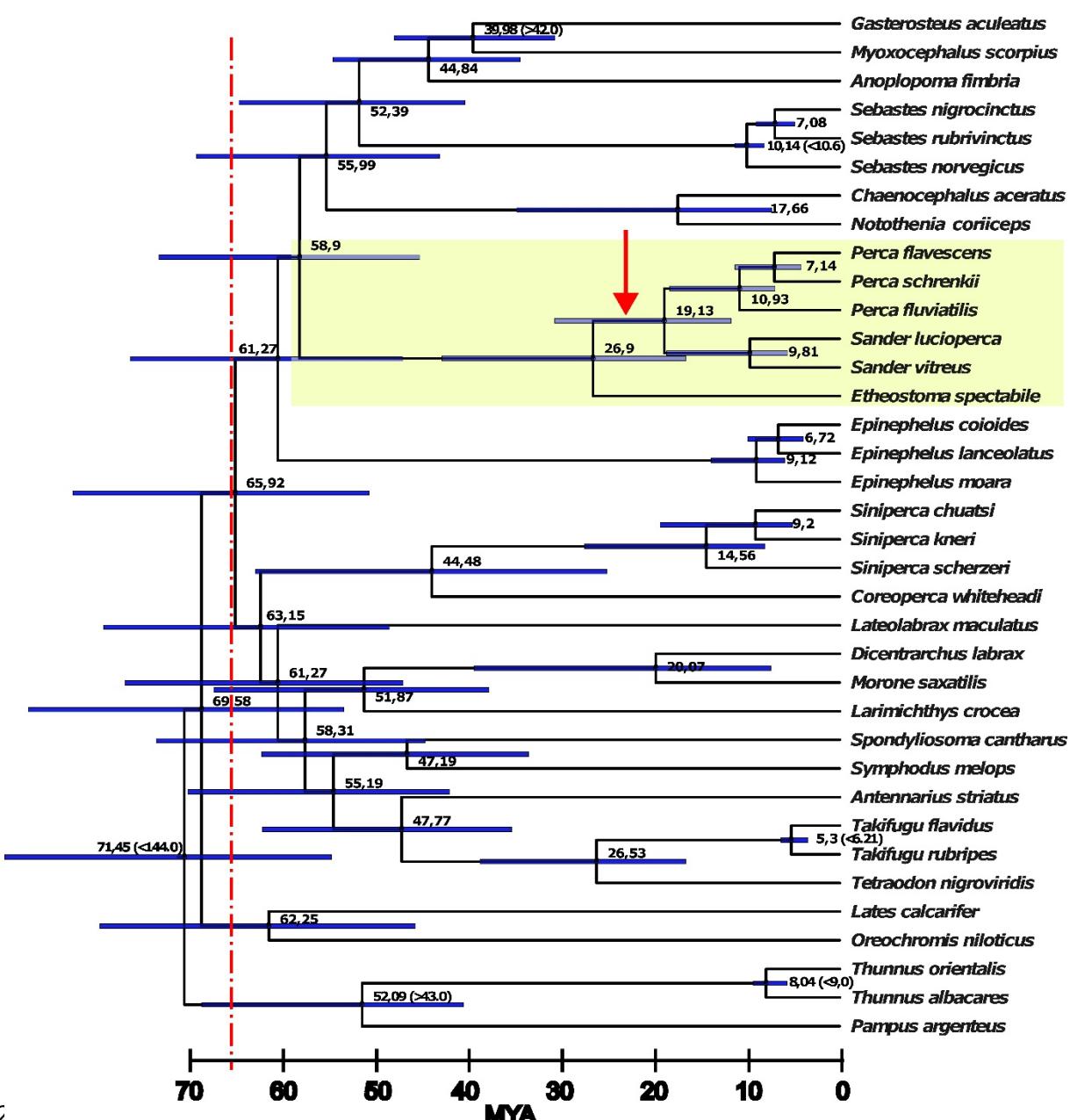
540 **AUTHOR CONTRIBUTIONS**

541 YG and HK designed the project. PE, WL provided sexed samples of *S. vitreus*. MW, BS, TL, ST, SV, DZ
542 collected and sexed the *P. fluviatilis* samples. MS in collaboration with ST did field work in Kazakhstan,
543 collected and sexed the tissue samples of *P. schrenkii*. EJ, MW, CR, CI, HP and HK extracted the gDNA,
544 made the genomic libraries and sequenced them. CeC, ChK, MZ, MW, CIK, RF, AH, HK and YG processed
545 the genome assemblies and/or analyzed the results. CP, LJ, CaC processed and analyzed the sex
546 genotyping tests. HK and YG wrote the manuscript with inputs from all other coauthors. JHP, CD, HK
547 and YG, supervised the project administration and raised funding. All the authors read and approved
548 the final manuscript.

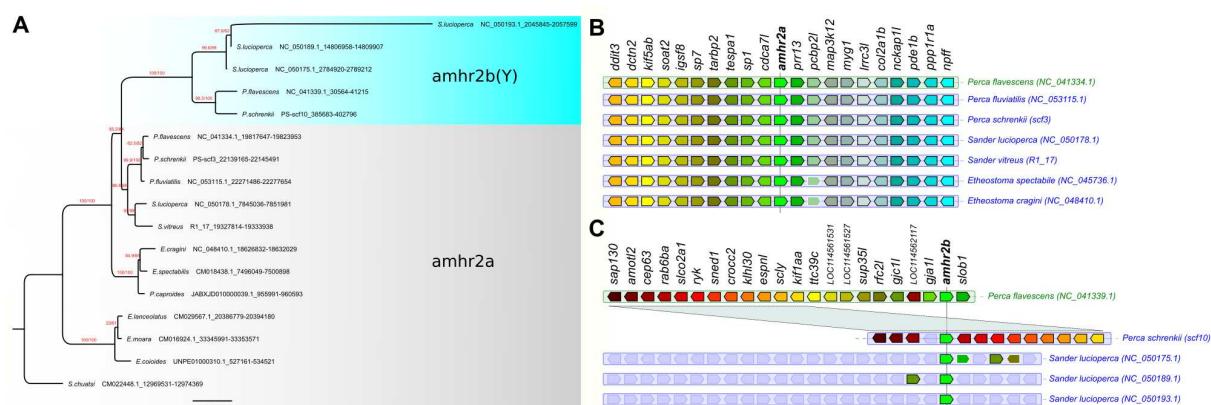
549 **COMPETING INTERESTS**

550 All authors declare no competing interests.

551 **FIGURES**

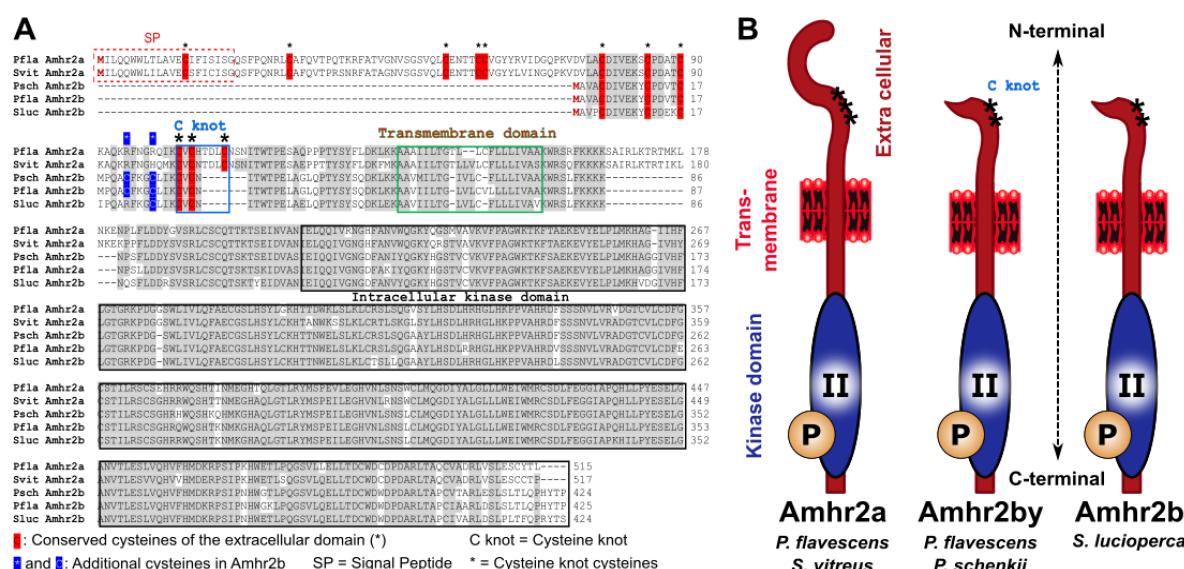


562



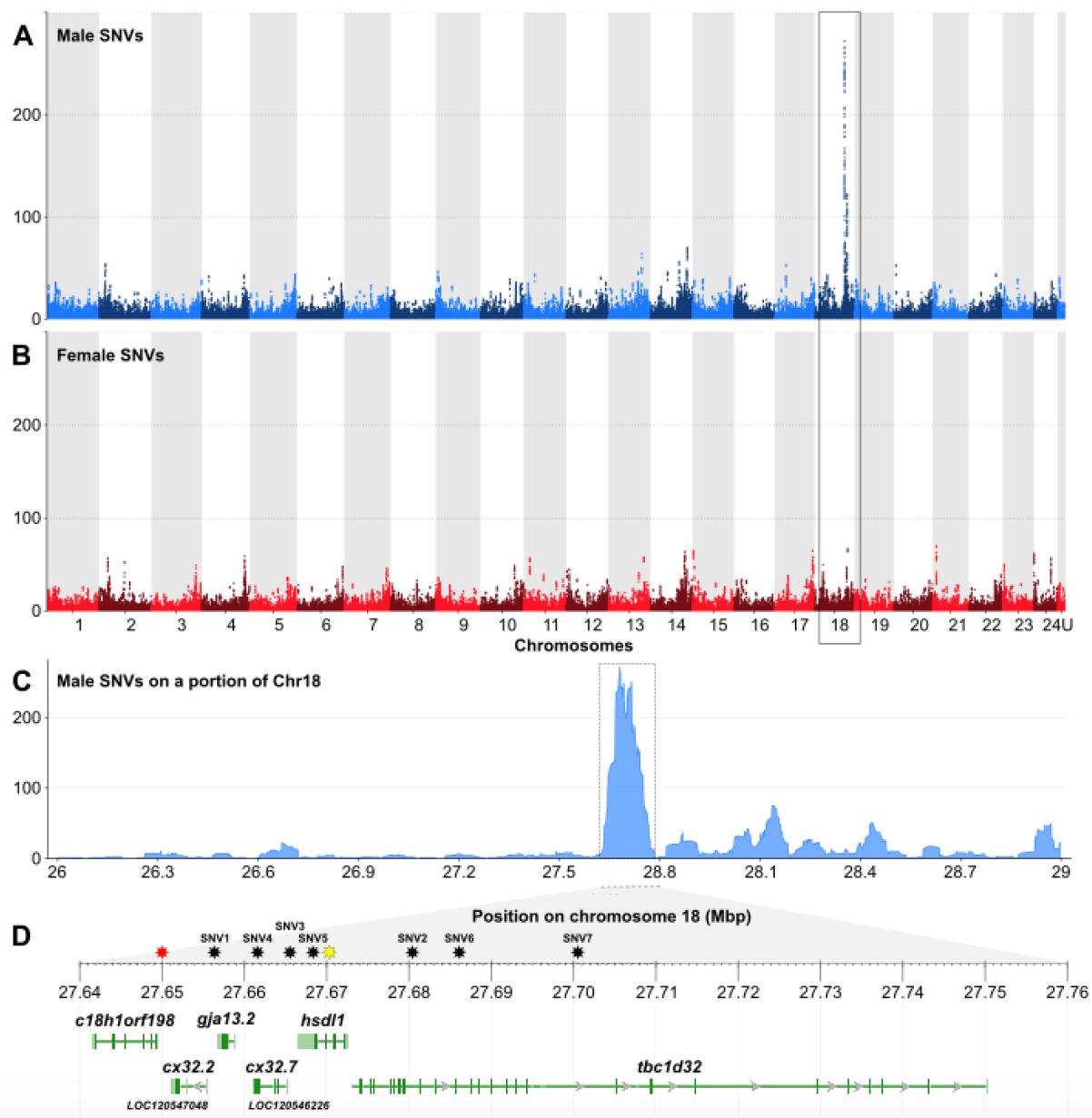
563

564 **Figure 2: Evolution of *amhr2* genes in Percidae.** **A.** Orthologs of *amhr2a* were identified in genome
565 assemblies and their gene tree is consistent with the species tree, which was found by the
566 phylogenomics approach (see Figure 1) The *amhr2b* duplications were only found in the genome
567 assemblies of *P. flavescens* (single copy, male specific), *P. schrenkii* (single copy, potentially male
568 specific) and *S. lucioperca* (three copies, no clear sex linkage) and they clustered together. Thus,
569 *amhr2b* stems from a gene duplication event that occurred at the origin of Percidae (19-27 Mya) and
570 the absence of *amhr2b* in *P. fluviatilis* and *S. vitreus* suggests a secondary loss event in these species.
571 This tree was calculated on codon position 1 and 2 alignments and achieved the best bootstrap support
572 (red numbers: SH-aLRT / UFBS support values) for the split of the *amhr2a* and *b* clades. Trees based on
573 complete CDS, CDS + Introns and amino acid sequences resulted in the same topology albeit with lower
574 bootstrap support for some splits (Suppl. Fig. 8). Numbers after species names depict the coordinates
575 of the respective *amhr2* genes in the genome assemblies. **B & C. Conserved synteny around the**
576 ***amhr2a* (B) and *amhr2b* (C) loci in some Percidae species.** These multiple duplications (in
577 *S. lucioperca*) or the loss of the *amhr2b* genes (in *P. fluviatilis* and *S. vitreus*) emphasize that *amhr2b*
578 may be considered a “jumping” gene locus, which is also supported by conserved synteny analysis.

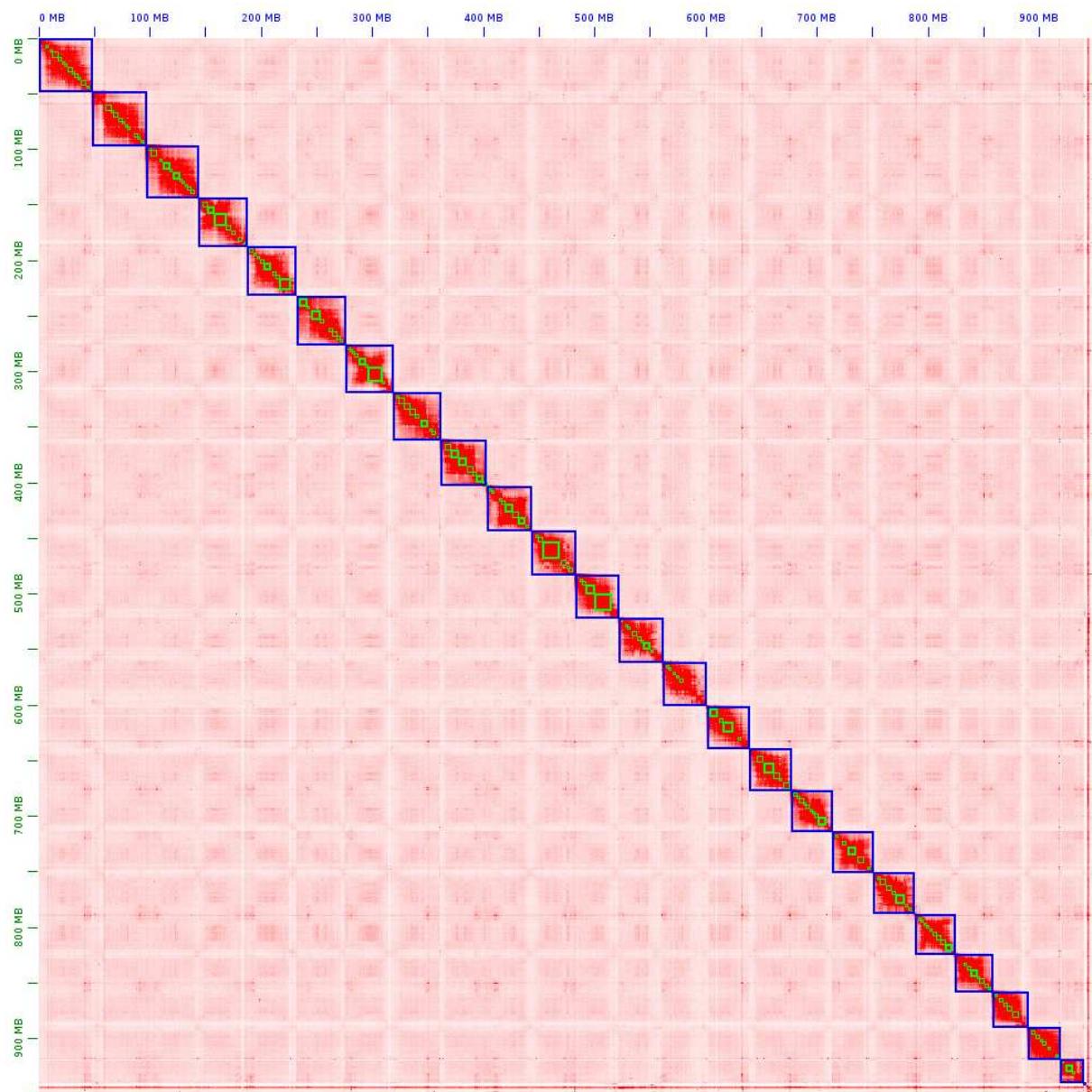


579

580 **Figure 3: Alignment and structure of Amhr2a and Amhr2b proteins in *Perca flavescens* (Pfla) and**
581 ***P. schrenkii* (Psch), *Sander lucioperca* (Sluc) and *S. vitreus* (Svit).** **A.** Alignment of some Percidae Amhr2a
582 (P. fluviatilis and S. vitreus) and Amhr2b (P. schrenkii, P. flavescens and S. lucioperca) proteins. **B.**
583 Schematic structure of Amhr2a and Amhr2b proteins.



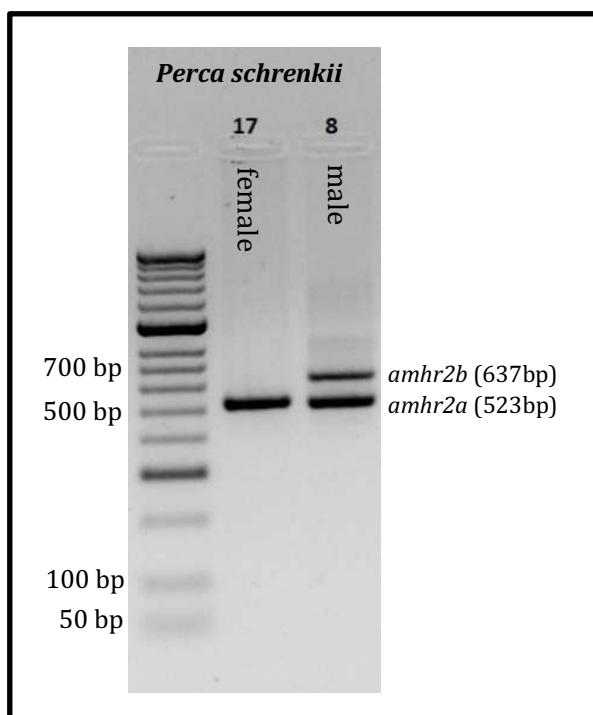
598 **Supplementary figures**



599

600 **Supplementary Figure 1:** Hi-C map of *P. fluviatilis* genome.

601

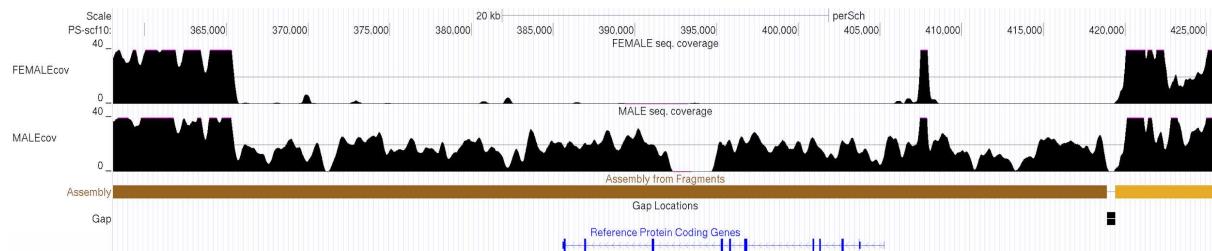


602

603 **Supplementary Figure 2:** A sex-specific 637 bp PCR product amplifies in *P. schrenkii* male (sample
604 id = 8), while it is absent in female (sample id = 17). The corresponding primer pair also works for sexing
605 of *P. flavescens*.

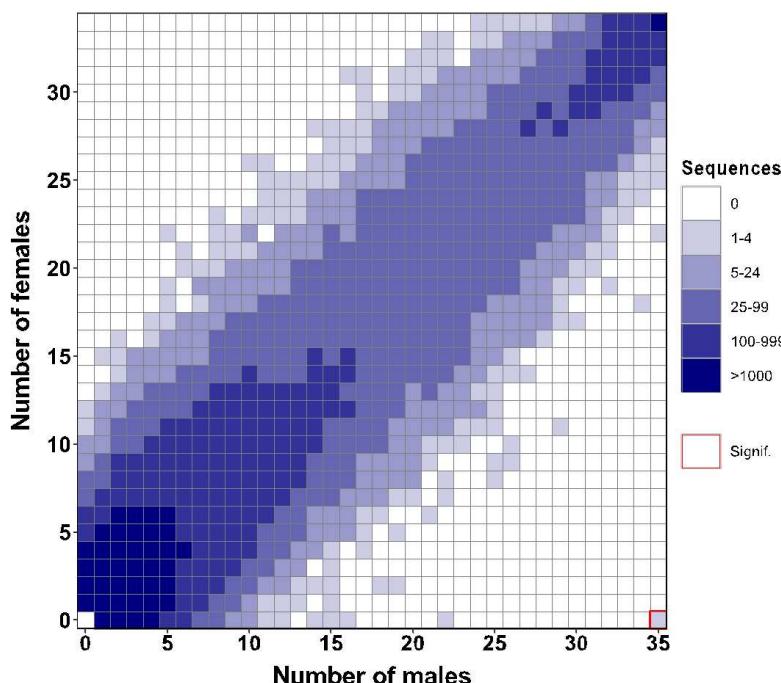
606

607



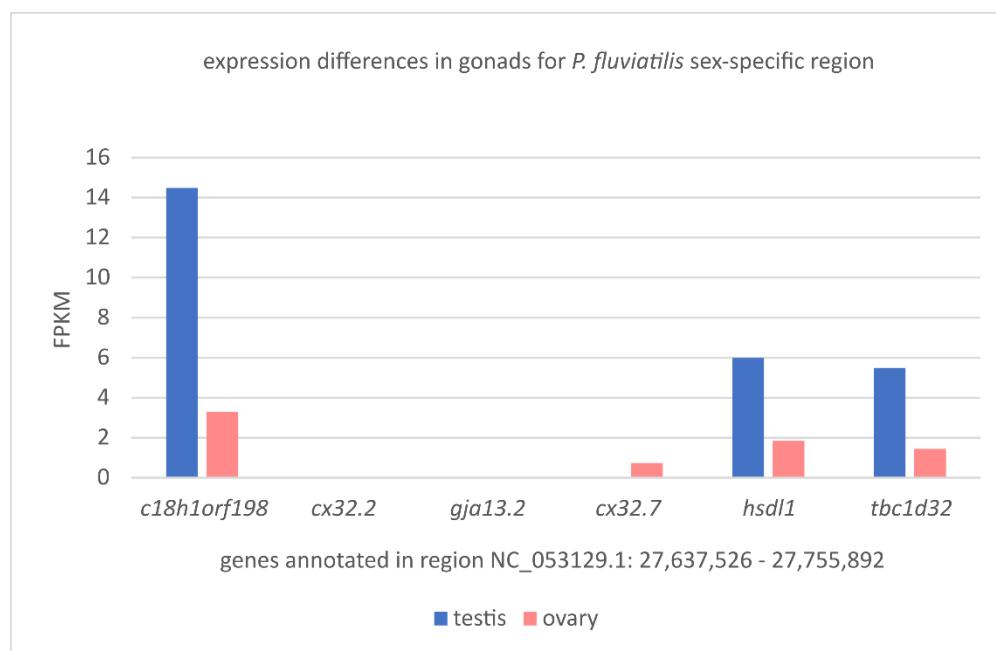
609 **Supplementary Figure 3:** Sex-specific sequencing coverage of *amhr2b* locus in *Perca schrenkii* in region
610 PS-scf10 / CM046795.1: 366,025-418,970. A female and a male genome were sequenced to
611 approximately 40x coverage using short-read sequencing. After filtering for unique mapping reads
612 (mapping quality 60), a clear coverage difference between females and males is visible. The ~53 kbp
613 region has virtually no coverage in females. In contrast males exhibit haploid coverage (about 20x),
614 which is in accordance with a X/Y SD system.

615



616

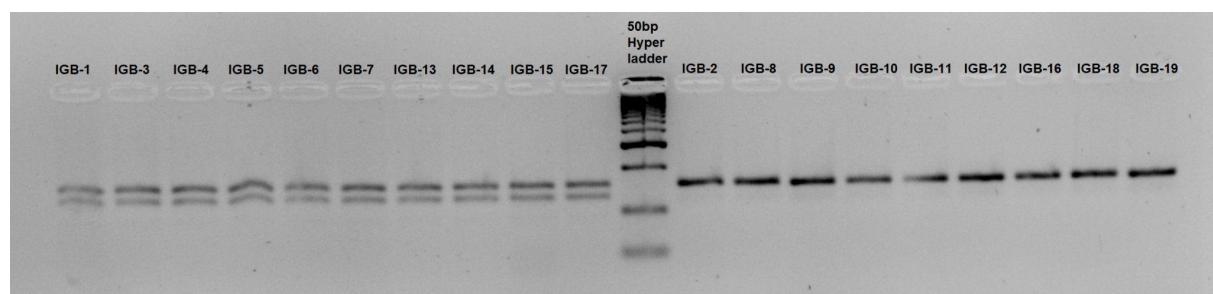
617 **Supplementary Figure 4: A single RAD sex-specific marker is significantly associated with male sex in**
618 ***P. fluviatilis*.** Tile plot of the distribution of RADSex markers between *Perca fluviatilis* males (horizontal
619 axis) and females (vertical axis) with a minimum read depth of 1 ($d = 1$). Color intensity (see color
620 legend on the right) indicates the number of markers present for each of the corresponding number
621 of males and females. A single significant marker at the lower right of the grid was present in all 35
622 males and absent from all 34 females and is boxed with a red border (Chi-squared test, $p < .05$ after
623 Bonferroni correction).



624

625 **Supplementary Figure 5: Expression of hsd1l and neighboring genes in public RNAseq datasets (testis:**
626 **SRR14461526, ovary: SRR14461527; age of both sampled individuals 9 month).** Here *hsd1l* expression
627 in testis is 3.25-fold higher than in ovary, for *tbc1d32* and *c18h1orf198* the ratio is 3.83 and 4.41,
628 respectively.

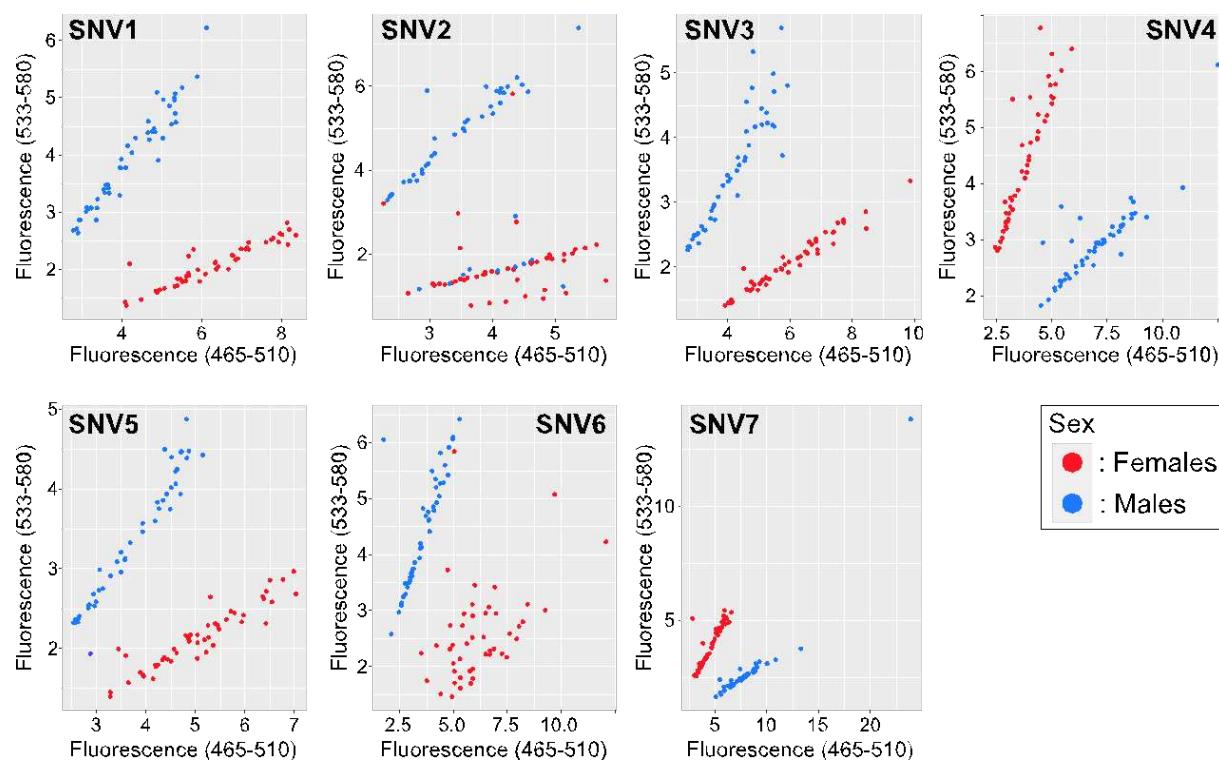
629



630

631 **Supplementary Figure 6:** Sexing of *P. fluviatilis* using a 27 bp male specific deletion in Intron 3 of the *hsd1* gene
632 (10 males (left) and 9 females (right) from Lake Mueggelsee, Berlin. The simultaneous amplification in both males
633 and females of the X allele without the 27 bp deletion provides an internal control for this PCR. All XY male
634 samples (N = 10) produce two amplicons due to the small size difference of the X and Y amplified alleles, and all
635 XX females (N = 9) produced only the larger X amplicon. This *hsd1* intronic indel variant is located near the variant
636 SNV5 (distance < 1.5 kbp).

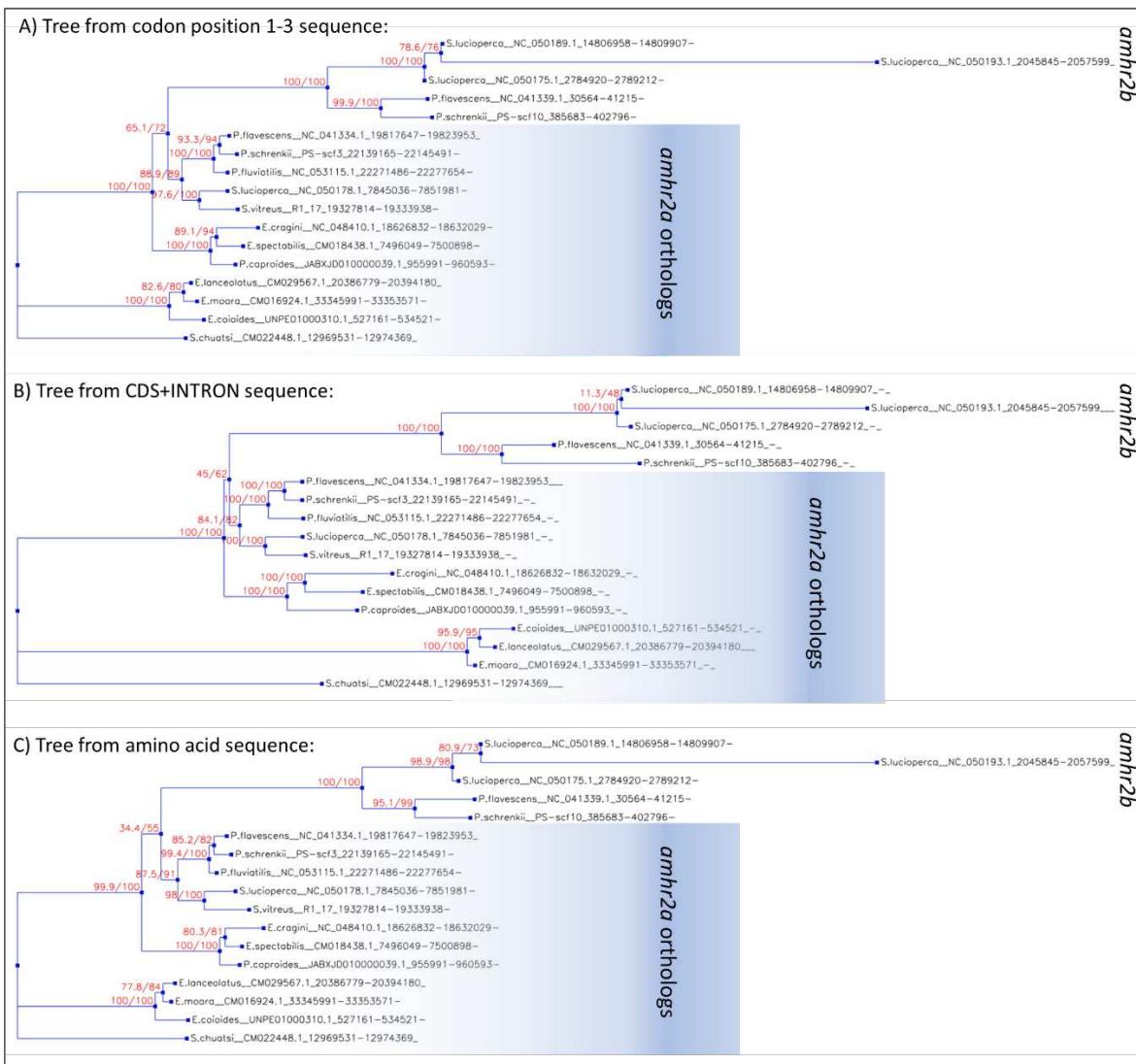
637



638

639 **Supplementary Figure 7: KASpar allele-specific PCR assays on seven single sex-specific nucleotide**
640 **variations (SNV ID#) in *P. fluviatilis*.** For each Single Nucleotide Variation (SNV), primer AL1 was
641 coupled to FAM fluorescent dye and primer AL2 was coupled to VIC fluorescent dye and the end-point
642 fluorescence of these two fluorescent dyes was respectively on the x- and y- axes. Male individuals are
643 represented by blue dots and females by red dots. Primers used for analysis can be found in Table 1.

644



645

646 **Supplementary Figure 8: Additional gene trees for amhr2. A) Tree calculated from coding sequence.**
 647 **B) Tree calculated from coding plus intron sequence. C) Tree calculated from amino acid sequence.**
 648 **All trees share the same topology but differ in support values for some splits (SH-aLRT and UFBS**
 649 **tests).**

650 **TABLES**

651 **Table 1: KASpar allele-specific PCR-primers.** For each allele (AL1 and AL2) primers, the X and Y sex
 652 chromosome-specific alleles are provided along with their sequences.

Primer ID#	Allele (X or Y)	Sequence (5' – 3')
SNV1_AL1	X	GAAGGTGACCAAGTTCATGCTGACACATATTGTCCATCTGATGTAAATG
SNV1_AL2	Y	GAAGGTCGGAGTCAACGGATTATGACACATATTGTCCATCTGATGTAAATT
SNV1_C	Common	CACCACCAGTACTGAAGAATAATATGAA
SNV2_AL1	X	GAAGGTGACCAAGTTCATGCTGGACTGATTGTGCTGCTTCTCTC
SNV2_AL2	Y	GAAGGTCGGAGTCAACGGATTGGACTGATTGTGCTGCTTCTCTT
SNV2_C	Common	CAGATGAGGAAGGAGGAGATGCAT
SNV3_AL1	X	GAAGGTGACCAAGTTCATGCTTCACCACCATAGAACCAACC
SNV3_AL2	Y	GAAGGTCGGAGTCAACGGATTGCTCACCACCATAGAACCAACT
SNV3_C	Common	GGGATGAGATGCCATTCTCCAATAATA
SNV4_AL1	Y	GAAGGTGACCAAGTTCATGCTCGCCCTCAGCCTGGTTGAT
SNV4_AL2	X	GAAGGTCGGAGTCAACGGATTGCCCTCAGCCTGGTTGAG
SNV4_C	Common	TCGTATGCACCCCTCACAGCTT
SNV5_AL1	X	GAAGGTGACCAAGTTCATGCTGGAATTGCCCTGAAATAATGAATGAATATG
SNV5_AL2	Y	GAAGGTCGGAGTCAACGGATTGCTGGAATTGCCCTGAAATAATGAATGAATATA
SNV5_C	Common	AGGACATTACAGATTGGTCAGACCATATA
SNV6_AL1	X	GAAGGTGACCAAGTTCATGCTTACCCCTCTCACCACCTGTT
SNV6_AL2	Y	GAAGGTCGGAGTCAACGGATTCTTACCCCTCTCACCACCTGTTG
SNV6_C	Common	CATAGTTCTACCCCTACTGTACCAAGATA
SNV7_AL1	Y	GAAGGTGACCAAGTTCATGCTCAGTGAACCTCCCTATGAGGCA
SNV7_AL2	X	GAAGGTCGGAGTCAACGGATTAGTGAACCTCCCTATGAGGCG
SNV7_C	Common	CTCGGTACAAGGTTGAAAGATGAAAGATA

653

654 **Table 2: Genome assembly statistics for *P. fluviatilis* and *P. schrenkii*.** For comparison, the table also
 655 provides numbers for two recently published *Perca* sp. assemblies [9,64]. Abbreviations: LR = long
 656 reads; HiC = chromosome conformation capture; CG = comparative genomics; 10X = linked-read
 657 sequencing; GLM = genetic linkage map.

Species	this study			earlier studies		
	<i>P. fluviatilis</i>	<i>P. schrenkii</i>	<i>S. vitreus</i>	<i>P. flavescens</i>	<i>P. fluviatilis</i>	<i>S. lucioperca</i>
Strategy	LR+HiC	LR+CG	LR+CG	LR+10X+HiC	10X	LR+GLM
total length	951,362,726	908,224,480	791,708,797	877,456,336	958,225,486	901,238,333
longest 24 sequences assembly fraction	99.0%	94.7%	96.5%	98.8%	41.4%	99.5%
scaffold/chromosome N50	39,550,354	36,400,992	33,333,317	37,412,490	6,260,519	41,060,379
contig N50	4,101,751	3,162,456	6,206,245	4,268,950	12,991	6,160,542

658 **Table 3: BUSCO scoring of annotations of five Percidae genome assemblies (*P. fluviatilis*, *P. schrenkii*,
659 *S. vitreus* from this study). The comparative mapping of high quality NCBI/GNOMON annotations onto
660 closely related species' genome assemblies is a cost-effective and fast procedure to annotate new
661 genomes. Abbreviations used: C = complete; S = single copy; D = duplicated; F = fragmented; M =
662 missing.**

	Species	Annotation type	BUSCO scoring code
# 1	<i>P. fluviatilis</i>	NCBI/GNOMON	C: 98.5% [S: 95.3%, D: 3.2%], F: 1.0%, M: 0.5%, n: 4584
# 2	<i>P. flavescens</i>	NCBI/GNOMON	C: 99.3% [S: 95.9%, D: 3.4%], F: 0.5%, M: 0.2%, n: 4584
# 3	<i>P. schrenkii</i>	mapping of annot. #2	C: 95.9% [S: 92.6%, D: 3.3%], F: 2.8%, M: 1.3%, n: 4584
# 4	<i>S. lucioperca</i>	NCBI/GNOMON	C: 99.3% [S: 95.9%, D: 3.4%], F: 0.5%, M: 0.2%, n: 4584
# 5	<i>S. vitreus</i>	mapping of annot. #4	C: 98.6% [S: 95.5%, D: 3.1%], F: 0.8%, M: 0.6%, n: 4584

663

664 **Table 4: Sex-linkage of different sex-markers in *Perca fluviatilis* and *Perca schrenkii*.** Associations
665 between each sex-specific marker and sex phenotypes are provided for both males and females
666 (number of positive individuals / total number of individuals) along with the p-value of association with
667 sex that was calculated for each species and population based on the Pearson's Chi-square test with
668 Yates' continuity correction. ns = not statistically significant.

Species	Population	marker	males	females	p value
<i>Perca schrenkii</i>	Alakol Lake, Kazakhstan	<i>amhr2bY</i>	1/1	0/1	ns
<i>Perca fluviatilis</i>	Mueggelsee Lake Germany	<i>hsd1</i>	10/10	0/9	9.667e-05
<i>Perca fluviatilis</i>	Lucas Perche, France	SNV1	48/48	0/47	< 2.2e-16
<i>Perca fluviatilis</i>	Kortowskie Lake, Poland	SNV1	17/17	0/20	8.83e-09

669 **Table 5: KASpar allele-specific PCR assays on seven single sex-specific nucleotide variations (SNV ID#)**
670 **in *P. fluviatilis*.** Numbers of homozygote (Ho), heterozygote (He) and uncalled genotypes (U). N = Total
671 numbers of genotyped individuals (M/F: Males/females), %N = percentage of genotyped individuals,
672 % As = percentage of correctly assigned genotypes i.e., male heterozygotes and female homozygotes.
673 The p-value of association with sex was calculated for each SNV based on the Pearson's Chi-square test
674 with Yates' continuity correction scoring heterozygote males and homozygote females as positives.

SNV ID#	N (M/F)	% N	M Ho/He/U	F Ho/He/U	% As	p value
SNV1	95 (48/47)	99,0	0/48/0	47/0/0	100,0	< 2.2e-16
SNV2	96 (48/48)	100,0	8/37/3	34/2/12	74,0	3.175e-11
SNV3	93 (45/48)	96,9	0/45/0	48/0/0	100,0	< 2.2e-16
SNV4	96 (48/48)	100,0	0/46/2	46/0/2	95,8	< 2.2e-16
SNV5	92 (44/48)	95,8	0/42/2	46/0/2	95,7	< 2.2e-16
SNV6	93 (48/45)	96,9	1/47/0	25/2/18	77,4	1.964e-14
SNV7	95 (47/48)	99,0	1/46/0	45/3/0	95,8	< 2.2e-16

675

676 **Supplementary Table 1:** Annotated repeats in *Perca* sp. and *Sander* sp. genomes (RepeatModeler *de*
677 *novo* analysis). Repeat elements mentioned in the manuscript have grey highlighting. The class “DNA”
678 is assigned to repeat elements that harbor signals of transposases, but miss further signals to classify
679 them with more detail.

	<i>P. fluviatilis</i>	<i>P. flavescens</i>	<i>P. schrenkii</i>	<i>S. lucioperca</i>	<i>S. vitreus</i>
length (no gaps)	951.053.269	877.041.836	893.440.234	901.028.039	785.350.625
Repeat class	bp masked	bp masked	bp masked	bp masked	bp masked
DNA	22.711.613	20.520.179	23.132.743	19.957.261	15.428.015
Academ	2.340.071	1.619.886	1.842.836	2.412.073	1.790.843
CMC-Chapaev-3	230.562	0	0	0	0
CMC-EnSpm	9.118.071	7.588.794	4.039.674	5.928.087	3.520.268
Crypton	841.031	547.307	748.707	772.498	402.911
Crypton-V	11.212	0	0	0	0
Dada	0	33.030	206.119	0	0
Ginger	0	15.288	0	0	0
IS3EU	1.282.033	531.773	496.067	288.665	509.614
Kolobok-Hydra	0	648.047	0	0	70.720
Kolobok-T2	1.252.506	1.976.511	2.668.034	1.586.132	1.222.046
MULE-MuDR	251.959	1.716.319	441.114	266.625	27.741
Maverick	2.187.699	1.402.756	266.487	1.342.717	0
Merlin	0	0	0	122.531	94.494
Novosib	0	0	0	0	82.199
P	467.516	442.149	411.688	765.209	555.234
PIF-Harbinger	12.989.790	12.205.497	12.181.221	12.830.055	8.037.073
PIF-ISL2EU	0	235.589	464.262	1.597.340	914.106
PiggyBac	3.844.887	2.011.288	2.742.549	3.776.961	2.299.194
Sola	1.885.532	229.560	163.435	77.871	0
TcMar	17.070	135.873	0	185.840	137.413
TcMar-Fot1	0	30.498	0	44.754	37.051
TcMar-ISRm11	4.570.634	3.171.866	2.722.899	2.158.133	1.830.923
TcMar-Stowaway	220.092	98.992	114.404	65.697	0
TcMar-Tc1	15.259.371	12.137.252	12.302.578	6.557.230	10.225.920
TcMar-Tc2	627.258	465.068	299.455	227.802	468.716
Zator	0	119.597	0	0	0
Zisupton	30.281	0	127.852	0	104.202

Zisupton-hAT-hybrid	0	0	0	0	339.346
hAT	9.251.059	7.244.672	9.113.922	8.460.295	6.356.798
hAT-Ac	26.161.464	27.815.352	29.153.525	30.578.355	27.361.032
hAT-Blackjack	1.252.051	459.611	338.327	514.383	437.930
hAT-Charlie	7.578.170	5.645.728	8.825.006	5.557.301	3.893.733
hAT-Tip100	3.892.297	2.083.169	5.060.148	2.183.415	2.136.454
hAT-Tol2	7.383.933	7.130.470	10.507.431	6.788.875	4.948.115
hAT-hAT5	1.008.628	1.141.196	1.066.203	873.439	560.575
hAT-hAT6	153.938	129.767	0	0	37.338
hAT-hobo	693.242	589.992	435.263	245.056	840.680
LINE	253.345	0	0	135.963	162.701
CR1	61.557	52.996	46.593	68.770	58.461
Dong-R4	0	161.938	0	0	0
I	902.480	1.009.950	1.721.862	451.473	448.921
I-Nimb	187.474	66.650	179.364	0	89.737
Jockey	1.281.097	0	0	249.376	0
L1	3.161.932	2.720.443	3.146.061	3.530.190	3.456.430
L1-Tx1	1.332.071	707.047	492.758	1.328.940	247.067
L2	24.295.832	18.778.236	18.931.729	15.671.956	10.981.470
Penelope	696.540	651.596	168.974	978.870	436.687
Proto2	432.031	406.555	335.607	288.595	210.869
R1	0	0	0	146.213	277.081
R2-Hero	258.797	37.021	152.400	101.778	204.860
R2-NeSL	139.492	0	191.055	106.574	62.871
RTE	1.012.896	0	0	0	0
RTE-BovB	3.915.100	11.058.368	4.053.280	5.607.772	8.522.404
RTE-RTE	0	0	0	169.573	0
RTE-RTEX	0	26.699	19.406	32.367	0
RTE-X	738.450	801.862	1.308.994	1.325.349	418.476
Rex-Babar	8.447.394	7.183.710	11.161.891	7.936.953	5.965.113
Tad1	0	0	0	138.056	320.275
LTR	569.316	146.042	61.769	0	0
Copia	180.685	275.175	422.246	471.447	0

DIRS	766.684	743.991	1.555.092	1.105.144	825.655
ERV1	750.875	1.123.315	980.225	1.469.763	126.774
ERVK	659.689	486.486	0	513.490	3.133.108
Gypsy	2.510.567	2.056.070	3.675.723	2.213.161	1.672.149
Ngaro	980.001	3.390.987	850.690	768.211	292.619
Pao	419.975	385.357	723.976	1.151.953	307.558
RC	0	0	0	0	0
Helitron	6.307.248	3.514.054	4.649.952	5.159.806	2.433.280
Retroposon	213.419	116.255	421.490	238.385	0
SINE	1.210.545	1.637.348	1.401.793	1.524.037	936.353
5S-Deu-L2	721.151	965.222	838.622	0	873.243
ID	0	0	19.012	29.728	116.970
MIR	1.185.268	1.020.008	1.163.355	2.103.257	382.739
tRNA	3.640.638	3.701.722	3.698.003	3.271.022	3.693.238
tRNA-Core	0	0	0	0	293.857
tRNA-Core-L2	0	0	0	89.805	106.872
tRNA-Deu-L2	0	0	0	42.616	0
tRNA-L2	461.844	0	103.308	223.271	301.117
tRNA-V	0	0	492.471	0	0
Unknown	149.784.304	121.893.498	134.020.433	144.621.637	113.225.894
total interspersed	354.992.667	305.241.677	326.860.083	319.430.101	255.255.533
Low_complexity	2.964.797	2.285.494	2.321.578	2.478.451	2.097.985
Satellite	700.110	1.425.777	1.023.206	656.699	1.718.620
Simple_repeat	30.391.090	32.172.324	29.527.186	34.124.020	27.469.062
rRNA	358.199	351.617	313.179	42.963	708.978
snRNA	0	0	0	9.265	0
Total	389.406.863	341.476.889	360.045.232	356.741.499	287.250.178

680

681

682 **REFERENCES**

683 1. Stepien CA, Haponski AE. Taxonomy, Distribution, and Evolution of the Percidae. In: Kestemont
684 P, Dabrowski K, Summerfelt RC, editors. *Biology and Culture of Percid Fishes: Principles and*
685 *Practices*. Dordrecht: Springer Netherlands;

686 2. Hermelink B, Wuertz S, Trubiroha A, Rennert B, Kloas W, Schulz C. Influence of temperature on
687 puberty and maturation of pikeperch, *Sander lucioperca*. *Gen Comp Endocrinol*. 2011; doi:
688 10.1016/j.ygcn.2011.03.013.

689 3. Schaefer FJ, Overton JL, Kloas W, Wuertz S. Length rather than year-round spawning, affects
690 reproductive performance of RAS-reared F-generation pikeperch, *Sander lucioperca* (Linnaeus, 1758)
691 - Insights from practice. *J Appl Ichthyol*. 2018; doi: 10.1111/jai.13628.

692 4. Policar T, Schaefer FJ, Panama E, Meyer S, Teerlinck S, Toner D, et al.. Recent progress in
693 European percid fish culture production technology-tackling bottlenecks (vol 27, pg 4, 2019).
694 *Aquacult Int*. 2019; doi: 10.1007/s10499-019-00457-4.

695 5. Malison JA, Kayes TB, Best CD, Amundson CH, Wentworth BC. Sexual Differentiation and Use of
696 Hormones to Control Sex in Yellow Perch (*Perca flavescens*). *Can J Fish Aquat Sci*. 1986; doi:
697 10.1139/f86-004.

698 6. Melard C, Kestemont P, Grignard JC. Intensive culture of juvenile and adult Eurasian perch (P-
699 fluviatilis): Effect of major biotic and abiotic factors on growth. *J Appl Ichthyol*. 1996; doi: DOI
700 10.1111/j.1439-0426.1996.tb00085.x.

701 7. Rougeot C, Jacobs B, Kestemont P, Melard C. Sex control and sex determinism study in Eurasian
702 perch, *Perca fluviatilis*, by use of hormonally sex-reversed male breeders. *Aquaculture*. 2002; doi: Pii
703 S0044-8486(01)00893-6 Doi 10.1016/S0044-8486(01)00893-6.

704 8. Rougeot C, Ngingo JV, Gillet L, Vanderplasschen A, Melard C. Gynogenesis induction and sex
705 determination in the Eurasian perch, *Perca fluviatilis*. *Aquaculture*. 2005; doi:
706 10.1016/j.aquaculture.2004.11.004.

707 9. Feron R, Zahm M, Cabau C, Klopp C, Roques C, Bouchez O, et al.. Characterization of a Y-
708 specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a
709 chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Mol Ecol Resour*. 2020; doi:
710 10.1111/1755-0998.13133.

711 10. Pan Q, Kay T, Depincé A, Adolfi M, Schartl M, Guiguen Y, et al.. Evolution of master sex
712 determiners: TGF- β signalling pathways at regulatory crossroads. *Philos Trans R Soc Lond B Biol Sci*.
713 2021; doi: 10.1098/rstb.2020.0091.

714 11. Pan Q, Feron R, Jouanno E, Darras H, Herpin A, Koop B, et al.. The rise and fall of the ancient
715 northern pike master sex-determining gene. Przeworski M, Weigel D, editors. *eLife*. eLife Sciences
716 Publications, Ltd; 2021; doi: 10.7554/eLife.62858.

717 12. Pan Q, Feron R, Yano A, Guyomard R, Jouanno E, Vigouroux E, et al.. Identification of the
718 master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome
719 differentiation. *PLOS Genetics*. 2019; doi: 10.1371/journal.pgen.1008013.

720 13. Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, et al.. A Tandem Duplicate of Anti-Müllerian

721 Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile
722 Tilapia, *Oreochromis niloticus*. *PLOS Genetics*. 2015; doi: 10.1371/journal.pgen.1005678.

723 14. Holborn MK, Einfeldt AL, Kess T, Duffy SJ, Messmer AM, Langille BL, et al.. Reference genome
724 of lumpfish *Cyclopterus lumpus* Linnaeus provides evidence of male heterogametic sex determination
725 through the AMH pathway. *Mol Ecol Resour*. 2022; doi: 10.1111/1755-0998.13565.

726 15. Song W, Xie Y, Sun M, Li X, Fitzpatrick CK, Vaux F, et al.. A duplicated amh is the master sex-
727 determining gene for *Sebastes* rockfish in the Northwest Pacific. *Open Biol*. 2021; doi:
728 10.1098/rsob.210063.

729 16. Rondeau EB, Laurie CV, Johnson SC, Koop BF. A PCR assay detects a male-specific duplicated
730 copy of Anti-Müllerian hormone (amh) in the lingcod (*Ophiodon elongatus*). *BMC Res Notes*. 2016;
731 doi: 10.1186/s13104-016-2030-6.

732 17. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, et al.. A Y-linked anti-Müllerian
733 hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci USA*. 2012;
734 doi: 10.1073/pnas.1018392109.

735 18. Wen M, Pan Q, Jouanno E, Montfort J, Zahm M, Cabau C, et al.. An ancient truncated duplication
736 of the anti-Müllerian hormone receptor type 2 gene is a potential conserved master sex determinant in
737 the Pangasiidae catfish family. *Mol Ecol Resour*. 2022; doi: 10.1111/1755-0998.13620.

738 19. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, et al.. A trans-species missense
739 SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu).
740 *PLoS Genet*. 2012; doi: 10.1371/journal.pgen.1002798.

741 20. Nakamoto M, Uchino T, Koshimizu E, Kuchiishi Y, Sekiguchi R, Wang L, et al.. A Y-linked anti-
742 Müllerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLoS
743 Genet*. 2021; doi: 10.1371/journal.pgen.1009705.

744 21. Qu M, Liu Y, Zhang Y, Wan S, Ravi V, Qin G, et al.. Seadragon genome analysis provides
745 insights into its phenotype and sex determination locus. *Sci Adv*. 2021; doi: 10.1126/sciadv.abg5196.

746 22. Nguinkal JA, Brunner RM, Verleih M, Rebl A, de los Rios-Perez L, Schafer N, et al.. The First
747 Highly Contiguous Genome Assembly of Pikeperch (*Sander lucioperca*), an Emerging Aquaculture
748 Species in Europe. *Genes-Basel*. 2019; doi: ARTN 708 10.3390/genes10090708.

749 23. Malmstrom M, Matschiner M, Torresen OK, Jakobsen KS, Jentoft S. Whole genome sequencing
750 data and de novo draft assemblies for 66 teleost species. *Sci Data*. 2017; doi: ARTN 160132
751 10.1038/sdata.2016.132.

752 24. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al.. Multi-species annotation of
753 transcriptome and chromatin structure in domesticated animals. *BMC Biology*. 2019; doi:
754 10.1186/s12915-019-0726-5.

755 25. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex
756 MinION sequencing. *Microb Genom*. 2017; doi: 10.1099/mgen.0.000132.

757 26. Hailin Liu SW. SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte*. 2021; doi:
758 10.46471/gigabyte.15.

759 27. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018; doi:

760 10.1093/bioinformatics/bty191.

761 28. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long
762 uncorrected reads. *Genome Res.* 2017; doi: 10.1101/gr.214270.116.

763 29. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
764 *arXiv:13033997 [q-bio]*. 2013;

765 30. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al.. Pilon: an integrated
766 tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*.
767 2014; doi: 10.1371/journal.pone.0112963.

768 31. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al.. Juicer Provides a
769 One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* 2016; doi:
770 10.1016/j.cels.2016.07.002.

771 32. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al.. De novo assembly
772 of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017; doi:
773 10.1126/science.aal3327.

774 33. Robinson JT, Turner D, Durand NC, Thorvaldsdottir H, Mesirov JP, Aiden EL. Juicebox.js
775 Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst.* 2018; doi:
776 10.1016/j.cels.2018.01.001.

777 34. Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, et al.. LR_Gapcloser: a tiling path-based gap
778 closer that uses long reads to complete genome assembly. *Gigascience*. 2019; doi: ARTN giy157
779 10.1093/gigascience/giy157.

780 35. Garrison E MG. Haplotype-based variant detection from short-read sequencing. *arXiv preprint*
781 *arXiv:12073907 [q-bioGN]*. 2012;

782 36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
783 *Bioinformatics*. 2014; doi: 10.1093/bioinformatics/btu170.

784 37. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and
785 metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012; doi:
786 10.1093/bioinformatics/bts174.

787 38. Kuhl H, Li L, Wuertz S, Stock M, Liang XF, Klopp C. CSA: A high-throughput chromosome-
788 scale assembly pipeline for vertebrate genomes. *Gigascience*. 2020; doi: 10.1093/gigascience/giaa034.

789 39. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat
790 graphs. *Nat Biotechnol.* 2019; doi: 10.1038/s41587-019-0072-8.

791 40. Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended
792 version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* 2012; doi:
793 ARTN e161 10.1093/nar/gks708.

794 41. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing
795 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; doi:
796 10.1093/bioinformatics/btv351.

797 42. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief*

798 798 *Bioinform.* 2013; doi: 10.1093/bib/bbs038.

799 799 43. Kent WJ. BLAT - The BLAST-like alignment tool. *Genome Res.* 2002; doi: 10.1101/gr.229202.

800 800 44. He S, Li L, Lv LY, Cai WJ, Dou YQ, Li J, et al.. Mandarin fish (Siniperidae) genomes provide
801 insights into innate predatory feeding. *Commun Biol.* 2020; doi: ARTN 361 10.1038/s42003-020-
802 1094-y.

803 803 45. Frith MC, Kawaguchi R. Split-alignment of genomes finds orthologies more accurately. *Genome
804 Biol.* 2015; doi: ARTN 106 10.1186/s13059-015-0670-9.

805 805 46. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, et al.. Aligning multiple
806 genomic sequences with the threaded blockset aligner. *Genome Res.* 2004; doi: 10.1101/gr.1933104.

807 807 47. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-
808 friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 2019; doi:
809 10.1093/bioinformatics/btz305.

810 810 48. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al.. IQ-
811 TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol
812 Evol.* 2020; doi: 10.1093/molbev/msaa015.

813 813 49. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007; doi:
814 10.1093/molbev/msm088.

815 815 50. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic
816 maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome
817 duplication. *Genetics.* 2011; doi: 10.1534/genetics.111.127324.

818 818 51. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and
819 genotyping Loci de novo from short-read sequences. *G3 (Bethesda).* 2011; doi:
820 10.1534/g3.111.000240.

821 821 52. Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, et al.. RADSex: A computational
822 workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol Ecol
823 Resour.* 2021; doi: 10.1111/1755-0998.13360.

824 824 53. Imarazene B, Du K, Beille S, Jouanno E, Feron R, Pan Q, et al.. A supernumerary “B-sex”
825 chromosome drives male sex determination in the Pachón cavefish, *Astyanax mexicanus*. *Curr Biol.*
826 2021; doi: 10.1016/j.cub.2021.08.030.

827 827 54. Wen M, Pan Q, Larson W, Eché C, Guiguen Y. Characterization of the sex determining region of
828 channel catfish (*Ictalurus punctatus*) and development of a sex-genotyping test. *Gene.* 2023; doi:
829 10.1016/j.gene.2022.146933.

830 830 55. Jasonowicz AJ, Simeon A, Zahm M, Cabau C, Klopp C, Roques C, et al.. Generation of a
831 chromosome-level genome assembly for Pacific halibut (*Hippoglossus stenolepis*) and
832 characterization of its sex-determining genomic region. *Mol Ecol Resour.* 2022; doi: 10.1111/1755-
833 0998.13641.

834 834 56. Smith SM, Maughan PJ. SNP genotyping using KASPar assays. *Methods Mol Biol.* 2015; doi:
835 10.1007/978-1-4939-1966-6_18.

836 57. Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, et al.. Gene evolution and gene
837 expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics*. 2016; doi:
838 10.1186/s12864-016-2709-z.

839 58. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements.
840 *Nat Methods*. 2015; doi: 10.1038/nmeth.3317.

841 59. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables
842 improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015; doi:
843 10.1038/nbt.3122.

844 60. Braun EL, Kimball RT. Data Types and the Phylogeny of Neoaves. *Birds*. 2:1–222021;

845 61. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al.. Whole-genome analyses resolve
846 early branches in the tree of life of modern birds. *Science*. 2014; doi: 10.1126/science.1253451.

847 62. Kuhl H, Frankl-Vilches C, Bakker A, Mayr G, Nikolaus G, Boerno ST, et al.. An Unbiased
848 Molecular Approach Using 3 'UTRs Resolves the Avian Family-Level Tree of Life. *Mol Biol Evol*.
849 2021; doi: 10.1093/molbev/msaa191.

850 63. Heldin C-H, Moustakas A. Signaling Receptors for TGF- β Family Members. *Cold Spring Harb
851 Perspect Biol*. 2016; doi: 10.1101/cshperspect.a022053.

852 64. Ozerov MY, Ahmad F, Gross R, Pukk L, Kahar S, Kisand V, et al.. Highly Continuous Genome
853 Assembly of Eurasian Perch (*Perca fluviatilis*) Using Linked-Read Sequencing. *G3: Genes, Genomes,
854 Genetics*. 2018; doi: 10.1534/g3.118.200768.

855 65. Berv JS, Field DJ. Genomic Signature of an Avian Lilliput Effect across the K-Pg Extinction. *Syst
856 Biol*. 2018; doi: 10.1093/sysbio/syx064.

857 66. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al.. Origins and
858 functional evolution of Y chromosomes across mammals. *Nature*. Nature Publishing Group; 2014;
859 doi: 10.1038/nature13151.

860 67. Pinto BJ, Gamble T, Smith CH, Wilson MA. A lizard is never late: squamate genomics as a recent
861 catalyst for understanding sex chromosome and microchromosome evolution. bioRxiv;

862 68. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, et al.. The sexually dimorphic on
863 the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many
864 salmonids. *Evol Appl*. 2013; doi: 10.1111/eva.12032.

865 69. Bertho S, Herpin A, Schartl M, Guiguen Y. Lessons from an unusual vertebrate sex-determining
866 gene. *Philos Trans R Soc Lond B Biol Sci*. 2021; doi: 10.1098/rstb.2020.0092.

867 70. Hart KN, Stocker WA, Nagykery NG, Walton KL, Harrison CA, Donahoe PK, et al.. Structure of
868 AMH bound to AMHR2 provides insight into a unique signaling pair in the TGF- β family. *Proc Natl
869 Acad Sci U S A*. 2021; doi: 10.1073/pnas.2104809118.

870 71. Meier M, Tokarz J, Haller F, Mindnich R, Adamski J. Human and zebrafish hydroxysteroid
871 dehydrogenase like 1 (HSDL1) proteins are inactive enzymes but conserved among species. *Chemico-
872 biological interactions*. Chem Biol Interact; 2009; doi: 10.1016/j.cbi.2008.10.036.

873 72. Xiao L, Guo Y, Wang D, Zhao M, Hou X, Li S, et al.. Beta-Hydroxysteroid Dehydrogenase Genes

874 in Orange-Spotted Grouper (*Epinephelus coioides*): Genome-Wide Identification and Expression
875 Analysis During Sex Reversal. *Front Genet.* 2020; doi: 10.3389/fgene.2020.00161.

876 73. Fan B, Xie D, Li Y, Wang X, Qi X, Li S, et al.. A single intronic single nucleotide polymorphism
877 in splicing site of steroidogenic enzyme *hsd17b1* is associated with phenotypic sex in oyster pompano,
878 *Trachinotus anak*. *Proc Biol Sci.* 2021; doi: 10.1098/rspb.2021.2245.

879 74. Koyama T, Nakamoto M, Morishima K, Yamashita R, Yamashita T, Sasaki K, et al.. A SNP in a
880 Steroidogenic Enzyme Is Associated with Phenotypic Sex in Seriola Fishes. *Current Biology*. 2019;
881 doi: 10.1016/j.cub.2019.04.069.

882 75. Purcell CM, Seetharam AS, Snodgrass O, Ortega-García S, Hyde JR, Severin AJ. Insights into
883 teleost sex determination from the *Seriola dorsalis* genome assembly. *BMC Genomics*. 2018; doi:
884 10.1186/s12864-017-4403-1.

885 76. Ko HW, Norman RX, Tran J, Fuller KP, Fukuda M, Eggenschwiler JT. Broad-Minded Links Cell
886 Cycle-Related Kinase to Cilia Assembly and Hedgehog Signal Transduction. *Developmental Cell*.
887 2010; doi: 10.1016/j.devcel.2009.12.014.

888 77. HUANG C-CJ, YAO HH-C. Diverse Functions of Hedgehog Signaling in Formation and
889 Physiology of Steroidogenic Organs. *Mol Reprod Dev.* 2010; doi: 10.1002/mrd.21174.

890 78. Guiguen Y, Fostier A, Piferrer F, Chang C-F. Ovarian aromatase and estrogens: a pivotal role for
891 gonadal sex differentiation and sex change in fish. *Gen Comp Endocrinol.* 2010; doi:
892 10.1016/j.ygcn.2009.03.002.

893 79. Baroiller JF, Guiguen Y. Endocrine and environmental aspects of sex differentiation in
894 gonochoristic fish. *EXS.* :177–201 2001;

895