



HAL
open science

Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*

Romain Feron, Margot Zahm, Cédric Cabau, Christophe Klopp, Céline Roques, Olivier Bouchez, Camille Eché, Sophie Valière, Cécile Donnadiou, Pierrick Haffray, et al.

► To cite this version:

Romain Feron, Margot Zahm, Cédric Cabau, Christophe Klopp, Céline Roques, et al.. Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. 2019. hal-02790966

HAL Id: hal-02790966

<https://hal.inrae.fr/hal-02790966v1>

Preprint submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Characterization of a Y-specific duplication/insertion of the anti-Mullerian**
2 **hormone type II receptor gene based on a chromosome-scale genome**
3 **assembly of yellow perch, *Perca flavescens*.**

4
5 Romain Feron^{1,2}, Margot Zahm³, Cédric Cabau⁴, Christophe Klopp^{3,4}, Céline Roques⁵, Olivier
6 Bouchez⁵, Camille Eché⁵, Sophie Valière⁵, Cécile Donnadiou⁵, Pierrick Haffray⁶, Anastasia
7 Bestin⁶, Romain Morvezen⁶, Hervé Acloque⁷, Peter T. Euclide⁸, Ming Wen¹, Elodie Jouano¹,
8 Manfred Schartl⁹, John H. Postlethwait¹⁰, Claire Schraidt¹¹, Mark R. Christie^{11,12}, Wes
9 Larson⁸, Amaury Herpin¹, Yann Guiguen^{1*}.

10
11 **AFFILIATIONS**

12
13 ¹ INRA, UR 1037 Fish Physiology and Genomics, F-35000 Rennes, France. ² Department of
14 Ecology and Evolution, University of Lausanne, and Swiss Institute of Bioinformatics, 1015
15 Lausanne, Switzerland. ³ Plate-forme bio-informatique Genotoul, Mathématiques et
16 Informatique Appliquées de Toulouse, INRA, Castanet Tolosan, France. ⁴ SIGENAE,
17 GenPhySE, Université de Toulouse, INRA, ENVT, Castanet Tolosan, France. ⁵ INRA, US
18 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France. ⁶ SYSAAF, French poultry and
19 aquaculture breeders, 35042, Rennes Cedex, France. ⁷ GenPhySE, Université de Toulouse,
20 INRA, INPT, ENVT, Castanet-Tolosan, France. ⁸ U. S. Geological Survey, Wisconsin
21 Cooperative Fishery Research Unit, University of Wisconsin-Stevens Point, 800 Reserve St.,
22 Stevens Point, WI 54481, USA. ⁹ Developmental Biochemistry, Biozentrum, University of
23 Würzburg, Würzburg, Germany and The *Xiphophorus* Genetic Stock Center, Department of
24 Chemistry and Biochemistry, Texas State University, San Marcos, Texas, USA. ¹⁰ Institute of
25 Neuroscience, University of Oregon, Eugene, Oregon, USA. ¹¹ Department of Forestry and
26 Natural Resources, Purdue University; 715 W. State St., West Lafayette, Indiana 47907-2054
27 USA. ¹² Department of Biological Sciences, Purdue University; 915 W. State St., West
28 Lafayette, Indiana 47907-2054 USA.

29
30 *corresponding author (yann.guiguen@inra.fr)

31
32

33 **ABSTRACT**

34

35 **Background:** Yellow perch, *Perca flavescens*, is an ecologically and commercially important
36 species native to a large portion of the northern United States and southern Canada. It is also a
37 promising candidate species for aquaculture. No yellow perch reference genome, however, has
38 been available to facilitate improvements in both fisheries and aquaculture management
39 practices.

40 **Findings:** By combining Oxford Nanopore Technologies long-reads, 10X genomics Illumina
41 short linked reads and a chromosome contact map produced with Hi-C, we generated a high-
42 continuity chromosome scale yellow perch genome assembly of 877.4 Mb. It contains, in
43 agreement with the known diploid chromosome yellow perch count, 24 chromosome-size
44 scaffolds covering 98.8% of the complete assembly (N50 = 37.4 Mb, L50 = 11). Genome
45 annotation identified 41.7% (366 Mb) of repeated elements and 24,486 genes including 16,579
46 genes (76.3%) significantly matching with proteins in public databases. We also provide a first
47 characterization of the yellow perch sex determination locus that contains a male-specific
48 duplicate of the anti-Mullerian hormone type II receptor gene (*amhr2by*) inserted at the
49 proximal end of the Y chromosome (chromosome 9). Using this sex-specific information, we
50 developed a simple PCR genotyping test which accurately differentiates XY genetic males
51 (*amhr2by*⁺) from XX genetic females (*amhr2by*⁻).

52 **Conclusions:** Our high-quality genome assembly is an important genomic resource for future
53 studies on yellow perch ecology, toxicology, fisheries, and aquaculture research. In addition,
54 the characterization of the *amhr2by* gene as a candidate sex determining gene in yellow perch
55 provides a new example of the recurrent implication of the transforming growth factor beta
56 pathway in fish sex determination, and highlights gene duplication as an important genomic
57 mechanism for the emergence of new master sex determination genes.

58

59 **KEYWORDS:** Yellow perch, evolution, whole genome sequencing, long-reads, sex-
60 determination, transforming growth factor beta, amhr2

61

62

63 DATA DESCRIPTION

64

65 Introduction and background

66

67

68 Yellow perch, *Perca flavescens* (Figure 1), is an ecologically and economically important
69 species native to a large portion of the northern United States and southern Canada. Yellow
70 perch supports recreational and commercial fisheries and is a major component of the food web
71 in many inland lakes, where they are often the most abundant prey for larger species such as
72 walleye (*Sander vitreus*), northern pike (*Esox lucius*), muskellunge (*Esox masquinongy*), and
73 lake trout (*Salvelinus namaycush*) [1]. In the Laurentian Great Lakes, yellow perch are an
74 important native species that has been heavily impacted by fishing pressure and environmental
75 changes over the last century [2,3]. Yellow perch is consistently among the most valuable
76 commercially harvested fish species in the Great Lakes (\$2.64/lb. dockside value in 2000 [4]),
77 with fillets selling as high as \$12/lb). However, many yellow perch fisheries have been forced
78 to close due to substantial population declines [5]. The mechanisms underlying these declines
79 are not fully understood but could be investigated using a combination of ecological and genetic
80 studies if adequate genomic information were available.

81 From an aquaculture perspective, yellow perch has many desirable attributes. For example,
82 yellow perch can tolerate high stocking densities, are relatively disease resistant, and can be
83 raised successfully under a variety of temperature and water conditions [6,7]. Furthermore,
84 yellow perch can be reared from hatching to marketable size in a relatively short period of time
85 (~1 year vs. 2+ years for most salmonids). Because yellow perch eat a diverse array of prey
86 items [8], their feed can be obtained from ecologically sustainable sources while remaining cost
87 effective (in contrast salmon are often fed a diet consisting primarily of other wild-caught
88 fishes, known as fish meal). Lastly, yellow perch fillets have a firm texture and a mild flavor
89 yielding a high market value.

90 The challenges faced by the yellow perch aquaculture include: increasing the spawning window
91 for broodstock, reducing early life stage mortality, and developing large-bodied strains with
92 faster growth rates [6]. Yellow perch spawn seasonally (typically in late spring to early summer)
93 during a relatively narrow period of time (1-2 weeks). From an aquaculture perspective, it can
94 be challenging to find males and females that are ready to spawn at the same time and, if the
95 fish are not monitored daily, the peak spawning period can be missed entirely [9]. Compared to

96 other aquaculture species, yellow perch also have a protracted free-swimming larval stage (~30
97 days), during which the fish require precise food and water conditions for optimal survival.
98 Developing broodstock that produce offspring with a shorter larval stage or that produce larger,
99 more robust offspring would allow perch to be successfully reared in a broader array of
100 facilities. Lastly, while yellow perch can already be grown to marketable size relatively quickly,
101 the relative lack of selective breeding means that there is considerable room for developing
102 yellow perch strains with faster growth rates and larger body sizes [6].

103 These challenges, which currently limit the wide-scale adoption of yellow perch as an
104 aquaculture species, can be addressed using cutting edge genomic resources, such as the
105 genome assembly described here. For example, one straightforward step towards obtaining fish
106 with faster growth rates and larger body size would be to produce genetically all-female
107 populations, as females grow considerably faster and larger than males [10–12]. More
108 generally, sequencing and characterizing the yellow perch genome will facilitate improvements
109 in both aquaculture and fisheries management practices.

110

111 **Results and Discussion**

112

113 **Genome characteristics**

114

115 By a combination of three approaches -- Oxford Nanopore Technologies (ONT) long-reads,
116 10X genomics Illumina short linked reads (PE150 chemistry), and a chromosome contact map
117 (Hi-C) -- we generated a high-continuity, chromosome length *de novo* genome assembly of the
118 yellow perch. Before the Hi-C integration step, the assembly yielded a genome size of 877 Mb
119 with 879 contigs, a N50 contig size of 4.3 Mb, and a L50 contig number of 60 (i.e. half of the
120 assembled genome is included in the 60 longest contigs). After Hi-C integration, the genome
121 assembled into 269 fragments with a total length of 877.4 Mb, including 24 chromosome-length
122 scaffolds representing 98.78 % of the complete genome sequence (N50 = 37.4 Mb, L50 = 11)
123 (see Table 1). Genome sizes are both very close to the 873 Mbp GenomeScope [13] estimation
124 based on short-read analysis with a repeat length of 266 Mbp (30.5%) and slightly lower than
125 the estimation of *P. flavescens* genome sizes based on C-values (900 Mbp and 1200 Mbp
126 records in the Animal Genome Size Database [14]). The 24 chromosome-length scaffolds
127 obtained after Hi-C integration are consistent with the diploid chromosome (Chr) number of
128 yellow perch ($2n = 48$) [15]. The genome completeness of these assemblies was estimated using

129 Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0 [16] based on the
130 Actinopterygii database. BUSCO scores (see Table 1) of the pre-Hi-C and post-Hi-C assemblies
131 were roughly similar (Complete BUSCOs between 97.6% and 97.8%) and with small values
132 for both fragmented (< 1%) and missing (< 1.5%) BUSCO genes.

133 Repeated elements accounted for 41.71% (366 Mbp) of our chromosomal assembly and these
134 regions were soft masked before gene annotation. Using protein, transcript, and *de novo* gene
135 prediction evidence we annotated 24,486 genes, including 16,579 (76.3%) that significantly
136 matched with a protein hit in the non-redundant NCBI database (Table 2). Our yellow perch
137 genome was also annotated with the NCBI Eukaryotic Genome Annotation Pipeline (NCBI
138 *Perca flavescens* Annotation Release 100 [17]), leading to a higher gene count (28,144) with
139 possibly multiple transcripts per gene (Table 2).

140

141 **Yellow perch sex-determination**

142

143 Yellow perch has a male monofactorial heterogametic sex determination system (XX/XY) [18]
144 with undifferentiated sex chromosomes [19]. Using a male-versus-female pooled gDNA whole
145 genome sequencing strategy [20], we identified a relatively small region of 100 kb localized at
146 the proximal end of chromosome 9 (Chr09:0-100,000 bp) with a complete absence of female
147 reads, excluding repeated elements (Fig. 2.A-B). This coverage bias strongly supports the
148 hypothesis that Chr09 is the yellow perch sex chromosome and contains a small Y-specific
149 region in phenotypic males that is completely absent from phenotypic females. Genome
150 annotation shows that this Y-specific insertion on Chr09 contains a duplicate copy (*amhr2by*)
151 of the autosomal anti-Mullerian hormone receptor gene located on Chr04 (*amhr2a*). The *amhr2*
152 gene has previously been characterized as a master sex-determining gene in some pufferfishes
153 [21,22] and the *hotei* mutation in the medaka *amhr2* gene induces a male-to-female sex-reversal
154 of genetically XY fish [23]. However, in contrast to pufferfishes, in which the differentiation
155 of X and Y chromosomes is extremely limited and originated from an allelic diversification
156 process, the yellow perch *amhr2by* sequence is quite divergent from its *amhr2a* autosomal
157 counterpart. Specifically, the *amhr2by* gene shows only 88.3 % identity with *amhr2a* in the
158 aligned coding sequence and 89.1 % in the aligned parts of the introns, but with many long gaps
159 and indels in the introns (Fig. 2C-D). This nucleotide sequence divergence impacts the protein
160 sequence of the yellow perch *amhr2by* gene (Fig. 2D-2E), but due to a complete absence of
161 exons 1 & 2 (Fig. 2C-2E) compared to its autosomal counterpart, the yellow perch Amhr2by

162 protein translates as a N-terminal-truncated type II receptor that lack most of the cysteine-rich
163 extracellular part of the receptor, which is crucially involved in ligand binding specificity [24].

164

165 To validate the male specificity of this potential Y-specific insertion, we designed primers
166 specific for both *amhr2by* and *amhr2a* and genotyped 25 male and 25 female yellow perch
167 collected from a Southeastern Lake Michigan population, which is geographically isolated from
168 the Plum Lake (Wisconsin) population of the 30 males and 30 females used for initial analysis
169 with pool-sequencing. The presence/absence of the *amhr2by* PCR product was perfectly
170 correlated with the determined phenotypic sex, with the amplification of an *amhr2by* fragment
171 only in the 25 males and no amplification in the 25 females (see Fig. 2F for 18 of the 50
172 individuals tested). The simultaneous amplification of the *amhr2a* fragment in both males and
173 in females provided an internal control preventing single-locus dropout in such a multiplexed
174 PCR reaction.

175

176 This complete sex-linkage result makes the yellow perch *amhr2by* an obvious candidate as a
177 sex determining gene. Interestingly, anti-Mullerian hormone (Amh) has been also characterized
178 as a male-promoting gene in zebrafish [25] and as a master sex determining gene both in
179 Patagonian pejerrey [26], Nile tilapia [27] and Northern pike [28]. The role of transforming
180 growth factor beta (TGF- β) members in sex determination is not limited to the Amh pathway;
181 additional TGF- β family genes have also been characterized as master sex determining genes,
182 including growth differentiation factor 6 (*gdf6*) in the turquoise killifish [29] and gonadal soma
183 derived factor (*gsdf*) in the Luzon medaka and the sablefish [30,31]. Additional evidence,
184 including loss of *amhr2by* function experiments in XY males and gain of *amhr2by* function
185 experiments in XX females, is necessary to critically test the hypothesis that this male-specific
186 *amhr2by* duplication really functions as a master sex determining gene in yellow perch.
187 However, given the known importance of the Amh pathway in fish sex determination, and that
188 no other gene in that small sex locus is known to play a role in sex differentiation, *amhr2by* is
189 a prime candidate for the yellow perch master sex determining gene. This finding provides
190 another example of the recurrent utilization of the TGF- β pathway in fish sex determination,
191 and thus supports the ‘limited option’ hypothesis [32], which states that some genes are more
192 likely than others to be selected as master sex determining genes. How this N terminal truncated
193 Amhr2 could trigger its function as a master sex determining gene is as yet unknown, but one
194 hypothesis is that this truncation constitutively activates the Amh receptor causing it to signal
195 in the absence of Amh ligand.

196

197 However, regardless of the precise role of the structural variation of *amhr2* in sex determination,
198 we have developed a simple molecular protocol for genotypically sexing perch of any life stage
199 and produced a fully annotated, chromosome-scale genome assembly that will undoubtedly aid
200 in the conservation and management of this species.

201

202 **MATERIAL AND METHODS**

203

204 **Sampling and genomic DNA extraction**

205

206 The male yellow perch used for whole genome sequencing was sampled from Plum Lake, Vilas
207 County, Wisconsin, USA. A 0.5 ml blood sample was taken from this animal and immediately
208 put in a TNES-Urea lysis buffer (TNES-Urea: 4 M urea; 10 mM Tris-HCl, pH 7.5; 125 mM
209 NaCl; 10 mM EDTA; 1% SDS) [33]. High molecular weight genomic DNA (gDNA) was then
210 purified by phenol-chloroform extraction. For the chromosome contact map (Hi-C), 1.5 ml of
211 blood was taken from a different male from a domesticated line of yellow perch raised at the
212 Farmory, an aquaculture facility in Green Bay, Wisconsin, USA. The fresh blood sample was
213 slowly cryopreserved with 15 % Dimethyl sulfoxide (DMSO) in a Mr. Frosty Freezing
214 Container (Thermo Scientific) at -80°C. Fin clip samples (30 males and 30 females) for whole-
215 genome sequencing of pools of individuals (Pool-seq) were collected from wild yellow perch
216 in Green Bay, Lake Michigan, Wisconsin, USA, placed in 90% ethanol and then stored dried
217 until gDNA extraction was performed using the NucleoSpin Kit for Tissue (Macherey-Nagel,
218 Duren, Germany). Genomic DNAs from individual fish were then quantified using a Qubit
219 fluorometer (ThermoFisher) and pooled in equimolar ratios by individual and sex, resulting in
220 one gDNA pool for males and one gDNA pool for females. For validation of *amhr2* sex-
221 linkage, 50 phenotypically sexed individuals (25 males and 25 females) wild perch from Lake
222 Michigan were collected in May of 2018 using gill net sets off the shore of Michigan City,
223 Indiana (41°42.5300'N, 86°57.5843'W). Upon collection, each individual fish was euthanized,
224 phenotypic sex was determined by visual inspection of gonads during necropsy, and caudal fin
225 clips were taken from each yellow perch individual and stored in 95% non-denatured ethanol.
226 Genomic DNA was extracted using the DNeasy extraction kit and protocol (Qiagen).

227

228 **DNA library construction and sequencing**

229 **Nanopore sequencing**

230 The quality and purity of gDNA was assessed using spectrophotometry, fluorometry and
231 capillary electrophoresis. Additional purification steps were performed using AMPure XP
232 beads (Beckman Coulter). All library preparations and sequencing were performed using
233 Oxford Nanopore Ligation Sequencing Kits SQK-LSK108 (Oxford Nanopore Technology) (14
234 flowcells) or SQK-LSK109 (2 flowcells) according to the manufacturer's instructions. For the
235 SQK-LSK108 sequencing Kit, 140 µg of DNA was purified and then sheared to 20 kb using
236 the megaruptor system (Diagenode). For each library, a DNA-damage-repair step was
237 performed on 5 µg of DNA. Then an END-repair+dA-tail-of-double-stranded-DNA-fragments
238 step was performed and adapters were ligated to DNAs in the library. Libraries were loaded
239 onto two R9.5 and twelve R9.4 flowcells and sequenced on a GridION instrument at a
240 concentration of 0.1 pmol for 48 h. For the SQK-LSK109 sequencing Kit, 10 µg of DNA was
241 purified and then sheared to 20 kb using the megaruptor system (Diagenode). For each library,
242 a one-step-DNA-damage repair+END-repair+dA-tail-of-double-stranded-DNA-fragments
243 procedure was performed on 2 µg of DNA. Adapters were then ligated to DNAs in the library.
244 Libraries were loaded on R9.4.1 flowcells and sequenced on either a GridION or PromethION
245 instrument at a concentration of 0.05 pmol for 48h or 64h respectively. The 15 GridION
246 flowcells produced 69.4 Gb of data and the PromethION flowcell produced 65.5 Gb of data.

247 **10X Genomics sequencing**

248 The Chromium library was prepared according to 10X Genomics' protocols using the Genome
249 Reagent Kit v2. Sample quantity and quality controls were validated by Qubit, Nanodrop and
250 Femto Pulse machines. The library was prepared from 10 ng of high molecular weight (HMW)
251 gDNA. Briefly, in the microfluidic Genome Chip, a library of Genome Gel Beads was
252 combined with HMW template gDNA in master mix and partitioning oil to create Gel Bead-
253 In-EMulsions (GEMs) in the Chromium apparatus. Each Gel Bead was then functionalized with
254 millions of copies of a 10x™ barcoded primer. Dissolution of the Genome Gel Bead in the
255 GEM released primers containing (i) an Illumina R1 sequence (Read 1 sequencing primer), (ii)
256 a 16 bp 10x Barcode, and (iii) a 6 bp random primer sequence. R1 sequence and the 10x™
257 barcode were added to the molecules during the GEM incubation. P5 and P7 primers, R2
258 sequence, and Sample Index were added during library construction. 10 cycles of PCR were
259 applied to amplify the library. Library quality was assessed using a Fragment Analyser and

260 library was quantified by qPCR using the Kapa Library Quantification Kit. The library was then
261 sequenced on an Illumina HiSeq3000 using a paired-end read length of 2x150 nt with the
262 Illumina HiSeq3000 sequencing kits and produced 315 million read pairs.

263 **Hi-C sequencing**

264 *In situ* Hi-C was performed according to previously described protocols [34]. Cryopreserved
265 blood cells were defrosted, washed with PBS twice and counted. 5 million cells were then cross-
266 linked with 1% formaldehyde in PBS, quenched with Glycine 0.125M and washed twice with
267 PBS. Membranes were then disrupted with a Dounce pestle, nuclei were permeabilized using
268 0.5% SDS and then digested with *HindIII* endonuclease. 5'-overhangs at *HindIII*-cut restriction
269 sites were filled-in, in the presence of biotin-dCTP with the Klenow large fragment, and then
270 re-ligated at a *NheI* restriction site. Nuclei were lysed and DNA was precipitated and then
271 purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Qubit
272 fluorometric quantification system (Thermo). T4 DNA polymerase was used to remove un-
273 ligated biotinylated ends. Then the Hi-C library was prepared according to Illumina's protocols
274 using the Illumina TruSeq Nano DNA HT Library Prep Kit with a few modifications: 1.4µg
275 DNA was fragmented to 550nt by sonication. Sheared DNA was then sized (200-600pb) using
276 Agencourt AMPure XP beads, and biotinylated ligation junctions were captured using M280
277 Streptavidin Dynabeads (Thermo) and then purified using reagents from the Nextera Mate Pair
278 Sample preparation kit (Illumina). Using the TruSeq nano DNA kit (Illumina), the 3' ends of
279 blunt fragments were adenylated. Next, adaptors and indexes were ligated and the library was
280 amplified for 10 cycles. Library quality was assessed by quantifying the proportion of DNA cut
281 by endonuclease *NheI* using a Fragment Analyzer (Advanced Analytical Technologies, Inc.,
282 Iowa, USA). Finally, the library was quantified by qPCR using the Kapa Library Quantification
283 Kit (Roche). Sequencing was performed on an Illumina HiSeq3000 apparatus (Illumina,
284 California, USA) using paired-end 2x150 nt reads. This produced 128 million read pairs (38.4
285 Gb of raw nucleotides).

286 **Pool sequencing**

287 Pool-sequencing libraries were prepared according to Illumina's protocols using the Illumina
288 TruSeq Nano DNA HT Library Prep Kit (Illumina, California, USA). In short, 200 ng of each
289 gDNA pool (males and females pools) was fragmented to 550 bp by sonication on M220
290 Focused-ultrasonicator (COVARIS). Size selection was performed using SPB beads (kit beads)

291 and the 3' ends of blunt fragments were mono-adenylated. Then, adaptors and indexes were
292 ligated and the construction was amplified with Illumina-specific primers. Library quality was
293 assessed using a Fragment Analyzer (Advanced Analytical Technologies, Inc., Iowa, USA) and
294 libraries were quantified by qPCR using the Kapa Library Quantification Kit (Roche).
295 Sequencing was performed on a NovaSeq (Illumina, California, USA) using a paired-end read
296 length of 2x150 nt with Illumina NovaSeq Reagent Kits. Sequencing produced 119 million
297 paired reads for the male pool library and 132 million paired reads for the female pool library.

298

299 **Genome assembly and analysis**

300 **Genome size estimation**

301 K-mer-based estimation of the genome size was carried out with GenomeScope [13]. 10X reads
302 were processed with Jellyfish v1.1.11 [35] to count 17-, 19-, 21-, 23- and 25-mers with a max
303 k-mer coverage of 10,000.

304 **Genome assembly**

305 GridION and PromethION data were trimmed using Porechop v0.2.1 [36], corrected using
306 Canu v1.6 [37] and filtered to keep only reads longer than 10 kbp. Corrected reads were then
307 assembled using SmartDeNovo version of May-2017 [38] with default parameters. The
308 assembly base pair quality was improved by several polishing steps including two rounds of
309 long read alignment to the draft genome with minimap2 v2.7 [39] followed by Racon v1.3.1
310 [40], as well as three rounds of 10X genomics short read alignments using Long Ranger v2.1.1
311 (10x Genomics 2018) followed by Pilon v1.22 [41]. The polished genome assembly was then
312 scaffolded using Hi-C as a source of linking information. Reads were aligned to the draft
313 genome using Juicer [42] with default parameters. A candidate assembly was then generated
314 with 3D de novo assembly (3D-DNA) pipeline [43] with the -r 0 parameter. Finally, the
315 candidate assembly was manually reviewed using Juicebox Assembly Tools [42]. Genome
316 completeness was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO)
317 v3.0 [16] based on 4,584 BUSCO orthologs derived from the Actinopterygii lineage.

318 **Genome annotation**

319 The first annotation step was to identify repetitive content using RepeatMasker v4.0.7 [43],
320 Dust (Kuzio et al., unpublished but described in [44]), and TRF v4.09 [45]. A species-specific
321 *de novo* repeat library was built with RepeatModeler v1.0.11 [46] and repeated regions were
322 located using RepeatMasker with the *de novo* and *Danio rerio* libraries. Bedtools v2.26.0 [47]
323 was used to merge repeated regions identified with the three tools and to soft mask the genome.
324 The MAKER3 genome annotation pipeline v3.01.02-beta [48] combined annotations and
325 evidence from three approaches: similarity with fish proteins, assembled transcripts (see
326 below), and *de novo* gene predictions. Protein sequences from 11 fish species (*Astyanax*
327 *mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*,
328 *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon*
329 *nigroviridis*, *Xiphophorus maculatus*) found in Ensembl were aligned to the masked genome
330 using Exonerate v2.4 [49]. As *Perca fluviatilis* is a relatively closely related species from *P.*
331 *flavescens* (divergence time is estimated to be 19.8 million years ago according to [50]), RNA-
332 Seq reads of *P. fluviatilis* (NCBI BioProject PRJNA256973) from the PhyloFish project [51]
333 were used for genome annotation and aligned to the chromosomal assembly using STAR
334 v2.5.1b [52] with outWigType and outWigStrand options to output signal wiggle files.
335 Cufflinks v2.2.1 [53] was used to assemble the transcripts which were used as RNA-seq
336 evidence. Braker v2.0.4 [54] provided *de novo* gene models with wiggle files provided by
337 STAR as hint files for GeneMark and Augustus training. The best supported transcript for each
338 gene was chosen using the quality metric called Annotation Edit Distance (AED) [55]. Genome
339 annotation gene completeness was assessed by BUSCO using the Actinopterygii group. Finally,
340 predicted genes were subjected to similarity searches against the NCBI NR database using
341 Diamond v0.9.22 [56]. The top hit with a coverage over 70% and identity over 80% was
342 retained.

343 **Pool-sequencing analysis**

344 Reads from the male and female pools were aligned to the chromosomal assembly with BWA
345 mem (version 0.7.17, [57]), and the resulting BAM files were sorted and PCR duplicates
346 removed using Picard tools (version 2.18.2). A file containing the nucleotide composition of
347 each pool for each genomic position was generated using samtools mpileup (version 1.8, [58])
348 and popoolation2 mpileup2sync (version 1201, [59]). This file was then analyzed with custom
349 software (PSASS version 2.0.0 [60]) to compute 1) the position and density of sex-specific
350 SNPs, defined as SNPs heterozygous in one sex while homozygous in the other sex, 2) absolute

351 and relative read depth for the male and female pools along the genome, and 3) F_{ST} between
352 males and females in windows along the genome. PSASS was run with default parameters
353 except --window-size which was set to 5000 and --output-resolution which was set to 1000.

354 **Validation of *amhr2by* sex-linkage**

355

356 To validate the sex-linkage of *amhr2by* in males suggested by the pool-sequencing results, two
357 primer sets were designed based on the alignment of yellow perch *amhr2a* and *amhr2by* genes
358 with one primer pair specific for the autosomal *amhr2a* gene (forward: 5'-
359 GGGAAACGTGGGAAACTCAC-3', and reverse: 5'-AGCAGTAGTTACAGGGCACA-3',
360 expected fragment size: 638 bp) and one primer pair specific for the Y chromosomal *amhr2by*
361 gene (forward: 5'-TGGTGTGTGGCAGTGATACT-3', and reverse: 5'-
362 ACTGTAGTTAGCGGGCACAT-3', expected fragment size: 443 bp). Gene alignments were
363 run with mVISTA [61]. Primers were sourced from Integrated Data Technologies (IDT). All
364 samples were run blind with respect to phenotypic sex; the male and female samples were
365 randomized, and their phenotypic sex was not cross referenced with field data until gel
366 electrophoresis was run on the final PCR products. Genotyping was carried out on each gDNA
367 sample using a multiplexed PCR approach. The PCR reaction solution was composed of 50 μ l
368 of PCR Master Mix (Quiagen), 10 μ l of each primer (40 μ l total), and 10 μ l of gDNA
369 (concentrations of gDNA ranging from 150 to 200 ng/ μ l) for a total reaction volume of 100 μ l.
370 Thermocycling conditions were 1 cycle of 3 min at 94°C, followed by 35 cycles of 30 sec at
371 94°C, 30 sec at 51°C, and 1 min at 72°C, and finishing with 10 min incubation at 72°C. PCR
372 products were loaded on a 1.5 % agarose gel, run at 100V for 45 minutes and visualized with a
373 UVP UVsolo touch UV box.

374

375 **Availability of supporting data**

376

377 This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the
378 accession SCKG00000000. The version described in this paper is version SCKG01000000. Hi-
379 C, 10X genomics and pool-sequencing Illumina reads, and Oxford Nanopore Technologies
380 genome raw reads are available in the Sequence Read Archive (SRA), under BioProject
381 reference PRJNA514308.

382

383 **Author contributions**

384

385 YG, MS, and JHP designed the project. WL, CS and MC collected the samples, EJ, MW, CR,
386 OB, SV and HA extract the gDNA, made the genomic libraries and sequenced them. RF, CC,
387 CK, MZ, PE, AH and YG processed the genome assemblies and / or analyzed the results. CS
388 and MC checked sex-linkage of *amhr2by* on yellow perch samples. YG, RF, WL, MC, JHP,
389 CC, CK, and CR wrote the manuscript. MS, JHP, CD, PH, AB, RM, MC and YG, supervised
390 the project administration and raised funding. All the authors read and approved the final
391 manuscript.

392

393 **Competing interests**

394

395 All authors declare no competing interests.

396

397 **Acknowledgements**

398

399 This project was supported by funds from the “Agence Nationale de la Recherche” and the
400 “Deutsche Forschungsgemeinschaft” (ANR/DFG, PhyloSex project, 2014-2016), the CRB-
401 Anim “Centre de Ressources Biologiques pour les animaux domestiques” project
402 PERCH’SEX, the FEAMP “Fonds européen pour les affaires maritimes et la pêche” project
403 SEX’NPERCH, and R01 GM085318 from the National Institutes of Health, USA. Additional
404 funding was provided to MRC from the Great Lakes Fishery Commission, Project ID:
405 2018_CHR_44072. The GeT core facility, Toulouse, France was supported by France
406 Génomique National infrastructure, funded as part of “Investissement d’avenir” program
407 managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). We are grateful
408 to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for
409 providing computing and/or storage resources. Any use of trade, product, or company name is
410 for descriptive purposes only and does not imply endorsement by the U.S. Government.

411

412

413 **REFERENCES**

414

415 1. Becker GC. Fishes of Wisconsin. Madison, Wisconsin: University of Wisconsin Press, 1983,
416 1052 p. [http://digioll.library.wisc.edu/cgi-bin/EcoNatRes/EcoNatRes-](http://digioll.library.wisc.edu/cgi-bin/EcoNatRes/EcoNatRes-idx?id=EcoNatRes.FishesWI)
417 [idx?id=EcoNatRes.FishesWI](http://digioll.library.wisc.edu/cgi-bin/EcoNatRes/EcoNatRes-idx?id=EcoNatRes.FishesWI). Accessed 7 Feb 2019.

- 418 2. Evans MS. Recent Major Declines in Zooplankton Populations in the Inshore Region of Lake
419 Michigan: Probable Causes and Implications. *Can J Fish Aquat Sci.* 1986;43:154–9.
- 420 3. Baldwin N, Saalfeld R, Dochoda M, Buettner H, Eshenroder L. Commercial Fish Production
421 in the Great Lakes 1867–2009. Great Lakes Fishery Commission, Ann Arbor, MI. 2009.
- 422 4. Kinnunen RE. Great Lakes Commercial Fisheries. 2003. Michigan Sea Grant Extension,
423 Marquette.
424 <https://pdfs.semanticscholar.org/d243/74bc18e240e29a8418e96e766ea8e8be50ed.pdf>.
425 Accessed 19 Jul 2019.
- 426 5. Marsden JE, Robillard SR. Decline of Yellow Perch in Southwestern Lake Michigan, 1987–
427 1997. *North American Journal of Fisheries Management.* 2004;24:952–66.
- 428 6. Malison JA. A white paper on the status and needs of yellow perch aquaculture in the North
429 Central Region. North Central Regional Aquaculture Center, Michigan State University, East
430 Lansing. 2003.
- 431 7. Malison JA, Held JA. Effects of fish size at harvest, initial stocking density and tank lighting
432 conditions on the habituation of pond-reared yellow perch (*Perca flavescens*) to intensive
433 culture conditions. *Aquaculture.* 1992;104:67–78.
- 434 8. Keast A. Diet overlaps and feeding relationships between the year classes in the yellow perch
435 (*Perca flavescens*). *Environ Biol Fish.* 1977;2:53–70.
- 436 9. Źarski D, Horváth A, Held JA, Kucharczyk D. Artificial Reproduction of Percid Fishes. In:
437 Kestemont P, Dabrowski K, Summerfelt RC, editors. *Biology and Culture of Percid Fishes:*
438 *Principles and Practices.* Dordrecht: Springer Netherlands; 2015. p. 123–61.
- 439 10. Malison JA, Garcia-Abiado M a. R. Sex control and ploidy manipulations in yellow perch
440 (*Perca flavescens*) and walleye (*Stizostedion vitreum*). *Journal of Applied Ichthyology.*
441 1996;12:189–94.
- 442 11. Rougeot C. Sex and Ploidy Manipulation in Percid Fishes. In: Kestemont P, Dabrowski K,
443 Summerfelt RC, editors. *Biology and Culture of Percid Fishes: Principles and Practices.*
444 Dordrecht: Springer Netherlands; 2015. p. 625–34.
- 445 12. Malison JA, Kayes TB, Wentworth BC, Amundson CH. Growth and Feeding Responses of
446 Male versus Female Yellow Perch (*Perca flavescens*) Treated with Estradiol-17 β . *Can J Fish*
447 *Aquat Sci.* 1988;45:1942–8.
- 448 13. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al.
449 GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics.*
450 2017;33:2202–4.
- 451 14. Animal Genome Size Database. <http://www.genomesize.com/index.php>. Accessed 7 Feb
452 2019.
453
- 454 15. Ráb P, Roth P, Mayr B. Karyotype Study of Eight Species of European Percid Fishes
455 (Pisces, Percidae). *Caryologia.* 1987;40:307–18.
- 456 16. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:

- 457 assessing genome assembly and annotation completeness with single-copy orthologs.
458 *Bioinformatics*. 2015;31:3210–2.
- 459 17. *Perca flavescens* Annotation Report.
460 https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Perca_flavescens/100/. Accessed 5 Jul
461 2019.
- 462 18. Malison JA, Kayes TB, Best CD, Amundson CH, Wentworth BC. Sexual Differentiation
463 and Use of Hormones to Control Sex in Yellow Perch (*Perca flavescens*). *Can J Fish Aquat*
464 *Sci*. 1986;43:26–35.
- 465 19. Beçak ML, Beçak W, Roberts FL, Shoffner RN, Volpe EP. *Perca flavescens* (Mitchill)
466 (Yellow perch) 2n = 48. In: Beçak ML, Beçak W, Roberts FL, Shoffner RN, Volpe EP, editors.
467 *Chromosome Atlas: Fish, Amphibians, Reptiles, and Birds: Volume 2*. Berlin, Heidelberg:
468 Springer Berlin Heidelberg; 1973.
- 469 20. Gammerding WJ, Conte MA, Baroiller J-F, D’Cotta H, Kocher TD. Comparative analysis
470 of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*. *BMC Genomics*.
471 2016;17:808.
- 472 21. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, et al. A trans-species
473 missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu*
474 *rubripes* (fugu). *PLoS Genet*. 2012;8:e1002798.
- 475 22. Ieda R, Hosoya S, Tajima S, Atsumi K, Kamiya T, Nozawa A, et al. Identification of the
476 sex-determining locus in grass puffer (*Takifugu niphobles*) provides evidence for sex-
477 chromosome turnover in a subset of *Takifugu* species. *PLOS ONE*. 2018;13:e0190635.
- 478 23. Morinaga C, Saito D, Nakamura S, Sasaki T, Asakawa S, Shimizu N, et al. The *hotei*
479 mutation of medaka in the anti-Müllerian hormone receptor causes the dysregulation of germ
480 cell and sexual development. *Proc Natl Acad Sci USA*. 2007;104:9691–6.
- 481 24. Heldin C-H, Moustakas A. Signaling Receptors for TGF- β Family Members. *Cold Spring*
482 *Harb Perspect Biol*. 2016;8:a022053.
- 483 25. Yan Y-L, Batzel P, Titus T, Sydes J, Desvignes T, Bremiller R, et al. The roles of *Amh* in
484 zebrafish gonad development and sex determination. *bioRxiv*. 2019;650218.
- 485 26. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, et al. A Y-linked anti-
486 Müllerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad*
487 *Sci USA*. 2012;109:2955–9.
- 488 27. Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, et al. A Tandem Duplicate of Anti-Müllerian
489 Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination
490 in Nile Tilapia, *Oreochromis niloticus*. *PLOS Genetics*. 2015;11:e1005678.
- 491 28. Pan Q, Feron R, Yano A, Guyomard R, Jouanno E, Vigouroux E, et al. Identification of the
492 master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome
493 differentiation. *bioRxiv*. 2019;549527.
- 494 29. Reichwald K, Petzold A, Koch P, Downie BR, Hartmann N, Pietsch S, et al. Insights into
495 Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell*.

- 496 2015;163:1527–38.
- 497 30. Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, et al. Tracing the
498 emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics*.
499 2012;191:163–70.
- 500 31. Rondeau EB, Messmer AM, Sanderson DS, Jantzen SG, von Schalburg KR, Minkley DR,
501 et al. Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny,
502 linkage map and identification of a putative sex gene. *BMC Genomics*. 2013;14:452.
- 503 32. Marshall Graves JA, Peichel CL. Are homologies in vertebrate sex determination due to
504 shared ancestry or to limited options? *Genome Biol*. 2010;11:205.
- 505 33. Asahida T, Kobayashi T, Saitoh K, Nakayama I. Tissue Preservation and Total DNA
506 Extraction from Fish Stored at Ambient Temperature Using Buffers Containing High
507 Concentration of Urea. *Fisheries science*. 1996;62:727–730.
- 508 34. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Transcriptome and
509 chromatin structure annotation of liver, CD4+ and CD8+ T cells from four livestock species.
510 *bioRxiv*. 2019;316091.
- 511 35. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of
512 occurrences of k-mers. *Bioinformatics*. 2011;27:764–70.
- 513 36. Wick R. adapter trimmer for Oxford Nanopore reads. <https://github.com/rrwick/Porechop>.
514 Accessed 7 Feb 2019.
- 515 37. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and
516 accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*.
517 2017;27:722–36.
- 518 38. Ruan J. Ultra-fast de novo assembler using long noisy reads: ruanjue/smarddenovo.
519 <https://github.com/ruanjue/smarddenovo>. Accessed 7 Feb 2019.
- 520 39. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*.
521 2018;34:3094–100.
- 522 40. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from
523 long uncorrected reads. *Genome Res*. 2017;27:737–46.
- 524 41. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an
525 integrated tool for comprehensive microbial variant detection and genome assembly
526 improvement. *PLoS ONE*. 2014;9:e112963.
- 527 42. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer
528 Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*.
529 2016;3:95–8.
- 530 43. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo
531 assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds.
532 *Science*. 2017;356:92–5.

- 533 44. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A fast and symmetric DUST
534 implementation to mask low-complexity DNA sequences. *J Comput Biol.* 2006;13:1028–40.
- 535 45. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids*
536 *Res.* 1999;27:573–80.
- 537 46. RepeatModeler Download Page. <http://www.repeatmasker.org/RepeatModeler/>. Accessed
538 23 Mar 2019.
- 539 47. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
540 features. *Bioinformatics.* 2010;26:841–2.
- 541 48. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management
542 tool for second-generation genome projects. *BMC Bioinformatics.* 2011;12:491.
- 543 49. Slater GSC, Birney E. Automated generation of heuristics for biological sequence
544 comparison. *BMC Bioinformatics.* 2005;6:31.
- 545 50. Couture P, Pyle G, Pyle G. Evolutionary Relationships, Population Genetics, and Ecological
546 and Genomic Adaptations of Perch (*Perca*). *Biology of Perch.* 2015. Available from:
547 <https://www.taylorfrancis.com/>. Accessed 19 July 2019.
- 548 51. Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, et al. Gene evolution and
549 gene expression after whole genome duplication in fish: the PhyloFish database. *BMC*
550 *Genomics.* 2016;17:368.
- 551 52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
552 universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
- 553 53. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript
554 assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform
555 switching during cell differentiation. *Nat Biotechnol.* 2010;28:511–5.
- 556 54. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised
557 RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics.*
558 2016;32:767–9.
- 559 55. Eilbeck K, Moore B, Holt C, Yandell M. Quantitative measures for the management and
560 comparison of annotated genomes. *BMC Bioinformatics.* 2009;10:67.
- 561 56. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat*
562 *Methods.* 2015;12:59–60.
- 563 57. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
564 arXiv:13033997 [q-bio]. <http://arxiv.org/abs/1303.3997>. 2013. Accessed 24 May 2019.
- 565 58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
566 Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
- 567 59. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between
568 populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.*
569 2011;27:3435–6.

570 60. Feron R. RomainFeron/PSASS: Version 2.0.0. Zenodo; 2019.
571 https://zenodo.org/record/2615936#XTyIS3s6_AI. Accessed 27 Jul 2019.

572 61. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for
573 comparative genomics. *Nucleic Acids Res.* 2004;32:W273-279.

574 62. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment
575 on a microcomputer. *Gene.* 1988;73:237-44.

576

577

578 **Tables**

579

580 **Table 1. Yellow perch assembly statistics.**

Assembly metrics	Pre Hi-C	Post Hi-C
Number of reads	3,118,677	3,118,677
Total size of reads	49,450,446,732	49,450,446,732
Number of contigs	879	269
Total size of the assembly	877,025,633	877,440,133
Longest fragment	18,280,501	44,580,961
Shortest fragment	160	160
Mean fragment size	997,754	3,261,859
Median fragment size	216,440	15,167
N50 fragment length	4,304,620	37,412,490
L50 fragment count	60	11
Assembly completeness	Pre Hi-C	Post Hi-C
Complete BUSCOs	4,482 (97.8%)	4,472 (97.6%)
Complete and single-copy BUSCOs	4,371 (95.4%)	4,363 (95.2%)
Complete and duplicated BUSCOs	111 (2.4%)	109 (2.4%)
Fragmented BUSCOs	47 (1%)	41 (0.9%)
Missing BUSCOs	55 (1.2%)	71 (1.5%)

581

582

583 **Table 2. Yellow perch annotation statistics.**

Gene annotation	This paper	NCBI
Number of genes	24,486	28,144
Number of mRNA	21,723	42,926
Number of tRNA	2,763	1,250
Transcriptome size	56,137,542 bp	138,437,341 bp
Mean transcript length	2,292 bp	2,938 bp
Longest transcript	67,783 bp	94,494 bp
Number of coding genes with significant hit against NCBI NR	16,579 (76.3%)	20,992 (88.4%)
Gene completeness (Actinopterygii dataset)		
Complete BUSCOs	4,287 (93.5%)	4,555 (99.4%)
Fragmented BUSCOs	87 (1.9%)	18 (0.4%)
Missing BUSCOs	210 (4.6%)	11 (0.2%)

584

585

586

587

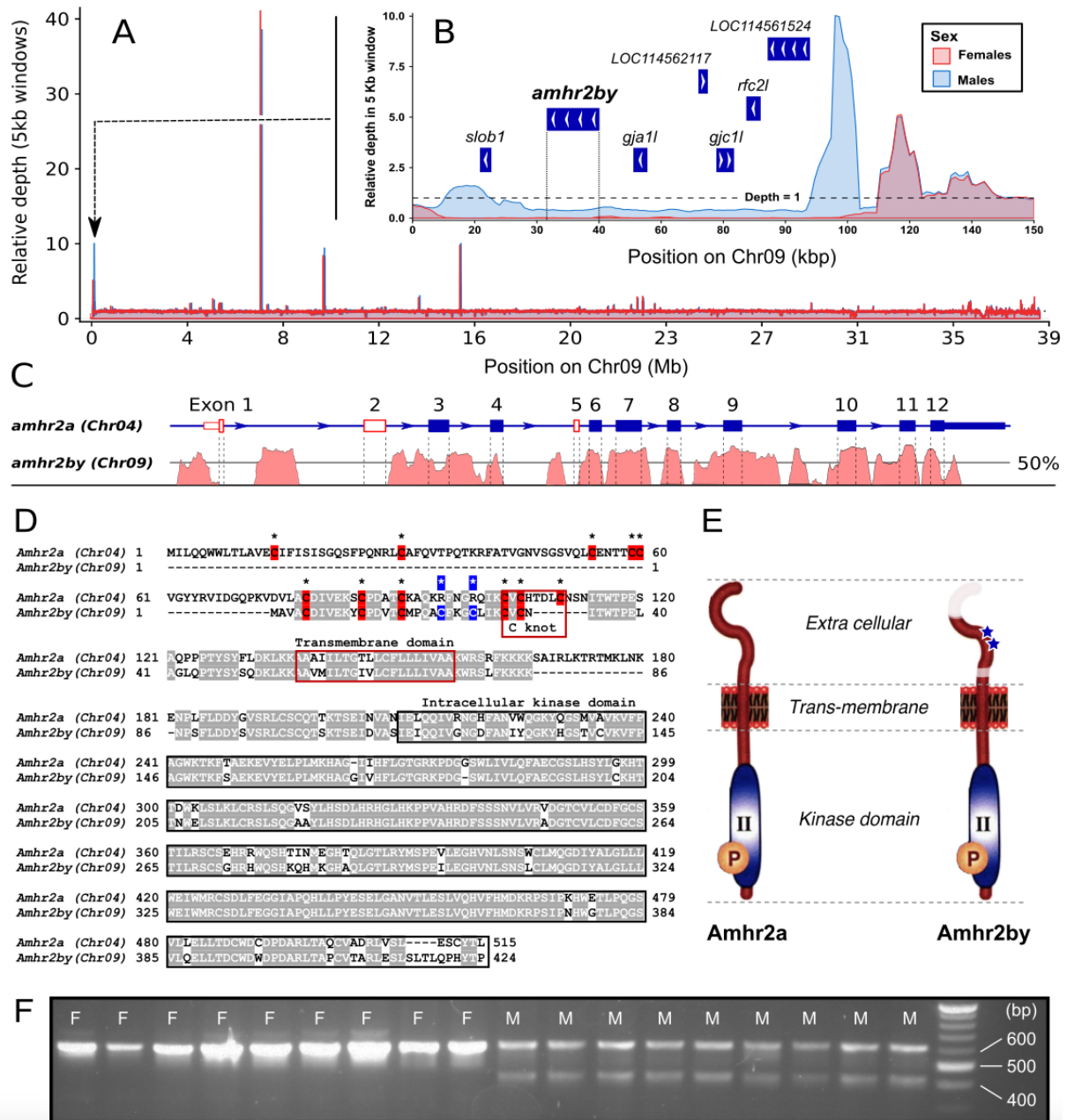
588 **Figures**

589



590

591 **Figure 1: Adult yellow perch (*Perca flavescens*).**



592
593

594 **Figure 2: Characterization of a Y-specific duplication/insertion of the anti-Mullerian**
 595 **hormone receptor gene (*amhr2by*) in yellow perch.** **A.** Pool-seq data illustrating relative read
 596 depth across chromosome 9 (Chr09) for the male (blue line) and female (red line) pools showing
 597 a coverage difference between males (blue area) and females (red area) in the first 100 kb of
 598 Chr09. **B.** Zoom-in on the read depth difference between males and females in the first 150 kb
 599 of Chr09. Gene annotation is represented by blue boxes with arrows to indicate transcript
 600 orientation (NCBI *Perca flavescens* Annotation Release 100 [13]). Abbreviations: *slob1*
 601 (probable inactive serine/threonine-protein kinase *slob1*, *LOC114561790*), *amhr2by* (anti-
 602 Mullerian hormone type-2 receptor-like, *LOC114561927*), *gja1* (gap junction alpha-1 protein-
 603 like, *LOC114562210*), *gjc1* (gap junction gamma-1 protein-like, *LOC114562012*), *rfc2l*

604 (replication factor C subunit 2-like, *LOC114561955*). **C.** Identity plot of the alignment of
605 *amhr2by* gene sequence (Chr09, bottom panel) with the autosomal *amhr2a* gene sequence
606 (Chr04, top panel). The structure of the *amhr2a* gene is depicted with blue boxes (exons, E1 to
607 E12) and blue lines (introns) with arrows indicating the transcription orientation. The solid line
608 on the identity plot (bottom panel) represents 50% nucleotide identity between the two
609 sequences. **D.** ClustalW [62] alignment of Amhr2a and Amhr2by proteins. Identical amino-
610 acids are shaded and the cysteines in the extracellular domain of Amhr2 are shown with bolded
611 black asterisks. Additional cysteines specific to Amhr2by are highlighted in blue. The different
612 domains of the receptor are boxed. **E.** Schematic representation of the two yellow perch Amhr2
613 proteins showing that the main differences impact the extracellular domain with parts missing
614 in Amhr2by represented in white and the two additional cysteines represented by blue asterisks.
615 **F.** Validation of *amhr2by* sex linkage in yellow perch. Agarose gel electrophoresis of multiplex
616 PCR of *amhr2a* (higher size PCR fragment, 638 bp), and *amhr2by* (lower size PCR fragment,
617 443 bp) in nine females (F, left side) and nine males (M, right side) genomic DNA.
618
619