



HAL
open science

Multi-genome comparisons reveal gain-and-loss evolution of anti-Mullerian hormone receptor type 2 as a candidate master sex-determining gene in Percidae

Heiner Kuhl, Peter Euclide, Christophe Klopp, Cédric Cabau, Margot Zahm, Céline Lopez-Roques, Carole Iampietro, Claire Kuchly, Cécile Donnadiou, Romain Feron, et al.

► To cite this version:

Heiner Kuhl, Peter Euclide, Christophe Klopp, Cédric Cabau, Margot Zahm, et al.. Multi-genome comparisons reveal gain-and-loss evolution of anti-Mullerian hormone receptor type 2 as a candidate master sex-determining gene in Percidae. *BMC Biology*, 2024, 22 (1), pp.141. 10.1186/s12915-024-01935-9 . hal-04638789

HAL Id: hal-04638789

<https://hal.inrae.fr/hal-04638789>

Submitted on 8 Jul 2024

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.







Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Open Access



Multi-genome comparisons reveal gain-and-loss evolution of *anti-Mullerian hormone receptor type 2* as a candidate master sex-determining gene in Percidae

Heiner Kuhl^{1*} , Peter T. Euclide² , Christophe Klopp³ , Cédric Cabau⁴ , Margot Zahm³ , Céline Lopez-Roques⁵ , Carole Iampietro⁵ , Claire Kuchly⁵ , Cécile Donnadieu⁵ , Romain Feron^{6,7} , Hugues Parrinello⁸ , Charles Poncet⁹ , Lydia Jaffrelo⁹, Carole Confolent⁹, Ming Wen^{10,11} , Amaury Herpin¹⁰ , Elodie Jouanno¹⁰ , Anastasia Bestin¹² , Pierrick Haffray¹² , Romain Morvezen¹² , Taina Rocha de Almeida¹³ , Thomas Lecocq¹³ , Bérénice Schaerlinger¹³ , Dominique Chardard¹³ , Daniel Źarski¹⁴ , Wesley A. Larson¹⁵ , John H. Postlethwait¹⁶ , Serik Timirkhanov¹⁷ , Werner Kloas¹ , Sven Wuertz¹ , Matthias Stöck¹ and Yann Guiguen^{10*}

Abstract

Background The Percidae family comprises many fish species of major importance for aquaculture and fisheries. Based on three new chromosome-scale assemblies in *Perca fluviatilis*, *Perca schrenkii*, and *Sander vitreus* along with additional percid fish reference genomes, we provide an evolutionary and comparative genomic analysis of their sex-determination systems.

Results We explored the fate of a duplicated anti-Mullerian hormone receptor type-2 gene (*amhr2bY*), previously suggested to be the master sex-determining (MSD) gene in *P. flavescens*. Phylogenetically related and structurally similar *amhr2* duplicates (*amhr2b*) were found in *P. schrenkii* and *Sander lucioperca*, potentially dating this duplication event to their last common ancestor around 19–27 Mya. In *P. fluviatilis* and *S. vitreus*, this *amhr2b* duplicate has been likely lost while it was subject to amplification in *S. lucioperca*. Analyses of the *amhr2b* locus in *P. schrenkii* suggest that this duplication could be also male-specific as it is in *P. flavescens*. In *P. fluviatilis*, a relatively small (100 kb) non-recombinant sex-determining region (SDR) was characterized on chromosome 18 using population-genomics approaches. This SDR is characterized by many male-specific single-nucleotide variations (SNVs) and no large duplication/insertion event, suggesting that *P. fluviatilis* has a male heterogametic sex-determination system (XXXY), generated by allelic diversification. This SDR contains six annotated genes, including three (*c18h1orf198*, *hsl1*, *tbc1d32*) with higher expression in the testis than in the ovary.

Conclusions Together, our results provide a new example of the highly dynamic sex chromosome turnover in teleosts and provide new genomic resources for Percidae, including sex-genotyping tools for all three known *Perca* species.

Keywords Sex-determination, Genome, Perches, Pikeperches, Sex chromosomes

*Correspondence:

Heiner Kuhl
Heiner.Kuhl@igb-berlin.de
Yann Guiguen
yann.guiguen@inrae.fr

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

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

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

Genes encoding many members of the transforming growth factor beta (TGF- β) gene family, including anti-Mullerian hormone (*amh*) and anti-Mullerian hormone receptor type-2 (*amhr2*), have repeatedly and independently evolved as MSD genes in vertebrates [10]. For instance, *amh* has been characterized or suspected to be the MSD gene in pikes [11, 12], Nile tilapia [13], lumpfish

[14], *Sebastes* rockfish [15], lingcod [16], and Patagonian pejerrey [17]. The cognate receptor gene of Amh, *amhr2*, has also been found as a potential MSD gene in Pangasiidae [18], Takifugu [19], Ayu [20], common seadragon and alligator pipefish [21], and in yellow perch [9]. The repeated and independent recruitment of TGF- β receptors, including *Amhr2*, in teleost fish sex determination is even more puzzling as many of these MSD genes, encoding a TGF β receptor, share a similar N-terminal truncation [9, 18, 21], supporting their evolution towards a ligand-independent mechanism of action [18]. Therefore, the extent of evolutionary conservation of the Y-linked *amhr2bY* gene found in yellow perch in closely related species (genus *Perca* and *Sander*) is an important question with implications for better sex control in these aquaculture species and for understanding the evolution of sex linkage and protein structure.

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

Results

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

The genome of *P. fluviatilis* was sequenced to high coverage using Oxford Nanopore long-read sequencing (estimated coverage, 67-fold/N50 read length, 11.9 kbp), and Hi-C data was generated to allow for chromosome-level assembly (coverage, 52-fold/alignable pairs, 89.1%/Hi-C map, see Additional file 1: Fig. S1). The final assembly yielded a highly complete reference genome (99.0% of sequence assigned to 24 chromosomes (N50 length, 39.6 Mbp) and highly continuous contigs (N50 length: 4.1 Mbp)). Compared to a previously published genome assembly of *P. fluviatilis*, obtained from “linked-short-reads” (10X Genomics), these numbers represent a 316-fold improvement of contig continuity and a 6.3-fold increase of scaffold continuity. The better continuity resulted in an increased percentage of predictable genes (BUSCO results below). Genome assembly statistics are also highly congruent with the previously published reference quality *P. flavescens* genome, except for obvious

size differences as the *P. fluviatilis* assembly is about 8.1% larger than the *P. flavescens* assembly and represents the largest genome known in the genus *Perca* (Table 1).

The genome of *P. schrenkii* was assembled by a hybrid assembly method, which was highly efficient regarding long-read sequencing coverage and read length needed (here only 30-fold/N50 read length, 4.95 kb). De novo assembled contigs from short reads were combined with long reads and scaffolded using our CSA pipeline [25], with the *P. flavescens* as the closest reference genome (Fig. 1: divergence time about 7.1 Mya) for genome comparison and inferring chromosomal-level sequences. Using this approach, we were able to assemble the genome of *P. schrenkii* to similar quality as those obtained for *P. flavescens* and *P. fluviatilis* (94.7% assigned to 24 chromosomes/contig N50 length, 3.2 Mbp; Table 1). The genome assembly size of *P. schrenkii* was in between the other two *Perca* sp. genomes (877 Mbp < 908 Mbp < 951 Mbp).

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

American *Sander* and *Perca* species tend to have smaller genome sizes than their Eurasian relatives (Table 1).

De novo repeat analysis showed that 60.1% of the genome assembly size difference between *Sander vitreus* and *S. lucioperca* could be explained by repeat expansion/reduction. Similarly, for *Perca flavescens* and *P. fluviatilis* about 64.8% of the genome size differences were due to repeat expansion/reduction. In both species pairs, most repeat expansions/reductions were observed in repeat elements classified as “unknown.” Regarding annotated repeat element classes, L2, DNA, Helitron, CMC-EnSpm, hAT, Rex-Babar, hAT-Charlie, and Piggy-Bac elements expanded the most in both Eurasian species (together adding roughly 20 Mbp to the genomes of *S. lucioperca* and *P. fluviatilis*), while a clear expansion of only a single repeat element, called RTE-BovB, was found in both North American species (adding about 3 and 7 Mbp of sequence to *S. vitreus* and *P. flavescens*, respectively; Additional file 2: Table S1).

The genome of *P. fluviatilis* was annotated by NCBI/GNOMON, which included ample public RNAseq data and protein homology evidence. For *P. schrenkii* and *S. vitreus*, we transferred the NCBI/GNOMON annotations of *P. flavescens* and *S. lucioperca*, respectively. BUSCO analysis (Table 2) revealed values larger than 95.9% for complete BUSCOs (category “C:”) for all annotations. The comparative annotation approach resulted only in small losses (category “M:”) of a few BUSCO genes in the range of 0.4–1.1%. In this regard, the *S. vitreus* assembly performed better than the *P. schrenkii* assembly, possibly due to the higher N50 read length of the underlying long-read data.

Percomorpha phylogenomics and divergence time estimation

To calculate the phylogenetic tree of 36 Percomorpha species and their divergence times, we aligned whole

Table 1 Genome assembly statistics for *P. fluviatilis* and *P. schrenkii*. For comparison, the table also provides numbers for two recently published *Perca* sp. assemblies [9, 24]

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

Abbreviations: LR long reads, HiC chromosome conformation capture, CG comparative genomics, 10X linked-read sequencing, GLM genetic linkage map

genomes and extracted the non-coding alignments (Fig. 1). The use of non-coding sequences is preferable to calculate difficult-to-resolve phylogenetic trees that occur after massive radiations [26–28]. Our multiple alignment consisted of 6,594,104 nt residues (2,256,299 distinct patterns; 1,652,510 parsimony-informative;

1,136,496 singleton sites; 3,805,098 constant sites) and resulted in a highly supported tree (raxml-ng and IQtree2 topologies were identical; all IQtree2’s SH-aLRT and ultrafast bootstrapping (UFBS) values were 100). According to the divergence time analysis, most Percomorpha orders emerged after

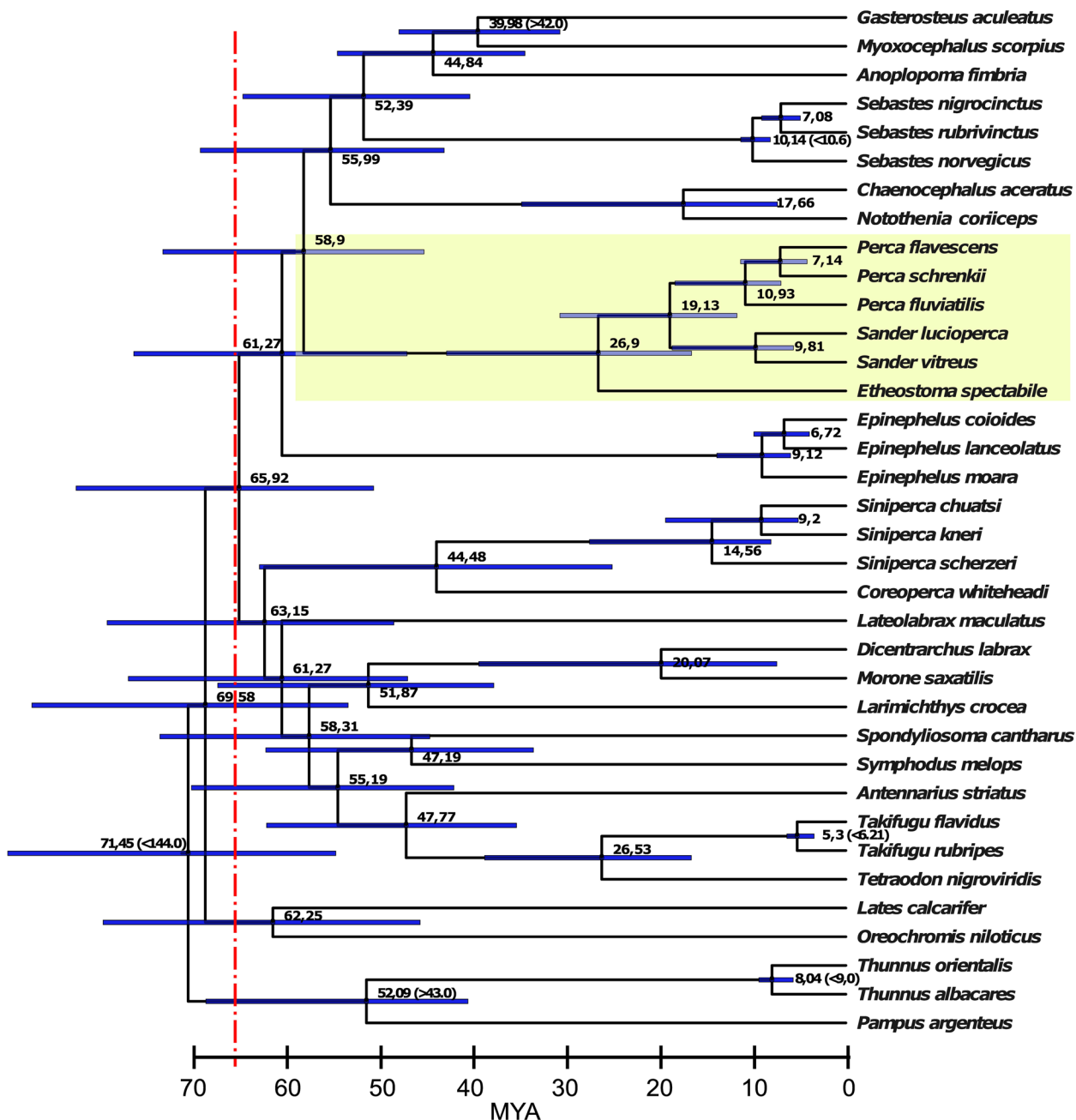


Fig. 1 Time calibrated phylogenomic tree constructed from non-coding alignments of 36 *Percomorpha* genome assemblies reveals a massive radiation after the Cretaceous–Paleogene (K–Pg) boundary (66 Mya, dotted red line). The family Percidae is highlighted in yellow. Node numbers depict the median ages in Mya calculated by Mctree (values in brackets were taken from www.timetree.org and used for calibration). Blue bars depict the 95% confidence intervals for the node ages. All branches of the tree obtained 100% support using the SH-aLRT (Shimodaira–Hasegawa-like approximate likelihood ratio test) and UFBS (ultra-fast bootstrap) tests

Table 2 BUSCO scoring of annotations of five Percidae genome assemblies (*P. fluviatilis*, *P. schrenkii*, *S. vitreus* from this study). The comparative mapping of high-quality NCBI/GNOMON annotations onto closely related species' genome assemblies is a cost-effective and fast procedure to annotate new genomes

	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

Abbreviations: C complete, S single copy, D duplicated, F fragmented, M missing

the Cretaceous–Paleogene (K-Pg) boundary about 65.9 Million years ago (Mya) (CI, 51.3–83.6). The lineage leading to the Percidae (represented with species of *Perca*, *Sander*, and *Etheostoma*) emerged about 58.9 Mya (CI, 45.8–74.2), and the extant Percidae species analyzed in this study diverged from the last common ancestor (LCA) about 26.9 Mya (CI, 16.8–43.4). The *Perca* and *Sander* genera split about 19.1 Mya (CI, 11.8–31.1), and *S. vitreus* and *S. lucioperca* splitted at 9.8 Mya (CI, 5.7–18.9) similar to the divergence of *P. fluviatilis* from *P. flavescens* and *P. schrenkii* 10.9 Mya (CI, 7.1–18.6). The closest *Perca* species are *P. flavescens* and *P. schrenkii* which diverged about 7.1 Mya (CI, 4.2–11.4), although today both species are occurring in completely different global ranges.

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

In the genome of *P. flavescens*, two *amhr2* paralogs were previously described, i.e., an autosomal gene, *amhr2a*, present in both males and females on chromosome 04 (Chr04), and a male-specific duplication on the Y-chromosome (Chr09), *amhr2bY* [9]. A similar *amhr2* gene duplication was also found in our male *P. schrenkii* assembly, and sequence homologies and conserved synteny analyses show that these two *P. schrenkii* *amhr2* genes are orthologs of *P. flavescens amhr2a* and *amhr2bY*, respectively (Fig. 2A and AB). Genotyping of one male and one female also suggests that the *amhr2bY* gene could be male-specific in *P. schrenkii* (Additional file 3: Fig. S2, Table 3), as described in *P.*

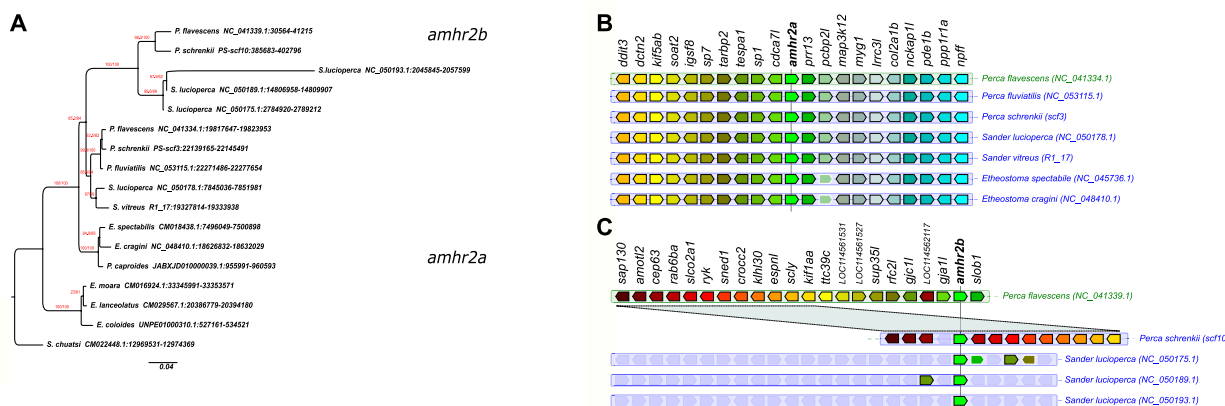


Fig. 2 Evolution of *amhr2* genes in Percidae. **A** Orthologs of *amhr2a* were identified in genome assemblies, and their gene tree is consistent with the species tree, which was found by the phylogenomics approach (see Fig. 1). The *amhr2b* duplicates were only found in the genome assemblies of *P. flavescens* (single copy, male-specific), *P. schrenkii* (single copy, potentially male-specific), and *S. lucioperca* (three copies, no clear sex linkage), and they clustered together. Thus, *amhr2b* stems from a gene duplication event that occurred at the origin of Percidae (19–27 Mya), and the absence of *amhr2b* in *P. fluviatilis* and *S. vitreus* suggests a secondary loss event in these species. This tree was calculated on codon position 1 and 2 alignments and achieved the best bootstrap support (red numbers: SH-aLRT/UFBS support values) for the split of the *amhr2a* and *amhr2b* clades. Trees based on complete CDS, CDS + introns, and amino acid sequences resulted in the same topology albeit with lower bootstrap support for some splits (Additional file 10: Fig. S9). Numbers after species names depict the coordinates of the respective *amhr2* genes in the genome assemblies. **B, C** Conserved synteny around the *amhr2a* (**B**) and *amhr2b* (**C**) loci in some Percidae species. These multiple duplications (in *S. lucioperca*) or the loss of the *amhr2b* genes (in *P. fluviatilis* and *S. vitreus*) emphasize that *amhr2b* may be considered a “jumping” gene locus, which is also supported by conserved synteny analysis

Table 3 Sex linkage of different sex markers in *Perca fluviatilis* and *Perca schrenkii*. Associations between each sex-specific marker and sex phenotypes are provided for both males and females (number of positive individuals/total number of individuals) along with the *p*-value of association with sex that was calculated for each species and population based on Pearson’s chi-square test with Yates’ continuity correction

Species	Population	Marker	Males	Females	<i>p</i> -value
<i>Perca schrenkii</i>	Alakol Lake, Kazakhstan	<i>amhr2bY</i>	1/1	0/1	ns
<i>Perca fluviatilis</i>	Mueggelsee Lake Germany	<i>hssl1</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

ns not statistically significant

flavescens [9]. This potential sex linkage is also supported by a half coverage of reads in the genomic region containing the *amhr2bY* locus in our male *P. schrenkii* genome assembly (Additional file 4: Fig. S3), in agreement with the hemizyosity of a male-specific region on the Y in species with a XX/XY sex-determination system. Alignment of the *P. schrenkii amhr2bY* ortholog shows that its coding sequence (CDS) shares 98% identity with the *P. flavescens amhr2bY* CDS and 95.7% identity at the protein level the *P. flavescens Amhr2bY*. As in *P. flavescens* [9], the *P. schrenkii amhr2bY* gene encodes a N-terminal-truncated type II receptor protein that lacks most of the cysteine-rich extracellular part of the receptor, which is crucially involved in ligand-binding specificity [29] (Fig. 3).

In contrast, in our *P. fluviatilis* male genome assembly, only one copy of *amhr2* could be identified, and sequence

homologies and conserved synteny analysis (Fig. 2A, B) show that this *amhr2* gene is the ortholog of *P. flavescens* and *P. schrenkii amhr2a*. In addition, PCR with primers designed to amplify both *amhr2a* and *amhr2bY* from *P. flavescens* and *P. schrenkii* did not show any sex differences in *P. fluviatilis*. Altogether, these results support the absence of an *amhr2b* gene in *P. fluviatilis*.

In the *Sander lucioperca* male genome assembly, four copies of *amhr2* were detected, and like in *P. flavescens* and *P. schrenkii*, the deduced duplicated *S. lucioperca Amhr2b* proteins have the same structural modifications with a deletion of part of their N-terminal region extracellular domain (see Fig. 3). Using the same primers as those used for the amplification of both *amhr2a* and *amhr2bY* in *P. flavescens* and *P. schrenkii*, PCR genotyping on males and females of *S. lucioperca* produced complex amplification patterns with multiple

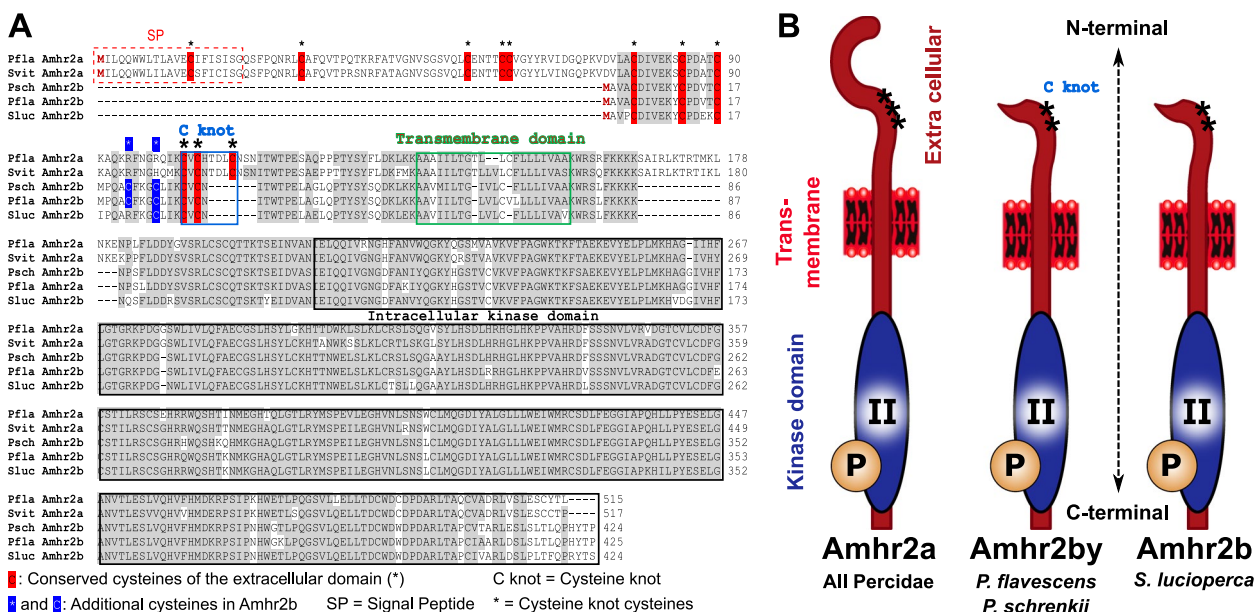


Fig. 3 Alignment and structure of Amhr2a and Amhr2b proteins in *Perca flavescens* (Pfla) and *P. schrenkii* (Psch), *Sander lucioperca* (Sluc), and *S. vitreus* (Svit). **A** Alignment of some Percidae Amhr2a (*P. fluviatilis* and *S. vitreus*) and Amhr2b (*P. schrenkii*, *P. flavescens*, and *S. lucioperca*)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

bands and no visible association with sex. In the publicly available genome assembly of *Sander vitreus*, sequence homologies and/or conserved synteny analysis (Fig. 2A, B) allowed the identification of a single autosomal *amhr2a* gene.

A phylogenetic analysis of the sequences with similarity to *amhr2* in *Perca* and *Sander* (Fig. 2A, B) shows that the different *amhr2* genes likely originated from a gene duplication event that happened in the branch leading to the LCA of these species, dated around 19–27 Mya. This hypothesis of a common origin of this *amhr2* duplication in the last common ancestor of these species is also corroborated by the fact that these *amhr2b* duplications all share the same deletion in their extracellular N-terminal domain (see Fig. 3). The most parsimonious hypothesis is then a common origin of the duplication of the *amhr2b* gene that has been conserved in *P. schrenkii* and *P. flavescens*, lost in *P. fluviatilis* and *S. vitreus*, and amplified in *S. lucioperca*.

Evolution of sex determination in *P. fluviatilis*

Because *P. fluviatilis* sex determination does not rely on an *amhr2* duplication like what has been found in *P. flavescens* [9] and *P. schrenkii* (this study), we used genome-wide approaches to better characterize its sex-determination system. Restriction site-associated DNA sequencing (RAD-Seq) analysis on 35 males and 35 females of *P. fluviatilis*, carried out with a minimum read depth of one, allowed the characterization of a single significant sex marker (Additional file 5: Fig. S4). This 94 bp RAD sequence matches (Blast Identities, 93/95%) a portion of *P. fluviatilis* chromosome 18 (GENO_Pfluv_1.0, Chr18: 27656212–27656305). This RAD-Seq analysis suggests that Chr18 could be the *P. fluviatilis* sex chromosome and that its sex-determining region could be very small because we only detected a single significant sex-linked RAD sequence. To get a better characterization of the *P. fluviatilis* sex chromosome and sex-determining region, we then used pool sequencing (Pool-Seq) to re-analyze the same samples used for RAD-Seq by pooling together DNA from the males in one pool and DNA from females in a second pool. Using these Pool-Seq datasets, we identified a small 100-kb region on *P. fluviatilis* Chr18 with a high density of male-specific SNVs (Fig. 4), confirming the RAD-Seq hypothesis that Chr18 is the sex chromosome in that species. No male-specific duplication/insertion event was found in this sex-determining region on Chr18, which contains six annotated genes (Fig. 4D). These genes encode a protein of unknown function (*C18h1orf198*), three gap-junction proteins (*Cx32.2*, *Gja13.2*, and *Cx32.7*), a protein annotated as inactive hydroxysteroid dehydrogenase-like protein (*Hsd11*), and a protein known as protein broad-minded or Tcb1

domain family member 32 protein (*Tbc1d32*). Three of these six annotated genes, i.e., *c18h1orf198*, *hsd11* and *tbc1d32*, display a higher expression in testis than in the ovary (Additional file 6: Fig. S5).

To provide a better support for the sex linkage of the male-specific variants found within this sex-determining region on Chr18, we designed different types of genotyping assays (classical PCR and KASpar PCR) that have been applied to different *P. fluviatilis* individuals which were phenotypically sexed with confidence. A classical PCR assay was first developed based on the detection of a Y-allele-specific 27-bp deletion in the third intron of the *P. fluviatilis* *hsd11* gene (Additional file 7: Fig. S6, Table 3), and this assay successfully identified all males of a Lake Mueggelsee from Germany (10 males and 9 females; p -value of association with sex = $9.667e-05$). In addition, KASpar allele-specific PCR assays were developed based on 7 single nucleotide sex-specific variants, located at different positions within the Chr18 sex-determining region of *P. fluviatilis* (Fig. 4D). Tests of 48 males and 48 females showed that of seven KASpar allele-specific PCR-assays, 5 resulted in a high proportion of correctly genotyped individuals with males being heterozygote and females being homozygote (> 95%). Two of the targeted SNVs (SNV1 and SNV3) that displayed 100% sex linkage accuracy (Additional file 8: Fig. S7, Tables 3 and 4). Sex linkage of SNV1 was then checked on a wild-type population from Kortowskie Lake in Poland for which the association of male phenotype and SNV1 heterozygosity was also complete (17 males and 20 females; p -value = $8.83e-09$, Table 3).

Discussion

The Percidae family is represented by 239 species and 11 genera. The genera *Perca* and *Sander* are especially important for aquaculture and fisheries. By providing new genome sequence assemblies for *Perca fluviatilis*, *P. schrenckii*, and *Sander vitreus*, we provide for the first-time access to all economically important species of the *Percidae* at the DNA-level. Assembly statistics for these three new genome sequence assemblies are reference grade with N50 continuity in the megabase range and chromosomal length scaffolds, obtained either by Hi-C scaffolding or by conserved synteny analysis involving the closest relatives. Importantly, completeness on the gene-level is significantly higher than in a short read-based draft genome for *P. fluviatilis*, published earlier [30], allowing much stronger conclusions to be drawn regarding the presence or absence of possible sex-determining genes.

The origin of *Percidae*

A phylogenomic approach using aligned non-coding sequences of 36 genomes resulted in a highly supported tree showing a rapid radiation of fish families. It has recently

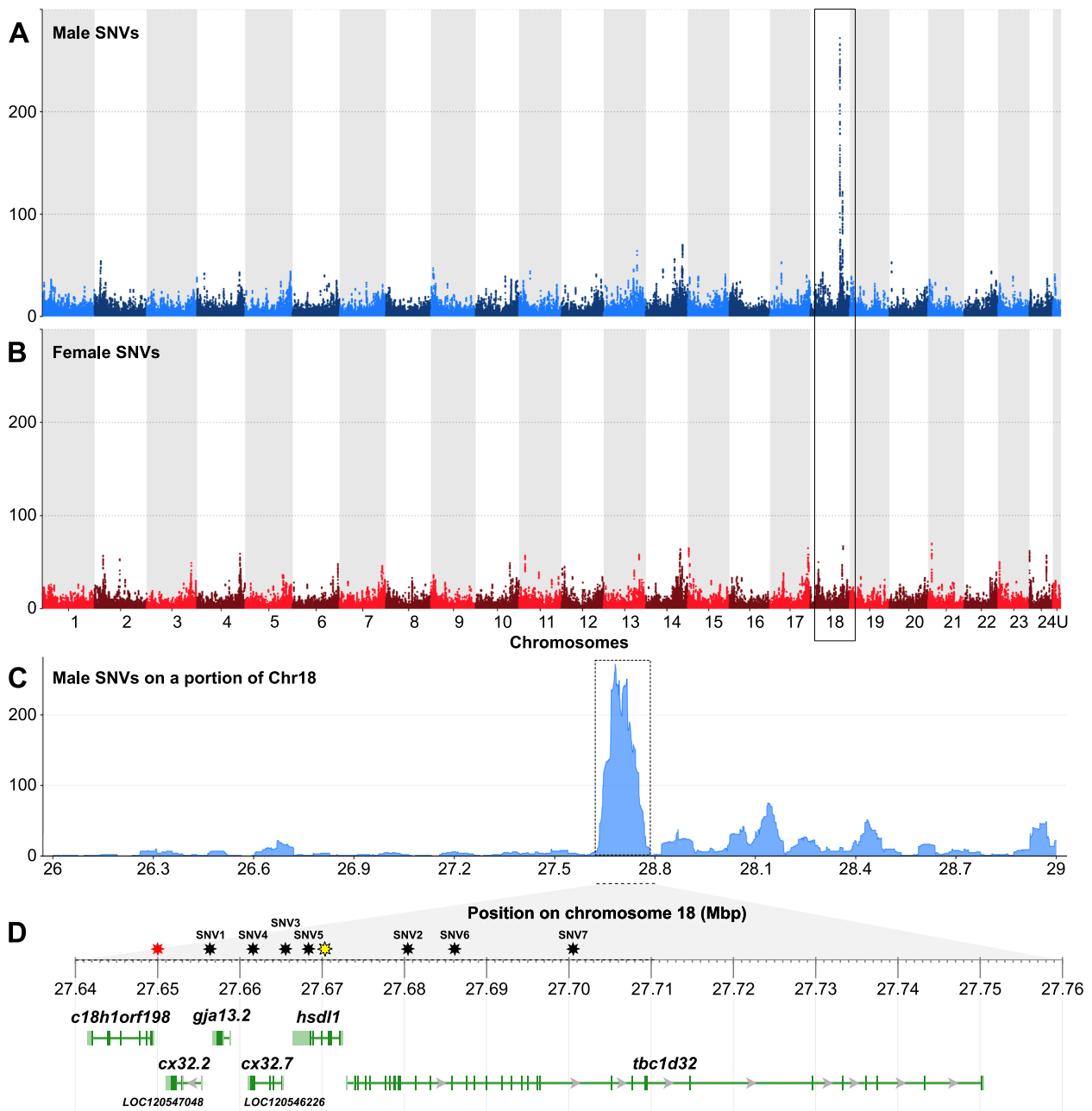


Fig. 4 Chromosome 18 is the sex chromosome of *Perca fluviatilis*. **A, B** Genome-wide Manhattan plots of **A** male- and **B** female-specific single-nucleotide variations (SNVs), showing that chromosome 18 (Chr18) contains a 100-kb region, enriched for male-specific SNVs. Male- and female-specific SNVs are represented as dots (total of SNVs per 50-kb window size) of alternating colors on adjacent chromosomes. **C** Zoomed view of the male-specific SNVs on the sex-biased region of Chr18 with its gene annotation content **(D)**: *c18h1orf198* = *c18h1orf198* homolog; *cx32.2* (LOC120547048) = gap junction Cx32.2 protein-like; *gja13.2* = gap junction protein alpha 13.2; *cx32.7* (LOC120546226) = gap junction Cx32.7 protein-like; *hsd1* = inactive hydroxysteroid dehydrogenase-like protein 1; *tbc1d32* = tcb1 domain family member 32 (also known as protein broad-minded). Stars over the ruler of **D** are the locations of the single male-specific RAD maker (red star) and of the SNVs used for designing KASPar assays (black stars; SNV1–7). The location of the sex-specific intronic indel inside the *hsd1*-gene used for sex-specific PCR is shown by a yellow star

been shown that phylogenomics based on non-coding sequences may be more reliable at resolving difficult-to-resolve radiations in species trees (i.e., in Aves). According to our time calibration, many taxonomic orders related to

the Percidae emerged shortly after the Cretaceous-Paleogene (K-Pg) mass extinction event, about 66 Mya, and gave rise to Percidae about 59 Mya. This is in contrast to many older studies, which have, for example, dated the split

Table 4 KASpar allele-specific PCR assays on seven single sex-specific nucleotide variations (SNV ID#) in *P. fluviatilis*. Numbers of homozygote (Ho), heterozygote (He), and uncalled genotypes (U). The *p*-value of association with sex was calculated for each SNV based on Pearson's chi-square test with Yates' continuity correction scoring heterozygote males and homozygote females as positives

SNV ID no.	N (M/F)	% N	M Ho/He/U	F Ho/He/U	% As	<i>p</i> -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

N total numbers of genotyped individuals (M/F: males/females), %N percentage of genotyped individuals, % As percentage of correctly assigned genotypes, i.e., male heterozygotes and female homozygotes

of *Perca* ssp. and *Gasterosteus* ssp. back to the Cretaceous (73–165 Mya; 18 of 23 studies listed at www.timetree.org). Similar patterns in rapid radiations have been observed in the avian tree of life and have likewise been attributed to the K-Pg mass extinction [27, 28]. In context, it has been argued that the so-called Lilliput Effect, which describes the selection in favor of species with small body sizes and fast generation times after mass extinction events, can lead to an increase in substitution rates and results in overestimations of node-ages for molecular clocks [31].

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

Evolution and turnover of *Amhr2*

Sex-determination systems with MSD evolved from duplications of the *amh* [10–17], or *amhr2* genes have now been characterized in many fish species, all with a male-heterogametic system (XX/XY). In addition, the fact that *Amh* in monotremes [32], or *Amhr2* in some lizards [33], are Y-linked also makes them strong MSD gene candidates in other vertebrate species. In Percidae, sex determination has only been explored in some species of *Perca* [8, 9], and *Amhr2* has been characterized as a potential MSD gene in yellow perch, *P. flavescens* [9]. Our results suggest that this duplication of *amhr2* in *P. flavescens* is also shared by *P. schrenkii* and *S. lucioperca*, implying an origin of duplication in their last common ancestor, dated around 19–27 Mya. However, the fate of this duplication seems to be complex—with multiple duplications/insertions on different chromosomes with no clear sex linkage in *S. lucioperca*, a secondary loss in *P. fluviatilis* and *S. vitreus*, contrasting with a single potentially sex-linked duplication/insertion in *P. schrenkii* and in *P. flavescens*. This finding suggests that the shared ancestral *amhr2b*-duplicated locus (Fig. 2A) might be a jumping locus that has been moving around during its evolution as found for the *sdY* MSD jumping sex locus in salmonids [34, 35]. Additional evidence that these

amhr2b genes originated a single ancestor also rely on the fact that the *Amhr2b* proteins of *P. flavescens* [9], *P. schrenkii* and *S. lucioperca* share a similar gene structure with an N-terminal truncation that results in the absence of the cysteine-rich extracellular part of the receptor. This part of the receptor is known to be crucial for ligand binding [36] and similar N-terminal truncations of duplicated *amhr2* involved in sex determination were also described in catfishes from the Pangasidae family [18] and in the common seadragon [21]. A N-terminal truncation has also been observed in a BMP receptor (*Bmpr1b*) acting as a master sex determining gene in the Atlantic herring, *Clupea harengus*, where this receptor retains the capacity to propagate specific intracellular signals through kinase activity and Smad protein phosphorylation despite its N-terminal truncation [37]. Collectively, these results suggest that certain TGFβ receptors, even when truncated in their N-terminal extracellular ligand-binding domain, can still elicit a biological response independent of ligand activation. The convergent evolution of numerous fish MSD genes encoding a TGFβ receptor with analogous N-terminal truncation suggests the likelihood of this ligand-independent action as a crucial adaptive mechanism. In the genus *Sander*, the situation might be similar to that in *Perca* regarding the changes of the sex-determination systems between species. In *S. lucioperca*, *amhr2b* might still serve as the MSD gene, but the several recent *amhr2b* duplications have complicated our analysis so far. Similar to *P. fluviatilis*, *S. vitreus* has lost *amhr2b* and likely another factor took over as a potential MSD gene.

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

The fact that *P. fluviatilis* sex does not rely on an *amhr2* duplication like *P. flavescens* [9] and *P. schrenkii* does indicates that *P. fluviatilis* evolved a completely different and new MSD gene. Our results also show that this sex locus on *P. fluviatilis* Chr18 is very small compared

to what is observed in many fish species, with a potentially non-recombining size around 100 kb. This locus however is not the smallest SD locus described in teleost fish: in the pufferfish, *Takifugu rubripes*, the sex locus is limited to a few SNPs that differentiate *amhr2* alleles on the X- and Y-chromosomes [19]. Because we did not find any sign of a sex chromosome-specific duplication/insertion event in the *P. fluviatilis* SDR, this sex locus seems to result from pure allelic diversification and is thus in contrast to *P. flavescens* [9] and probably also *P. schrenkii* (this study). The *P. fluviatilis* sex-specific region on Chr18 contains 6 annotated genes, which encode a protein of unknown function (*C18h1orf198*), 3 gap junction proteins (*Cx32.2*, *Gja13.2*, and *Cx32.7*), a protein annotated as inactive hydroxysteroid dehydrogenase-like protein (*Hsd11*), and a protein known as protein broad-minded or Tcb1 domain family member 32 protein (*Tbc1d32*). Of these genes, *hsd11* and *tbc1d32* are interesting potential MSD candidates in *P. fluviatilis*, based on their potential functions and the fact that both display a higher testicular expression compared to the ovary. The *Hsd11* protein is indeed annotated as “inactive” [38], but this annotation only refers to its lack of enzymatic activity against substrates so far tested, leaving other potential functional roles for a protein that is highly conserved in vertebrates [38]. In *Epinephelus coioides*, *hsd11* has been shown to be differentially expressed during female-to-male sex reversal, and its expression profile clustered with *hsd17b1* [39], which plays a central role in converting sex steroids and has recently been identified as a potential MSD gene in different species with a female heterogametic (ZZ/ZW) sex-determination system, like in oyster pompano, *Trachinotus anak* [40] and different amberjack species [41, 42]. The *Tbc1d32* protein has been shown to be required for a high Sonic hedgehog (*Shh*) signaling in the mouse neural tube [43]. Given the role of *Shh* signaling downstream of steroidogenic factor 1 (*nr5a1*) for the proper steroidogenic lineage fate [44] and the importance of steroids in gonadal sex differentiation [45, 46], *tbc1d32* would be also an interesting candidate as a potential MSD gene in *Perca fluviatilis*.

Conclusions

Our study shows that *Percidae* exhibit a remarkable high variation in sex determination mechanisms. This variation is connected to deletion or amplification of *amhr2bY*, which if lost in certain species (like *Perca fluviatilis* or *Sander vitreus*) should cause rewiring of the sex determining pathways and result in the rise of new sex-determination systems. The mechanisms behind a “jumping” *amhr2bY* expansion or loss and which genes

replace it as the MSD remain to be elucidated. The new *Percidae* reference sequence assemblies presented here and the highly reliable sex markers developed for *P. fluviatilis* can now be applied for sex genotyping in basic science as well as in aquaculture.

Methods

Biological samples

In *Perca fluviatilis*, high molecular weight (HMW) genomic DNA (gDNA) for genome sequencing was extracted from a blood sample of a male called “Pf_M1” (BioSample ID SAMN12071746 [47]) from the aquaculture facility of the University de Lorraine, Nancy, France. Blood (0.5 ml) was sampled and directly stored in 25 ml of a TNES-Urea lysis buffer (TNES-Urea, 4 M urea; 10 mM Tris-HCl, pH 7.5; 125 mM NaCl; 10 mM EDTA; 1% SDS). HMW gDNA was extracted from the TNES-urea buffer using a slightly modified phenol/chloroform protocol as described [12]. For the chromosome contact map (Hi-C), 1.5 ml of blood was taken from the same animal and slowly (1 K/min) cryopreserved with 15% dimethyl sulfoxide (DMSO) in a Mr. Frosty Freezing Container (Thermo Fisher) at -80°C . Additional fin clip samples for RAD sequencing (RAD-Seq), pool sequencing (Pool-Seq), or sex-genotyping assays were collected and stored in 90% ethanol, either at the Lucas Perche aquaculture facility (Le Moulin de Cany, 57170 Hampont, France), at Kortowskie Lake in Poland, or at Mueggelsee Lake in Germany.

Samples of *Perca schrenkii* were obtained for genome sequencing and sex genotyping from male and female wild catches at lake Alakol, Kazakhstan (46.328 N, 81.374 E, Additional file 9: Fig. S8). Different organs and tissues (brain, liver, muscle, ovary, testis) were sampled for genome and transcriptome sequencing (BioSample ID SAMN15143703, [48]) and stored in RNAlater. HMW gDNA for genome sequencing was extracted from brain tissue of the male *P. schrenkii* individual, using the MagAttract HMW DNA Kit (Qiagen, Germany). Total RNA for transcriptome sequencing was isolated using a standard TRIzol protocol, in combination with the RNeasy Mini Kit (Qiagen, Germany).

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

Sequencing

Genomic sequencing of *P. fluviatilis* was carried out using a combination of 2×250 bp Illumina short-reads, Oxford Nanopore long reads and a chromosome contact map (Hi-C). For long-read sequencing, DNA was sheared to 20 kb using the megaruptor system (Diagenode). ONT (Oxford nanopore technologies) library preparation and sequencing were performed using 5 μ g of sheared DNA and ligation sequencing kits SQK-LSK108 or SQK-LSK109, according to the manufacturer's instructions. The libraries were loaded at a concentration of 0.005 to 0.1 pmol and sequenced for 48 h on 11 GridION R9.4 or R9.4.1 flowcells. Short read wgs (whole-genome shotgun) sequencing for consensus polishing of noisy long read assemblies was carried out by shearing the HMW DNA to approximately 500-bp fragments and using the Illumina Truseq X kit, according to the manufacturer's instructions. The library was sequenced using a read length of 250 bp in paired-end mode (HiSeq 3000, Illumina, California, USA). Hi-C library generation for chromosome assembly was carried out according to a protocol adapted from Foissac et al. [49]. The blood sample was spun down, and the cell pellet was resuspended and fixed in 1% formaldehyde. Five million cells were processed for the Hi-C library. After overnight digestion with *HindIII* (NEB), DNA ends were labeled with Biotin-14-DCTP (Invitrogen), using Klenow fragment (NEB) and re-ligated. A total of 1.4 μ g of DNA was sheared to an average size of 550 bp (Covaris). Biotinylated DNA fragments were pulled down using M280 Streptavidin Dynabeads (Invitrogen) and ligated to PE adaptors (Illumina). The Hi-C library was amplified using PE primers (Illumina) with 10 PCR amplification cycles. The library was sequenced using a HiSeq3000 (Illumina, California, USA), generating 150 bp paired-end reads.

Genomic sequencing of *P. schrenkii* and *S. vitreus* was carried out using Oxford Nanopore long reads on a MinION nanopore sequencer (Oxford Nanopore Technologies, UK) in combination with the MinIT system. Several libraries were constructed using the tagmentation-based SQK-RAD004 kit with varying amounts of input DNA (0.4 to 1.2 μ g) from a male individual or using the ligation approach of the SQK-LSK109 kit (input DNA 2 μ g). Libraries were sequenced on R9.4.1 flowcells with variable run times and exonuclease washes by the EXP-WSH003 kit to remove pore blocks and improve the data yield. Short-read wgs sequencing of *P. schrenkii* was conducted at BGI (BGI Genomics Co., Ltd.). A *P. schrenkii* male and a female wgs library (300 bp fragment length) were constructed and paired end reads of 150 bp length were generated on an Illumina HiSeq4000 system. Public short-read wgs data of *S. vitreus* were obtained from the NCBI Sequence Read Archive (SRA) using the accession

SRR9711286 [50]. Transcriptome sequencing of six *P. schrenkii* samples (female brain, male brain, male muscle, female liver, ovary, and testis) was conducted at BGI. Transcriptome sequencing libraries were constructed from total ribonucleic acid (RNA), applying enrichment of mRNA with oligo(dT) hybridization, mRNA fragmentation, random hexamer cDNA synthesis, size selection, and PCR amplification. Sequencing of 150-bp paired-end reads was performed by an Illumina HiSeq X Ten system.

Genome assembly of *Perca fluviatilis*

Residual adaptor sequences in ONT GridION long reads were trimmed and split by Porechop (v0.2.1) [51]. Reads longer than 9999 bp were assembled by SmartDeNovo (May 2017) [52] using default parameters. Long reads were remapped to the SmartDeNovo contigs by Minimap2 (v2.7) [53] and Racon (v1.3.12) [54] was used to polish the consensus sequence. In a second round of polishing, Illumina short-reads were mapped by BWA mem (v0.7.12-r1039) [55] to the contigs, which were subsequently polished by Pilon (v1.223) [56]. The chromosome-scale assembly was performed by mapping Hi-C data to the assembled contigs, using the Juicer pipeline (v1.5.6) [57] and subsequent scaffolding by 3D-DNA (v180114) [58]. Juicebox (v1.8.8) [59] was used to manually review and curate the chromosome-level scaffolds. A final gap-closing step, applying long reads and LR_gapcloser (v1.1, default parameters) [60], further increased contig length. After gap-closing, a final consensus sequence polishing step was performed by mapping short reads to the scaffolds, sequence variants (1/1 genotypes were considered as corrected errors) were detected with FreeBayes (v0.9.7) [61] and written to a vcf-file. The final fasta file was then generated by vcf-consensus from Vcftools (v0.1.15, default parameters).

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

Illumina short reads were trimmed using Trimmomatic (v0.35) [62]. Short reads were assembled using a custom compiled high kmer version of idba-ud (v1.1.1) [63] with kmer size up to 252. The resulting contigs were mapped against available Percidae genomes (*P. flavescens*, *P. fluviatilis*, and *S. lucioperca*) by Minimap2 [53] and analysis of overall mapped sequence length resulted in *P. schrenkii* aligned best with *P. flavescens* and *S. vitreus* aligned best with *S. lucioperca*. According to the benchmarks published in [25], the publicly available chromosome-level assembly of *P. flavescens* (RefSeq: GCF_004354835.1, [64]) could be used to aid the chromosome assembly of *P. schrenkii* as follows: ONT MinION long reads (male sample) were trimmed and split using Porechop (v0.2.1) [51]. The inhouse developed CSA method (v2.6) [25] was used to assemble the *P. schrenkii* genome from long-read

data and short-read contigs and to infer chromosomal scaffolds using the *P. flavescens* reference genome. CSA parameters were optimized to account for relatively low long-read sequencing coverage and hybrid assembly of long reads and short-read contigs:

```
CSA2.6.pl -r longreads.fa.gz -g P.flavescens.fa -k 19 -s 2 -e 2 -l „-i shortreadcontigs.fa -L3000 -A”
```

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

```
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
```

The assemblies were manually curated, and the consensus sequences were polished using long reads and flye (v2.6) [65], with options: `--nanoraw --polish-target`, followed by two rounds of polishing by Pilon (v1.23) [56], using the short-read data, which had been mapped by Minimap2 (v2.17-r941) [53], to the genome assemblies.

Genome annotation

De novo repeat annotation was performed using Repeat-Modeler (version open-1.0.8) and Repeat Masker (version open-4.0.7). The *P. fluviatilis* genome has been assigned to the RefSeq assembly section of NCBI and has been annotated by GNOMON (www.ncbi.nlm.nih.gov/genome/annotation_euk/process), which included evidence from Actinopterygii proteins ($n = 154,659$) and *P. fluviatilis* RNAseq reads ($n = 3,537,868,978$) (www.ncbi.nlm.nih.gov/genome/annotation_euk/Perca_fluviatilis/100). To annotate our *P. schrenkii* and *S. vitreus* assemblies, we used the high-quality GNOMON annotations from their closest relatives *P. flavescens* (www.ncbi.nlm.nih.gov/genome/annotation_euk/Perca_flavescens/100) and *S. lucioperca* (www.ncbi.nlm.nih.gov/genome/annotation_euk/Sander_lucioperca/101), respectively. We performed high-throughput comparative protein coding gene annotation by spliced alignment of GNOMON mRNAs and proteins by Spaln (v2.06f [66]) to our assemblies and combined the resulting CDS and UTR matches into complete gene models by custom scripts. All annotations were benchmarked using BUSCO [24] with the Actinopterygii_odb9 database and obtained highly similar values as the reference annotations used for the comparative annotation approach.

Genome browsers and data availability

We provide UCSC genome browsers [67] for the five available *Perca* and *Sander* reference genomes (this study: *P. fluviatilis*, *P. schrenkii*, *S. vitreus*; earlier studies: *P. flavescens* [9] and *S. lucioperca* [22] at <http://genomes.igb-berlin.de/Percidae/>). These genome browsers provide

access to genomic sequences and annotations (either public NCBI GNOMON annotations or annotations resulting from our comparative approach). Blat [68] servers for each genome are available to align nucleotide or protein sequences.

Phylogenomics and divergence time estimation

We performed pair-wise whole-genome alignments of 36 teleost genome assemblies as in [69], using Last-aligner and Last-split [34] for filtering 1-to-1 genome matches and Multiz [35] for multiple alignment construction from pairwise alignments. We removed all annotated coding sequence (CDS) from that multiple alignment to retain non-coding alignments using the tool named mafsInRegion (<https://github.com/ENCODE-DCC/kentUtils>) and a bed file with CDS coordinates. Subsequently, we calculated the species tree using iqtree2 and raxml-ng [70, 71]. We added the genomes of *P. schrenkii*, *S. vitreus*, and *Etheostoma spectabile* (GCF_008692095.1 [72]) to this dataset and re-analyzed the highly supported subclade containing Percidae species using several outgroups (*Lates*, *Oreochromis*, *Pampus*, and *Thunnus* sp.). We estimated the divergence times using a large subset of our multiple alignment (10^6 nt residues) and the approximate method of Mctree (Paml package version, [73]). We calibrated 5 nodes of the tree by left or right CI values, obtained from www.timetree.org and applied independent rates or correlated rates clock models and the HKY85 evolutionary model. Approximately 10^8 samples were calculated, of which we used the top 50% for divergence time estimation. Each calculation was performed in two replicates, which were checked for convergence using linear regression. The final tree was plotted using FigTree (v1.4.4, <http://tree.bio.ed.ac.uk/software/figtree>).

Perca fluviatilis RAD sequencing

Perca fluviatilis gDNA samples from 35 males and 35 females were extracted with the NucleoSpin Kit for Tissue (Macherey-Nagel, Duren, Germany), following the manufacturer's instructions. Then, gDNA concentrations were quantified with a Qubit3 fluorometer (Invitrogen, Carlsbad, CA) using a Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA). RAD libraries were constructed from each individual's gDNA, using a previously described protocol with the single *Sbf*I restriction enzyme [74]. These libraries were sequenced on an Illumina HiSeq 2500. Raw reads were demultiplexed using the `process_radtags.pl` wrapper script of stacks, version 1.44, with default settings [75], and further analyzed with the RADSex analysis pipeline [76] to identify sex-specific markers.

Perca fluviatilis pool sequencing

Sequencing of pooled samples (Pool-Seq) was carried out in *Perca fluviatilis* to increase the resolution of RAD sequencing for the identification of sex-specific signatures characteristic of its sex-determining region. The gDNA samples used for RAD sequencing were pooled in equimolar quantities according to their sex. Pooled male and pooled female libraries were constructed using a Truseq nano kit (Illumina, ref. FC-121-4001) following the manufacturer's instructions. Each library was sequenced in an Illumina HiSeq2500 with 2× 250 reads. Pool-Seq reads were analyzed as previously described [9, 11, 77, 78] with the PSASS pipeline (psass version 2.0.0: https://zenodo.org/record/2615936#.XtyIS3s6_AI) that computes the position and density of SNVs, heterozygous

in one sex but homozygous in the other sex (sex-specific SNVs), and the read depths for the male and female pools along the genome to look for sex coverage differences. Psass was run with default parameters except –window-size, which was set to 5,000, and –output-resolution, which was set to 1000.

PCR-based sex diagnostics

A *Perca schrenkii* PCR-based sex diagnostic test was designed based on multiple alignments of the different *amhr2* genes in *P. fluviatilis* (one autosomal gene only), *P. flavescens* (two genes), and *Perca schrenkii* (two genes) to target a conserved region for all *Perca amhr2* genes, allowing the design of PCR-primers that amplify both the autosomal *amhr2a* and the male-specific *amhr2bY* with

Table 5 KASpar allele-specific PCR primers. For each allele (AL1 and AL2) primers, the X and Y sex chromosome-specific alleles are provided along with their sequences

Primer ID no.	Allele (X or Y)	Sequence (5'–3')
SNV1_AL1	X	GAAGGTGACCAAGTTCATGCTGACACATATTGTCCA TCTGATGATAATG
SNV1_AL2	Y	GAAGGTCCGAGTCAACGGATTATGACACATATTGTC CATCTGATGATAAT
SNV1_C	Common	CACCACCACTGACTGAAGAATAATATGAA
SNV2_AL1	X	GAAGGTGACCAAGTTCATGCTGGACTGATTGTGCTG CTTCTCTC
SNV2_AL2	Y	GAAGGTCCGAGTCAACGGATTGGACTGATTGTG CTGCTTCTCT
SNV2_C	Common	CAGATGAGGAAGGAGGAGATGCAT
SNV3_AL1	X	GAAGGTGACCAAGTTCATGCTTACCACCATAG AACCACC
SNV3_AL2	Y	GAAGGTCCGAGTCAACGGATTGCTTACCACCATA GAACCACT
SNV3_C	Common	GGGATGAGATGCCATCTTCCAAATAATA
SNV4_AL1	Y	GAAGGTGACCAAGTTCATGCTCGCCCTCAGCCT GGTTGAT
SNV4_AL2	X	GAAGGTCCGAGTCAACGGATTGCTTACCACCATAG GGTTGAT
SNV4_C	Common	TCGTCATGCACTCCTTACAGCTTT
SNV5_AL1	X	GAAGGTGACCAAGTTCATGCTGGAATTTGCCTGAAA TAATGAATGAATATG
SNV5_AL2	Y	GAAGGTCCGAGTCAACGGATTGTTGGAATTTGCC TGAAATAATGAATGAATATA
SNV5_C	Common	AGGACATTACAGATTGGTCAGACCATATA
SNV6_AL1	X	GAAGGTGACCAAGTTCATGCTTACCCTTCTACCAC CTGTTT
SNV6_AL2	Y	GAAGGTCCGAGTCAACGGATTCTTACCCTTCTACC ACCTGTTG
SNV6_C	Common	CATAGTTCTTACCCTTACTGTACCAGATA
SNV7_AL1	Y	GAAGGTGACCAAGTTCATGCTCAGTGAACCTCCCTA TGAGGCA
SNV7_AL2	X	GAAGGTCCGAGTCAACGGATTAGTGAACCTCCCTAT GAGGCG
SNV7_C	Common	CTCGGTACAAGGTTGAAAGATGAAAGATA

different and specific PCR-amplicon sizes. Selected PCR primer sequences were forward: 5'-AGTTTATTGTGT TAGTTTGGGCT-3' and reverse: 5'-CAAATAAAT CAGAGCAGCGCATC-3'. PCRs were carried out with 1U Platinum Taq DNA Polymerase and its corresponding Buffer (Thermo Fisher) supplemented with 0.8 mM dNTPs (0.2mM each), 1.5 mM MgCl₂, and 0.2 μM of each primer with the following cycling conditions, 96 °C for 3 min; 40 cycles of denaturation (96 °C, 30 s), annealing (54 °C, 30 s), and extension (72 °C, 1 min); final extension (72 °C, 5 min); storage at 4 °C. PCR amplicons were separated on 1.5% agarose gels (1.5% std. agarose, 1× TBE buffer, 5 V/cm, running time 40 min) and the systematic amplification of the autosomal (*amhr2a*) amplicon was used as a positive PCR control.

Perca fluviatilis primers were designed to amplify a 27-bp deletion variant in the third intron of the *P. fluviatilis* *hsd11* gene, which was identified as a male specific (Y-specific) variation based on the pool-seq analysis. Selected PCR primer sequences were forward 5'-ACACTGATC AACATTTTCTGTCTCA-3' and reverse 5'-TGTTAAC ATTTGAGAATTTTGCCTT-3'. PCRs were carried out as described above with the following cycling conditions: denaturation 96 °C for 3 min; 40 cycles of denaturation (96 °C, 30 s), annealing (60 °C, 30 s), and extension (72 °C, 30 min); final extension (72 °C, 5 min); storage at 4 °C. PCR amplicons were separated on 5% agarose gels (5% Biozym sieve 3:1 agarose, 1× TBE buffer, 5 V/cm, 1 h 40 min running time), and the amplicon derived from the amplification of the X-chromosome allele was used as a positive PCR control. In addition to this classical PCR sex-genotyping method, we also explored the sex linkage of some sex-specific SNVs in *P. fluviatilis* using Kompetitive Allele-Specific Polymerase chain reaction (KASPar) assays [79]. Seven sex-specific SNVs were selected at different locations within the *P. fluviatilis* sex-determining region. Primers (Table 5) were designed using the design service available on the 3CR Bioscience website (www.3crbio.com/free-assay-design). KASPar genotyping assays were carried out with a single end-point measure on a Q-PCR Light Cycler 480 (Roche) using the Agencourt[®] DNAdvance kit (Beckman), following the manufacturer's instructions.

RNAseq expression analyses

Already available gonadal datasets of *P. fluviatilis* (age 9 month) RNAseq [80] [SRA accessions: SRR14461526 [81] and SRR14461527 [82]] were used to compare gene expression between ovary and testes for the gene models annotated in the SD region. Reads were mapped to our *P. fluviatilis* reference genome using HISAT2 [83] and transcript assembly and fragments per kilobase million (FPKM) values were calculated using STRINGTIE [84].

Abbreviations

CDS	Coding sequence
Chr	Chromosome
DNA	Deoxyribonucleic acid
FPKM	Fragments per kilobase million
gDNA	Genomic deoxyribonucleic acid
HMW	High molecular weight
KASPar	Kompetitive allele specific
LCA	Last common ancestor
MSD	Master sex determining
Mya	Million years ago
PCR	Polymerase chain reaction
Pool-Seq	Pool sequencing
RAD-Seq	Restriction site-associated DNA sequencing
RAD	Restriction site-associated DNA
RNA	Ribonucleic acid
SDR	Sex-determining region
SNV	Single nucleotide variation
SRA	Sequence Read Archive
TGF-β	Transforming growth factor beta

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01935-9>.

Additional file 1: Fig. S1. Hi-C map of *P. fluviatilis* genome

Additional file 2: Table S1. Annotated repeats in *Perca sp.* and *Sander sp.* genomes (RepeatModeler *de novo* analysis). Repeat elements mentioned in the manuscript have grey highlighting. The class "DNA" is assigned to repeat elements that harbor signals of transposases, but miss further signals to classify them with more detail.

Additional file 3: Fig. S2. A sex-specific 637 bp PCR product amplifies in *P. schrenkii* male (sample id = 8), while it is absent in female (sample id = 17). The corresponding primer pair also works for sexing of *P. flavescens*.

Additional file 4: Fig. S3. Sex-specific sequencing coverage of *amhr2b* locus in *Perca schrenkii* in region PS-scf10/CM046795.1: 366,025–418,970. A female and a male genome were sequenced to approximately 40x coverage using short-read sequencing. After filtering for unique mapping reads (mapping quality 60), a clear coverage difference between females and males is visible. The ~53 kbp region has virtually no coverage in females. In contrast males exhibit haploid coverage (about 20x), which is in accordance with a X/Y SD system.

Additional file 5: Fig. S4. A single RAD sex-specific marker is significantly associated with male sex in *P. fluviatilis*. Tile plot of the distribution of RAD-Sex markers between *Perca fluviatilis* males (horizontal axis) and females (vertical axis) with a minimum read depth of 1 (d = 1). Color intensity (see color legend on the right) indicates the number of markers present for each of the corresponding number of males and females. A single significant marker at the lower right of the grid was present in all 35 males and absent from all 34 females and is boxed with a red border (Chi-squared test, $p < .05$ after Bonferroni correction).

Additional file 6: Fig. S5. Expression of *hsd11* and neighboring genes in public RNAseq datasets (testis: SRR14461526 [81], ovary: SRR14461527 [82]; age of both sampled individuals 9 month). Here *hsd11* expression in testis is 3.25-fold higher than in ovary, for *tbc1d32* and *c18h1orf198* the ratio is 3.83 and 4.41, respectively.

Additional file 7: Fig. S6. Sexing of *P. fluviatilis* using a 27 bp male specific deletion in Intron 3 of the *hsd11* gene (10 males (left) and 9 females (right) from Lake Mueggelsee, Berlin). The simultaneous amplification in both males and females of the X allele without the 27 bp deletion provides an internal control for this PCR. All XY male samples (N = 10) produce two amplicons due to the small size difference of the X and Y amplified alleles, and all XX females (N = 9) produced only the larger X amplicon. This *hsd11* intronic indel variant is located near the variant SNV5 (distance < 1.5 kbp).

Additional file 8: Fig. S7. KASpar allele-specific PCR assays on seven single sex-specific nucleotide variations (SNV ID#) in *P. fluviatilis*. For each Single Nucleotide Variation (SNV), primer AL1 was coupled to FAM fluorescent dye and primer AL2 was coupled to VIC fluorescent dye and the end-point fluorescence of these two fluorescent dyes was respectively on the x and y axes. Male individuals are represented by blue dots and females by red dots. Primers used for analysis can be found in Table 1.

Additional file 9: Fig. S8. Photo-voucher of the male *Perca schrenkii* sampled at Lake Alakol on 11 September 2018; Kazakhstan (46.328 N, 81.374 E). Photo: Matthias Stöck.

Additional file 10: Fig. S9. Additional gene trees for *amhr2*. A) Tree calculated from coding sequence. B) Tree calculated from coding plus intron sequence. C) Tree calculated from amino acid sequence. All trees share the same topology but differ in support values for some splits (SH-aLRT and UFBS tests).

Acknowledgements

We kindly acknowledge the NCBI/Genbank team for providing a GNOMON annotation of the *P. fluviatilis* genome. We thank Dr. Tatjana Dujsebajeva, Salamat Karlybai, and his colleague for the help during field work in Kazakhstan. We thank Eva Kreuz and Wibke Kleiner for the technical assistance. We thank Matt Faust (Ohio, DNR) for the help with samples of *S. vitreus*.

Author's contributions

YG and HK designed the project. PE and WL provided sexed samples of *S. vitreus*. MW, BS, TL, ST, SV, DZ, DC, AB, PH, RM, and TRA collected and sexed the *P. fluviatilis* samples. MS, ST, WK, and SW collected and sexed the tissue samples of *P. schrenkii*. EJ, MW, CL-R, CI, HP, and HK extracted the gDNA, made the genomic libraries, and sequenced them. CC1, CK1, MZ, MW, CK2, RF, AH, HK, and YG processed the genome assemblies and/or analyzed the results. CP, LJ, and CC2 processed and analyzed the sex genotyping tests. HK and YG wrote the manuscript with inputs from all other coauthors. JP, CD, HK, and YG supervised the project administration and raised funding. All the authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was funded by the German Research Foundation (DFG) "Eigene Stelle" grant KU 3596/1-1; project number: 324050651, by the Agence Nationale pour la Recherche (ANR)/DFG PhyloSex project (ANR-13-ISV7-0005), the CRB-Anim "Centre de Ressources Biologiques pour les animaux domestiques" project PERCHSEX, the FEAMP "Fonds européen pour les affaires maritimes et la pêche" project SEX'NPERCH, NIH (R35 GM139635), and NSF (2232891). The GeT-PlaGe and MGX sequencing facilities were supported by France Génomique National infrastructure as part of the "Investissement d'avenir" program managed by (ANR-10-INBS-09).

Availability of data and materials

All genome assemblies and raw sequence datasets have been submitted to NCBI/GENBANK under the BioProject accessions: PRJNA549142 [85], PRJNA637487 [86], and PRJNA808842 [87] (*P. fluviatilis*, *P. schrenkii*, *S. vitreus*, respectively).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

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

²Department of Forestry and Natural Resources | Illinois-Indiana Sea Grant, Purdue University, West Lafayette, USA. ³Sigenae, Plateforme Bioinformatique,

Genotoul, BioinfoMics, UR875 Biométrie et Intelligence Artificielle, INRAE, Castanet-Tolosan, France. ⁴Sigenae, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France. ⁵INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France. ⁶Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland. ⁷Swiss Institute of Bioinformatics, Lausanne, Switzerland. ⁸Montpellier GenomiX (MGX), c/o Institut de Génétique Fonctionnelle, 141 rue de la Cardonille, 34094 Montpellier Cedex 05, France. ⁹GDEC Gentyane, INRAE, Université Clermont Auvergne, Clermont-Ferrand, France. ¹⁰INRAE, LPGP, 35000 Rennes, France. ¹¹State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan Normal University, Changsha, China. ¹²SYSAAF, Station INRAE-LPGP, Campus de Beaulieu, 35042 Rennes Cedex, France. ¹³University of Lorraine, INRAE, L2A, Nancy, France. ¹⁴Department of Gamete and Embryo Biology, Institute of Animal Reproduction and Food Research, Polish Academy of Sciences, ul. Tuwima 10, 10-748, Olsztyn, Poland. ¹⁵National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, 17109 Point Lena Loop Road, Auke Bay Laboratories Juneau, AK 99801, USA. ¹⁶Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA. ¹⁷Sea Biology LLP, Almaty, Kazakhstan.

Received: 19 January 2024 Accepted: 6 June 2024

Published online: 26 June 2024

References

- Stepien CA, Haponski AE. Taxonomy, distribution, and evolution of the Percidae. In: Kestemont P, Dabrowski K, Summerfelt RC, editors. Biology and culture of percid fishes: principles and practices. Dordrecht: Springer, Netherlands; 2015. p. 3–60.
- Hermelink B, Wuertz S, Trubiroha A, Rennert B, Kloas W, Schulz C. Influence of temperature on puberty and maturation of pikeperch. *Sander lucioperca* Gen Comp Endocrinol. 2011;172:282–92.
- Schaefer FJ, Overton JL, Kloas W, Wuertz S. Length rather than year-round spawning, affects reproductive performance of RAS-reared F-generation pikeperch, *Sander lucioperca* (Linnaeus, 1758) - insights from practice. *J Appl Ichthyol*. 2018;34:617–21.
- Policar T, Schaefer FJ, Panana E, Meyer S, Teerlinck S, Toner D, et al. Recent progress in European percid fish culture production technology-tackling bottlenecks (vol 27, pg 4, 2019). *Aquacult Int*. 2019;27:1175–1175.
- Malison JA, Kayes TB, Best CD, Amundson CH, Wentworth BC. Sexual differentiation and use of hormones to control sex in yellow perch (*Perca flavescens*). *Can J Fish Aquat Sci*. 1986;43:26–35.
- Melard C, Kestemont P, Grignard JC. Intensive culture of juvenile and adult Eurasian perch (*P. fluviatilis*): effect of major biotic and abiotic factors on growth. *J Appl Ichthyol*. 1996;12:175–80.
- Rougeot C, Jacobs B, Kestemont P, Melard C. Sex control and sex determination study in Eurasian perch, *Perca fluviatilis*, by use of hormonally sex-reversed male breeders. *Aquaculture*. 2002;211:81–9.
- Rougeot C, Ngingo JV, Gillet L, Vanderplasschen A, Melard C. Gynogenesis induction and sex determination in the Eurasian perch. *Perca fluviatilis* Aquaculture. 2005;243:411–5.
- Feron R, Zahm M, Cabau C, Klopp C, Roques C, Bouchez O, et al. Characterization of a Y-specific duplication/insertion of the anti-Müllerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch. *Perca flavescens* Mol Ecol Resour. 2020;20:531–43.
- Pan Q, Kay T, Depince A, Adolphi M, Schartl M, Guiguen Y, et al. Evolution of master sex determiners: TGF-beta signalling pathways at regulatory crossroads. *Philos Trans R Soc Lond B Biol Sci*. 2021;376:20200091.
- Pan Q, Feron R, Jouanno E, Darras H, Herpin A, Koop B, et al. The rise and fall of the ancient northern pike master sex-determining gene. *Life*. 2021;10:e62858.
- Pan Q, Feron R, Yano A, Guyomard R, Jouanno E, Vigouroux E, et al. Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics*. 2019;15: e1008013.
- Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, et al. A tandem duplicate of anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia. *Oreochromis niloticus* PLOS Genetics. 2015;11: e1005678.

14. Holborn MK, Einfeldt AL, Kess T, Duffy SJ, Messmer AM, Langille BL, et al. Reference genome of lumpfish *Cyclopterus lumpus* Linnaeus provides evidence of male heterogametic sex determination through the AMH pathway. *Mol Ecol Resour*. 2022. <https://doi.org/10.1111/1755-0998.13565>.
15. Song W, Xie Y, Sun M, Li X, Fitzpatrick CK, Vaux F, et al. A duplicated amh is the master sex-determining gene for *Sebastes* rockfish in the Northwest Pacific. *Open Biol*. 2021;11: 210063.
16. Rondeau EB, Laurie CV, Johnson SC, Koop BF. A PCR assay detects a male-specific duplicated copy of anti-Müllerian hormone (amh) in the lingcod (*Ophiodon elongatus*). *BMC Res Notes*. 2016;9:230.
17. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, et al. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci USA*. 2012;109:2955–9.
18. Wen M, Pan Q, Jouanno E, Montfort J, Zahm M, Cabau C, et al. An ancient truncated duplication of the anti-Müllerian hormone receptor type 2 gene is a potential conserved master sex determinant in the Pangasiidae catfish family. *Mol Ecol Resour*. 2022. <https://doi.org/10.1111/1755-0998.13620>.
19. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, et al. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genet*. 2012;8: e1002798.
20. Nakamoto M, Uchino T, Koshimizu E, Kuchiishi Y, Sekiguchi R, Wang L, et al. A Y-linked anti-Müllerian hormone type-II receptor is the sex-determining gene in *ayu*. *Plecoglossus altivelis* *PLoS Genet*. 2021;17: e1009705.
21. Qu M, Liu Y, Zhang Y, Wan S, Ravi V, Qin G, et al. Seadragon genome analysis provides insights into its phenotype and sex determination locus. *Sci Adv*. 2021;7:eabg5196.
22. Nguinkal JA, Brunner RM, Verleih M, Rebl A, de los Rios-Perez L, Schafer N, et al. The first highly contiguous genome assembly of pikeperch (*Sander lucioperca*), an emerging aquaculture species in Europe. *Genes-Basel*. 2019;10(9):708. <https://doi.org/10.3390/genes10090708>.
23. Malmstrom M, Matschiner M, Torresen OK, Jakobsen KS, Jentoft S. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci Data*. 2017;4:160132.
24. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
25. Kuhl H, Li L, Wuertz S, Stock M, Liang XF, Klopp C. CSA: a high-throughput chromosome-scale assembly pipeline for vertebrate genomes. *Gigascience*. 2020;9:giaa034.
26. Braun EL, Kimball RT. Data types and the phylogeny of Neoaves. *Birds*. 2021;2:1–22.
27. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014;346:1320–31.
28. Kuhl H, Frank-Vilches C, Bakker A, Mayr G, Nikolaus G, Boerno ST, et al. An unbiased molecular approach using 3'-UTRs resolves the avian family-level tree of life. *Mol Biol Evol*. 2021;38:108–27.
29. Heldin C-H, Moustakas A. Signaling receptors for TGF- β family members. *Cold Spring Harb Perspect Biol*. 2016;8: a022053.
30. Ozerov MY, Ahmad F, Gross R, Pukk L, Kahar S, Kisand V, et al. Highly continuous genome assembly of Eurasian perch (*Perca fluviatilis*) using linked-read sequencing. *G3: Genes, Genomes, Genetics*. 2018;8:3737–43.
31. Berv JS, Field DJ. Genomic signature of an avian Lilliput Effect across the K-Pg extinction. *Syst Biol*. 2018;67:1–13.
32. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. *Nature*. 2014;508:488–93.
33. Pinto BJ, Gamble T, Smith CH, Wilson MA. A lizard is never late: squamate genomics as a recent catalyst for understanding sex chromosome and microchromosome evolution. *J Hered*. 2023;114:445–58.
34. Frith MC, Kawaguchi R. Split-alignment of genomes finds orthologies more accurately. *Genome Biol*. 2015;16:106.
35. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004;14:708–15.
36. Hart KN, Stocker WA, Nagykyery NG, Walton KL, Harrison CA, Donahoe PK, et al. Structure of AMH bound to AMHR2 provides insight into a unique signaling pair in the TGF- β family. *Proc Natl Acad Sci U S A*. 2021;118: e2104809118.
37. Rafati N, Chen J, Herpin A, Pettersson ME, Han F, Feng C, et al. Reconstruction of the birth of a male sex chromosome present in Atlantic herring. *PNAS*. 2020;117:24359–68.
38. M M, J T, F H, R M, J A. Human and zebrafish hydroxysteroid dehydrogenase like 1 (HSDL1) proteins are inactive enzymes but conserved among species. *Chemico-biological interactions*. 2009;178.
39. Xiao L, Guo Y, Wang D, Zhao M, Hou X, Li S, et al. Beta-hydroxysteroid dehydrogenase genes in orange-spotted grouper (*Epinephelus coioides*): genome-wide identification and expression analysis during sex reversal. *Front Genet*. 2020;11:161.
40. Fan B, Xie D, Li Y, Wang X, Qi X, Li S, et al. A single intronic single nucleotide polymorphism in splicing site of steroidogenic enzyme hsd17b1 is associated with phenotypic sex in oyster pompano. *Trachinotus anak* *Proc Biol Sci*. 2021;288:20212245.
41. Koyama T, Nakamoto M, Morishima K, Yamashita R, Yamashita T, Sasaki K, et al. A SNP in a steroidogenic enzyme is associated with phenotypic sex in *Seriola* fishes. *Current Biology*. 2019;29:1901–1909.e8.
42. Purcell CM, Seetharam AS, Snodgrass O, Ortega-García S, Hyde JR, Severin AJ. Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. *BMC Genomics*. 2018;19:31.
43. Ko HW, Norman RX, Tran J, Fuller KP, Fukuda M, Eggenschwiler JT. Broad-minded links cell cycle-related kinase to cilia assembly and Hedgehog signal transduction. *Developmental Cell*. 2010;18:237–47.
44. Huang C-C, Yao H-H. Diverse functions of Hedgehog signaling in formation and physiology of steroidogenic organs. *Mol Reprod Dev*. 2010;77:489–96.
45. Guiguen Y, Fostier A, Piferrer F, Chang C-F. Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. *Gen Comp Endocrinol*. 2010;165:352–66.
46. Baroiller JF, Guiguen Y. Endocrine and environmental aspects of sex differentiation in gonochoristic fish. *EXS*. 2001;177–201.
47. Schaeerlinger B. Model organism or animal sample from *Perca fluviatilis*. NCBI <https://www.ncbi.nlm.nih.gov/biosample/12071746> (2019).
48. Stock M, Timirkhanov S. Model organism or animal sample from *Perca schrenkii*. NCBI <https://www.ncbi.nlm.nih.gov/biosample/15143703> (2020).
49. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biology*. 2019;17:108.
50. Auburn University. Raw read of the assembled draft genome for walleye. NCBI <https://www.ncbi.nlm.nih.gov/sra/SRR9711286> (2019).
51. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom*. 2017;3: e000132.
52. Liu H, Wu S, Li A, Ruan J. SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte*. 2021;2021:gigabyte15. <https://doi.org/10.46471/gigabyte.15>.
53. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
54. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;27:737–46.
55. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:13033997 [q-bio]*. 2013.
56. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9: e112963.
57. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016;3:95–8.
58. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356:92–5.
59. Robinson JT, Turner D, Durand NC, Thorvaldsdottir H, Mesirov JP, Aiden EL. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst*. 2018;6:256–+.
60. Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, et al. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience*. 2019;8:gij157.
61. Garrison E MG. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:12073907 [q-bioGN]*. 2012.
62. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
63. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.

64. Feron R, Morvezen R, Bestin A, Haffray P, Klopp C, Zahm M, et al. Genome assembly PFLA_1.0. NCBI https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_004354835.1 (2019).
65. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37:540–6.
66. Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* 2012;40:e161.
67. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform.* 2013;14:144–61.
68. Kent WJ. BLAT - the BLAST-like alignment tool. *Genome Res.* 2002;12:656–64.
69. He S, Li L, Lv LY, Cai WJ, Dou YQ, Li J, et al. Mandarin fish (Sinipercaidae) genomes provide insights into innate predatory feeding. *Commun Biol.* 2020;3.
70. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 2019;35:4453–5.
71. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4.
72. Moran RL, Catchen JM, Fuller RC. Genome assembly UIUC_Espe_1.0. NCBI https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_008692095.1/ (2019).
73. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
74. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics.* 2011;188:799–808.
75. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda).* 2011;1:171–82.
76. Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, et al. RADSex: a computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol Ecol Resour.* 2021;21:1715–31.
77. Wen M, Pan Q, Larson W, Eché C, Guiguen Y. Characterization of the sex determining region of channel catfish (*Ictalurus punctatus*) and development of a sex-genotyping test. *Gene.* 2023;850: 146933.
78. Jasonowicz AJ, Simeon A, Zahm M, Cabau C, Klopp C, Roques C, et al. Generation of a chromosome-level genome assembly for Pacific halibut (*Hippoglossus stenolepis*) and characterization of its sex-determining genomic region. *Mol Ecol Resour.* 2022;22:2685–700.
79. Smith SM, Maughan PJ. SNP genotyping using KASPar assays. *Methods Mol Biol.* 2015;1245:243–56.
80. Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics.* 2016;17:368.
81. Bobe J. *Perca fluviatilis* testis Transcriptome. NCBI <https://www.ncbi.nlm.nih.gov/sra/SRX10811604> (2022).
82. Bobe J. *Perca fluviatilis* ovary transcriptome. NCBI <https://www.ncbi.nlm.nih.gov/sra/SRX10811603> (2022).
83. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357–60.
84. Perteza M, Perteza GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33:290–5.
85. Roques C, Zahm M, Klopp C, Bouchez O, Donnadiou C, Kuhl H, et al. *Perca fluviatilis* genome sequencing and assembly. NCBI https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_010015445.1/ (2020).
86. Kuhl H, Stock M. *Perca schrenkii* isolate:PSM8 genome sequencing and assembly. NCBI https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_025617405.1/ (2022).
87. Kuhl H, Euclide PT. *P. Sander vitreus* isolate:19-12246 genome sequencing and assembly. NCBI https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_031162955.1/ (2023).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.