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 C. Klopp, Céline Roques, Olivier Bouchez, Camille Eché, Sophie Valière, Cécile Donnadieu, Pierrick Haffray, et al.

HAL Id: hal-02623895

https://hal.inrae.fr/hal-02623895

Submitted on 26 Jul 2024

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[1,2], **Margot Zahm**[3], **Cédric Cabau**[4], **Christophe Klopp**[3,4], **Céline Roques**[5], **Olivier Bouchez**[5], **Camille Eché**[5], **Sophie Valière**[5], **Cécile Donnadieu**[5], **Pierrick Haffray**[6], **Anastasia Bestin**[6], **Romain Morvezen**[6], **Hervé Acloque**[7], **Peter T. Euclide**[8], **Ming Wen**[1,9], **Elodie Jouano**[1], **Manfred Schartl**[10], **John H. Postlethwait**[11], **Claire Schraidt**[12], **Mark R. Christie**[12,13], **Wesley A. Larson**[14], **Amaury Herpin**[1], **Yann Guiguen**[1,*]

[1]INRAE, UR 1037 Fish Physiology and Genomics, F-35000 Rennes, France. [2]Department of Ecology and Evolution, University of Lausanne, and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [3]Plate-forme bio-informatique Genotoul, Mathématiques et Informatique Appliquées de Toulouse, INRAE, Castanet Tolosan, France. [4]SIGENAE, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France. [5]INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France. [6]SYSAAF, French poultry and aquaculture breeders, 35042, Rennes Cedex, France. [7]GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, Castanet-Tolosan, France. [8]Wisconsin Cooperative Fishery Research Unit, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, USA. [9]State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Science, Hunan Normal University, Changsha, China. [10]Developmental Biochemistry, Biozentrum, University of Würzburg, Würzburg, Germany and The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas, USA. [11]Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA. [12]Department of Forestry and Natural Resources, Purdue University; 715 W. State St., West Lafayette, Indiana 47907-2054 USA. [13]Department of Biological Sciences, Purdue University; 915 W. State St., West Lafayette, Indiana 47907-2054 USA. [14]U.S. Geological Survey Wisconsin Cooperative Fishery Research Unit, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, USA.

[*]corresponding author (yann.guiguen@inra.fr).

Author Manuscript

Abstract

Yellow perch, *Perca flavescens*, is an ecologically and economically important species native to a large portion of the northern United States and southern Canada and is also a promising candidate species for aquaculture. No yellow perch reference genome, however, has been available to facilitate improvements in both fisheries and aquaculture management practices. By combining Oxford Nanopore Technologies long-reads, 10X genomics Illumina short linked reads and a chromosome contact map produced with Hi-C, we generated a high-continuity chromosome scale yellow perch genome assembly of 877.4 Mb. It contains, in agreement with the known diploid chromosome yellow perch count, 24 chromosome-size scaffolds covering 98.8% of the complete assembly (N50 = 37.4 Mb, L50 = 11). We also provide a first characterization of the yellow perch sex determination locus that contains a male-specific duplicate of the anti-Mullerian hormone type II receptor gene (*amhr2by*) inserted at the proximal end of the Y chromosome (chromosome 9). Using this sex-specific information, we developed a simple PCR genotyping assay which accurately differentiates XY genetic males (*amhr2by$^+$*) from XX genetic females (*amhr2by$^-$*). Our high-quality genome assembly is an important genomic resource for future studies on yellow perch ecology, toxicology, fisheries, and aquaculture research. In addition, the characterization of the *amhr2by* gene as a candidate sex determining gene in yellow perch provides a new example of the recurrent implication of the transforming growth factor beta pathway in fish sex determination, and highlights gene duplication as an important genomic mechanism for the emergence of new master sex determination genes.

Keywords

INTRODUCTION

Yellow perch, *Perca flavescens*, is an ecologically and economically important species native to a large portion of the northern United States and southern Canada. In the Laurentian Great Lakes, yellow perch, have bipartite life cycles that include a prolonged dispersive larval stage. Identifying patterns of population connectivity and local adaptation in yellow perch may provide important insights not only for the important yellow perch fishery, but also for a wide variety of Great Lakes fishes with similar life histories (e.g. walleye). Relevant life history characteristics for Great Lakes yellow perch populations include: a pelagic larval duration on the order of 30 to 40 days (Dettmers, Janssen, Pientka, Fulford, & Jude, 2005; Whiteside, Swindoll, & Doolittle, 1985), high fecundity (~ 10,000 to 150,000 eggs/female), high larval and juvenile mortality, and the potential for large population sizes – all characteristics shared with marine fishes (Brazo, Tack, & Liston, 1975; Forney, 1971; Ludsin, DeVanna, & Smith, 2014; Pritt, Roseman, & O'Brien, 2014). Adult yellow perch have modest home ranges and mark-recapture studies have demonstrated that most adult yellow perch and their congener, Eurasian perch (*Perca fluviatilis*), have relatively high site fidelity, particularly with respect to spawning grounds (Bergek & Björklund, 2009; Böhling & Lehtonen, 1984; Glover, Dettmers, Wahl, & Clapp, 2008; Schneeberger, 2000). Female yellow perch lay their eggs in large gelatinous mats, known as skeins, which are often found

entangled in plant material and woody debris or attached to rocky outcrops (Robillard & Marsden, 2001). The duration of egg development in skeins is temperature dependent and is thought to take approximately 15–20 days in Lake Michigan. Embryos subsequently hatch into minuscule larvae that are likely, at least early in their lives, to disperse passively in the currents (Beletsky et al., 2007; Höök, McCormick, Rutherford, Mason, & Carter, 2006). Similar to many marine fishes, yellow perch larvae may have some control over their dispersal trajectories simply by varying their vertical position within the water column (Leis, 2006) As the larvae develop, they may also become better at swimming, such that a combination of active and passive dispersal strategies may ultimately dictate where individual yellow perch are located when they transition to a demersal life stage.

Yellow perch also support recreational and commercial fisheries and are a major component of the food web in many inland lakes, where they are often the most abundant prey for larger species such as walleye (*Sander vitreus*), northern pike (*Esox lucius*), muskellunge (*Esox masquinongy*), and lake trout (*Salvelinus namaycush*) (Becker, 1983). In the Laurentian Great Lakes, yellow perch are an important native species that have been heavily impacted by fishing pressure and environmental changes over the last century (Baldwin, Saalfeld, Dochoda, Buettner, & Eshenroder, 2009; Evans, 1986). Historically, yellow perch supported both commercial and recreational fisheries throughout the Great Lakes region. In Lake Michigan alone, the annual commercial harvest has been as high as 5.8 million lbs. (in 1964; Baldwin et al. 2009) representing nearly $16 million (US$) in dockside value and much more in retail value. Moreover, yellow perch is consistently among the most valuable commercially harvested fish species in the Great Lakes [$2.64/lb. dockside value in 2000; (Kinnunen, 2003)], with fillets selling as high as $12/lb. However, beginning in the late 1900s, yellow perch populations in Lake Michigan collapsed, suffering from consistently poor recruitment, interactions with invasive species like alewife (*Alosa pseudoharengus*), and overharvest (Marsden & Robillard, 2004; Wilberg, Bence, Eggold, Makauskas, & Clapp, 2005). In response, commercial harvest was closed in the main basin of Lake Michigan in 1997 and recreational limits were tightened, but abundance and recruitment have not recovered (Clapp & Dettmers, 2004). The population history of yellow perch in the Great Lakes makes them an excellent system to investigate the genetic basis of fisheries induced evolution (FIE) as well as the genetic impacts of overfishing, both topics with broad importance for global food security. Previous research in other species has utilized a metanalysis approach to demonstrate that fishing pressure can reduce genetic diversity (Hauser, Adcock, Smith, Ramírez, & Carvalho, 2002; Pinsky & Palumbi, 2014) and analyzed laboratory crosses to determine that FIE can induce substantial divergence at the genomic and phenotypic levels (Nina O. Therkildsen et al., 2019). Yellow perch provide a unique opportunity to improve on these previous studies, as scale samples collected over the last four decades could facilitate analysis of the genetic impacts of overfishing and subsequent fisheries closures in "real time."

From an aquaculture perspective, yellow perch has many desirable attributes. For example, yellow perch can tolerate high stocking densities, are relatively disease resistant, and can be raised successfully under a variety of temperature and water conditions (Jeffrey A. Malison, 2003; Jeffrey A. Malison & Held, 1992). Furthermore, yellow perch can be reared from hatching to marketable size in a relatively short period of time (~1 year *vs.* 2+ years for most

salmonids). Because yellow perch eat a diverse array of prey items (Keast, 1977), their feed can be obtained from ecologically sustainable sources while remaining cost effective (in contrast salmon are often fed a diet consisting primarily of other wild-caught fishes, known as fish meal). Lastly, yellow perch fillets have a firm texture and a mild flavor yielding a high market value. Because of these advantages, there has been considerable interest in developing an industry for farm raised yellow perch for > 30 years. However, production levels are still relatively low, with farmers raising only ~100,000 kg per year (Wallat, Tiu, Wang, Rapp, & Leighfield, 2005).

A sequenced genome will be a vital resource for research on yellow perch ecology and fisheries, as new techniques for low coverage whole genome resequencing make it possible to screen nearly all genomic polymorphisms to elucidate even subtle signals of reduced diversity and adaptation in response to fisheries exploitation (Fuentes-Pardo & Ruzzante, 2017; Nina Overgaard Therkildsen & Palumbi, 2017). Such a yellow perch genome reference would be a valuable resource for researchers facing many challenges related to conservation of yellow perch and could also be leveraged to address current limitations that have prevented the wide-scale adoption of yellow perch as an aquaculture species. The genome could be used to improve researchers understanding of adaptive divergence in yellow perch, leading to the creation of more robust management units that better preserve important adaptive diversity in this species. Additionally, researchers could leverage the genome to conduct marker assisted selection with the goal of creating faster growing populations of yellow perch. For example, one straightforward step towards obtaining fish with faster growth rates and larger body size would be using the genomic resources and sex genotyping assay presented here to aid in the production of genetically all-female populations, as females grow considerably faster and larger than males (J. A. Malison & Garcia-Abiado, 1996; Jeffrey A. Malison, Kayes, Wentworth, & Amundson, 1988; Rougeot, 2015). More generally, sequencing and characterizing the yellow perch genome will facilitate improvements in both aquaculture and fisheries management practices. Finally, a yellow perch genome will be useful to enable research in other ecologically and commercially important percid fishes, such as walleye (*Sander vitreus*), allowing for example to anchor transcriptome sequences, facilitating important studies related to aquaculture and adaptation to thermally challenging habitats in that species.

MATERIAL AND METHODS

Sampling and genomic DNA extraction

The male yellow perch [sample (1) in Fig. 1] used for whole genome sequencing (long-reads Oxford Nanopore Technologies and 10X Genomics) was sampled in April 2017 from Plum Lake, Vilas County, Wisconsin, USA (46°00'01.5"N 89°31'44.3). A 0.5 ml blood sample was taken from this animal and immediately put in 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) (Asahida, Kobayashi, Saitoh, & Nakayama, 1996). High molecular weight genomic DNA (gDNA) was then purified by phenol-chloroform extraction. For the chromosome contact map (Hi-C), 1.5 ml of blood was taken in January 2018 from a different male [sample (2) in Fig. 1] from a domesticated line of yellow perch raised at the Farmory, an aquaculture facility in Green

Bay, Wisconsin, USA (44°30'23.2"N 88°00'35.9"W). The fresh blood sample was slowly cryopreserved with 15 % Dimethyl sulfoxide (DMSO) in a Mr. Frosty Freezing Container (Thermo Scientific) at −80°C. Fin clip samples (30 males and 30 females) for whole-genome sequencing of pools of individuals (Pool-seq) [samples (3) in Fig. 1] were collected in September 2009 from wild yellow perch in Green Bay, Lake Michigan, Wisconsin, USA (44°32'19.0"N 88°00'16.6"W), placed in 90% ethanol and then stored dried until gDNA extraction was performed using the NucleoSpin Kit for Tissue (Macherey-Nagel, Duren, Germany). Genomic DNAs from individual fish were then quantified using a Qubit fluorometer (Thermofisher), gDNA concentrations were standardized among all samples and pooled in equimolar ratios by individual and sex, resulting in one gDNA pool for males and one gDNA pool for females. For validation of *amhr2by* sex-linkage, 50 phenotypically sexed individuals (25 males and 25 females) wild perch were sampled from a geographically isolated population from the one used for the previous Pool-sequencing experiment. These fish were collected in Lake Michigan in May 2018 using gill net sets off the shore of Michigan City, Indiana (41°42.5300'N, 86°57.5843'W). Upon collection, each individual fish was euthanized, phenotypic sex was determined by visual inspection of gonads during necropsy, and caudal fin clips were taken from each yellow perch individual and stored in 95% non-denatured ethanol. Genomic DNA was extracted using the DNeasy extraction kit and protocol (Qiagen).

DNA library construction and sequencing

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

10X Genomics sequencing—The Chromium library was prepared according to 10X Genomics' protocols using the Genome Reagent Kit v2. Sample quantity and quality controls were validated by Qubit, Nanodrop and Femto Pulse machines. The library was prepared from 10 ng of high molecular weight (HMW) gDNA. Briefly, in the microfluidic Genome Chip, a library of Genome Gel Beads was combined with HMW template gDNA in

master mix and partitioning oil to create Gel Bead-In-EMulsions (GEMs) in the Chromium apparatus. Each Gel Bead was then functionalized with millions of copies of a 10x™ barcoded primer. Dissolution of the Genome Gel Bead in the GEM released primers containing (i) an Illumina R1 sequence (Read 1 sequencing primer), (ii) a 16 bp 10x Barcode, and (iii) a 6 bp random primer sequence. R1 sequence and the 10x™ barcode were added to the molecules during the GEM incubation. P5 and P7 primers, R2 sequence, and Sample Index were added during library construction. 10 cycles of PCR were applied to amplify the library. Library quality was assessed using a Fragment Analyser and library was quantified by qPCR using the Kapa Library Quantification Kit. The library was then sequenced on a single lane of Illumina HiSeq3000 using a paired-end read length of 2×150 nt with the Illumina HiSeq3000 sequencing kits and produced 315 million read pairs.

Hi-C sequencing—_In situ_ Hi-C was performed according to previously described protocols (Foissac et al., 2019). Cryopreserved blood cells were defrosted, washed with PBS twice and counted. 5 million cells were then cross-linked with 1% formaldehyde in PBS, quenched with Glycine 0.125M and washed twice with PBS. Membranes were then disrupted with a Dounce pestle, nuclei were permeabilized using 0.5% SDS and then digested with _Hind_III endonuclease. 5'-overhangs at _Hind_III-cut restriction sites were filled-in, in the presence of biotin-dCTP with the Klenow large fragment, and then re-ligated at a NheI restriction site. Nuclei were lysed and DNA was precipitated and then purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Qubit fluorometric quantification system (Thermo). T4 DNA polymerase was used to remove un-ligated biotinylated ends. Then, the Hi-C library was prepared according to Illumina's protocols using the Illumina TruSeq Nano DNA HT Library Prep Kit with a few modifications: 1.4μg DNA was fragmented to 550nt by sonication. Sheared DNA was then sized (200–600pb) using Agencourt AMPure XP beads, and biotinylated ligation junctions were captured using M280 Streptavidin Dynabeads (Thermo) and then purified using reagents from the Nextera Mate Pair Sample preparation kit (Illumina). Using the TruSeq nano DNA kit (Illumina), the 3' ends of blunt fragments were adenylated. Next, adaptors and indexes were ligated and the library was amplified for 10 cycles. Library quality was assessed by quantifying the proportion of DNA cut by endonuclease NheI using a Fragment Analyzer (Advanced Analytical Technologies, Inc., Iowa, USA). Finally, the library was quantified by qPCR using the Kapa Library Quantification Kit (Roche). Sequencing was performed on an Illumina HiSeq3000 apparatus (Illumina, California, USA) using paired-end 2×150 nt reads. This produced 128 million read pairs (38.4 Gb of raw nucleotides).

Pool sequencing—Pool-sequencing libraries were prepared according to Illumina's protocols using the Illumina TruSeq Nano DNA HT Library Prep Kit (Illumina, California, USA). In short, 200 ng of each gDNA pool (males and females pools) was fragmented to 550 bp by sonication on M220 Focused-ultrasonicator (COVARIS). Size selection was performed using SPB beads (kit beads) and the 3' ends of blunt fragments were mono-adenylated. Then, adaptors and indexes were ligated and the construction was amplified with Illumina-specific primers. Library quality was assessed using a Fragment Analyzer (Advanced Analytical Technologies, Inc., Iowa, USA) and libraries were quantified by qPCR using the Kapa Library Quantification Kit (Roche). Sequencing of the male and

female pools were performed on the same NovaSeq (Illumina, California, USA) lane using a paired-end read length of 2×150 nt with Illumina NovaSeq Reagent Kits. Sequencing produced 119 million paired reads for the male pool library and 132 million paired reads for the female pool library.

Genome assembly and analysis

Genome size estimation—K-mer-based estimation of the genome size was carried out with GenomeScope (Vurture et al., 2017). 10X reads were processed with Jellyfish v1.1.11 (Marçais & Kingsford, 2011) to count 17-, 19-, 21-, 23- and 25-mers with a max k-mer coverage of 10,000.

Genome assembly—GridION and PromethION data were trimmed using Porechop v0.2.1 (Wick, 2017/2019), corrected using Canu v1.6 (Koren et al., 2017) and filtered to keep only reads longer than 10 kbp. Corrected reads were then assembled using SmartDeNovo version of May-2017 (Ruan, 2015/2019) with default parameters. The assembly base pair quality was improved by several polishing steps including two rounds of long read alignment to the draft genome with minimap2 v2.7 (H. Li, 2018) followed by Racon v1.3.1 (Vaser, Sovi , Nagarajan, & Šiki , 2017), as well as three rounds of 10X genomics short read alignments using Long Ranger v2.1.1 (10x Genomics 2018) followed by Pilon v1.22 (Walker et al., 2014). The polished genome assembly was then scaffolded using Hi-C as a source of linking information. Reads were aligned to the draft genome using Juicer (Durand et al., 2016) with default parameters. A candidate assembly was then generated with 3D de novo assembly (3D-DNA) pipeline (Dudchenko et al., 2017) with the -r 0 parameter. Finally, the candidate assembly was manually reviewed using Juicebox Assembly Tools (Durand et al., 2016). Genome completeness was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) based on 4,584 BUSCO orthologs derived from the Actinopterygii lineage.

Genome annotation—The first annotation step was to identify repetitive content using RepeatMasker v4.0.7 (Tarailo-Graovac & Chen, 2009), Dust (Kuzio et al., unpublished but described in (Morgulis, Gertz, Schäffer, & Agarwala, 2006)), and TRF v4.09 (Benson, 1999). A species-specific *de novo* repeat library was built with RepeatModeler v1.0.11 (http://www.repeatmasker.org/RepeatModeler/) and repeated regions were located using RepeatMasker with the *de novo* and *Danio rerio* libraries. Bedtools v2.26.0 (Quinlan & Hall, 2010) was used to aggregate repeated regions identified with the three tools and to soft mask the genome. The MAKER3 genome annotation pipeline v3.01.02-beta (Holt & Yandell, 2011) combined annotations and evidence from three approaches: similarity with fish proteins, assembled transcripts (see below), and *de novo* gene predictions. Protein sequences from 11 fish species (*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Xiphophorus maculatus*) found in Ensembl were aligned to the masked genome using Exonerate v2.4 (Slater & Birney, 2005) with the alignment model protein2genome that allows translated alignments with modelling of introns. As *Perca fluviatilis* is a relatively closely related species from *P. flavescens*

[divergence time is estimated to be 19.8 million years ago according to (Couture, Pyle, & Pyle, 2015)], RNA-Seq reads of *P. fluviatilis* (NCBI BioProject PRJNA256973) from the PhyloFish project (Pasquier et al., 2016) were used for genome annotation and aligned to the chromosomal assembly using STAR v2.5.1b (Dobin et al., 2013) with outWigType and outWigStrand options to output signal wiggle files. Cufflinks v2.2.1 (Trapnell et al., 2010) was used to assemble the transcripts which were used as RNA-seq evidence. Braker v2.0.4 (Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2016) provided *de novo* gene models with wiggle files provided by STAR as hint files for GeneMark and Augustus training. The best supported transcript for each gene was chosen using the quality metric called Annotation Edit Distance (AED) (Eilbeck, Moore, Holt, & Yandell, 2009). Genome annotation gene completeness was assessed by BUSCO using the Actinopterygii group. Finally, predicted genes were subjected to similarity searches against the NCBI NR database using Diamond v0.9.22 (Buchfink, Xie, & Huson, 2015). The top hit with a coverage over 70% and identity over 80% was retained.

Pool-sequencing analysis—Reads from the male and female pools were aligned to the chromosomal assembly with BWA mem (version 0.7.17, (H. Li, 2013)), and the resulting BAM files were sorted and PCR duplicates removed using Picard tools (version 2.18.2). A file containing the nucleotide composition of each pool for each genomic position was generated using samtools mpileup [version 1.8, (H. Li et al., 2009)] and popoolation2 mpileup2sync [version 1201, (Kofler, Pandey, & Schlötterer, 2011)]. This file was then analyzed with custom software (PSASS version 2.0.0: https://zenodo.org/record/2615936#.XTyIS3s6_AI) to compute 1) the F_{ST} between males and females in a sliding window along the genome which is used to identify regions with strong differentiation between the male and female genomes, 2) the position and density of sex-specific SNPs, defined as SNPs heterozygous in one sex while homozygous in the other sex, which is correlated with F_{ST} between males and females, but specifically indicates that a region is heterozygous in one sex while homozygous in the other sex , and 3) the absolute and relative read depths for the male and female pools along the genome to look for regions present in one sex and absent in the other (e.g. sex-specific insertions). PSASS was run with default parameters except --window-size which was set to 5000 and --output-resolution which was set to 1000. In the yellow perch analysis, all metrics were computed but only the absolute and relative read depths for the male and female pools were found to be informative and used to characterize the sex locus region.

Validation of *amhr2by* sex-linkage—To validate the sex-linkage of *amhr2by* in males suggested by the pool-sequencing results, two primer sets were designed based on the alignment of yellow perch *amhr2a* and *amhr2by* genes with one primer pair specific for the autosomal *amhr2a* gene (forward: 5'-GGGAAACGTGGGAAACTCAC-3', and reverse: 5'-AGCAGTAGTTACAGGGCACA-3', expected fragment size: 638 bp) and one primer pair specific for the Y chromosomal *amhr2by* gene (forward: 5'-TGGTGTGTGGCAGTGATACT-3', and reverse: 5'-ACTGTAGTTAGCGGGCACAT-3', expected fragment size: 443 bp). Gene alignments were run with mVISTA (Frazer, Pachter, Poliakov, Rubin, & Dubchak, 2004). Primers were sourced from Integrated Data Technologies (IDT). All samples were run blind with respect to phenotypic sex; the male

and female samples were randomized, and their phenotypic sex was not cross referenced with field data until gel electrophoresis was run on the final PCR products. Genotyping was carried out on each gDNA sample using a multiplexed PCR approach. The PCR reaction solution was composed of 50 μl of PCR Master Mix (Quiagen), 10 μl of each primer (40 μl total), and 10 μl of gDNA (concentrations of gDNA ranging from 150 to 200 ng/ μl) for a total reaction volume of 100 μl. Thermocycling conditions were 1 cycle of 3 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 51°C, and 1 min at 72°C, and finishing with 10 min incubation at 72°C. PCR products were loaded on a 1.5 % agarose gel, run at 100V for 45 minutes and visualized with a UVP UVsolo touch UV box.

RESULTS AND DISCUSSION

Genome characteristics

Using a combination of Oxford Nanopore Technologies (ONT) long-reads, 10X genomics Illumina short linked reads (PE150 chemistry), and a chromosome contact map (Hi-C), we generated a high-continuity, chromosome length *de novo* genome assembly of the yellow perch (Fig. 1A). Before the Hi-C integration step, the assembly yielded a genome size of 877 Mb with 879 contigs, a N50 contig size of 4.3 Mb, and a L50 contig number of 60 (*i.e.*, half of the assembled genome is included in the 60 longest contigs). After Hi-C integration, the genome assembled into 269 fragments with a total length of 877.4 Mb, including 24 chromosome-length scaffolds representing 98.78 % of the complete genome sequence (N50 = 37.4 Mb, L50 = 11) (see Table 1). Genome sizes are both very close to the 873 Mbp GenomeScope (Vurture et al., 2017) estimation based on short-read analysis with a repeat length of 266 Mbp (30.5%) and slightly lower than the estimation of *P. flavescens* genome sizes based on C-values [900 Mbp and 1200 Mbp records in the Animal Genome Size Database (http://www.genomesize.com/index.php)]. The 24 chromosome-length scaffolds obtained after Hi-C integration are consistent with the diploid chromosome (Chr) number of yellow perch (2n = 48) (Ráb, Roth, & Mayr, 1987). The genome completeness of these assemblies was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0 (Simão et al., 2015) based on the Actinopterygii database. BUSCO scores (see Table 1) of the pre-Hi-C and post-Hi-C assemblies were similar (Complete BUSCOs between 97.6% and 97.8%), with an unexpected slight decrease of the Post-Hi-C BUSCO's scores and small values for both fragmented (< 1%) and missing (< 1.5%) BUSCO genes. The reason behind this slight decrease of the Post-Hi-C BUSCO's scores was explored in detail (see http://genoweb.toulouse.inra.fr/~sigenae/GenoFish_public/BUSCO_and_assembly.html) and is not related to gene fragmentation due to the Hi-C integration.

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

The comparison of our yellow perch assembly with the published Eurasian perch genome assembly (Ozerov et al., 2018) shows that the Eurasian perch genome assembly (31,105 scaffolds) is much more fragmented than our yellow perch assembly (267 scaffolds) with N50 and L50 metrics reflecting this fragmentation of the Eurasian perch genome assembly (Table 2). This difference is most likely technological as the Eurasian perch genome has been sequenced and assembled based on a single approach with the 10x Genomics methods (Ozerov et al., 2018). Although 10x Genomics alone has been shown to produce reasonably good quality genome assemblies (Hammond et al., 2017; C. Li et al., 2018; Louro et al., 2019; Ozerov et al., 2018), the current standard to produce highly contiguous reference genome assemblies consists in the combination of different approaches including long-reads technologies to produce large-size contigs (Jain et al., 2018) and chromosome contact maps to build chromosome-scale scaffolds (Dudchenko et al., 2017).

Yellow perch sex-determination

Yellow perch has a male monofactorial heterogametic sex determination system (XX/XY) (Jeffrey A. Malison, Kayes, Best, Amundson, & Wentworth, 1986) with undifferentiated sex chromosomes (Beçak, Beçak, Roberts, Shoffner, & Volpe, 1973). Using a male-versus-female pooled gDNA whole genome sequencing strategy (Gammerdinger, Conte, Baroiller, D'Cotta, & Kocher, 2016) (Fig. 1B), we identified a relatively small region of 100 kb localized at the proximal end of chromosome 9 (Chr09:0–100,000 bp) with a complete absence of female reads, excluding repeated elements (Fig. 2.A–B). This coverage pattern strongly supports the hypothesis that Chr09 is the yellow perch sex chromosome and contains a small Y-specific region in phenotypic males that is completely absent from phenotypic females. Genome annotation shows that this Y-specific insertion on Chr09 contains a duplicate copy (*amhr2by*) of the autosomal anti-Mullerian hormone receptor gene located on Chr04 (*amhr2a*). The absolute average read depth per base pair for *amhr2by* (gene sequence from start codon to stop codon) was 11.4 for males (compared to a whole genome average of 27.2) and 0.8 for females (whole genome average of 26.3). The *amhr2* gene has previously been characterized as a master sex-determining gene in some pufferfishes (Ieda et al., 2018; Kamiya et al., 2012) and the *hotei* mutation in the medaka *amhr2* gene induces a male-to-female sex-reversal of genetically XY fish (Morinaga et al., 2007). However, in contrast to pufferfishes, in which the differentiation of X and Y chromosomes is extremely limited and originated from an allelic diversification process, the yellow perch *amhr2by* sequence is quite divergent from its *amhr2a* autosomal counterpart. Specifically, the *amhr2by* gene shows only 88.3 % identity with *amhr2a* in the aligned coding sequence and 89.1 % in the aligned parts of the introns, but with many long gaps and indels in the introns (Fig. 2C–D). This nucleotide sequence divergence impacts the protein sequence of the yellow perch *amhr2by* gene (Fig. 2D–2E), but due to a complete absence of exons 1 & 2 (Fig. 2C–2E) compared to its autosomal counterpart, the yellow perch Amhr2by protein translates as a N-terminal-truncated type II receptor that lack most of the cysteine-rich extracellular part of the receptor, which is crucially involved in ligand binding specificity (Heldin & Moustakas, 2016).

To validate the male specificity of this potential Y-specific insertion, we designed primers specific for both *amhr2by* and *amhr2a* and genotyped 25 male and 25 female yellow perch

collected from a Southeastern Lake Michigan population, which is geographically isolated from the Plum Lake (Wisconsin) population of the 30 males and 30 females used for initial analysis with pool-sequencing. The presence/absence of the *amhr2by* PCR product was perfectly correlated with the determined phenotypic sex, with the amplification of an *amhr2by* fragment only in the 25 males and no amplification in the 25 females (see Fig. 2F for 18 of the 50 individuals tested and Fig. S1 for all animals). The simultaneous amplification of the *amhr2a* fragment in both males and in females provided an internal control preventing single-locus dropout in such a multiplexed PCR reaction.

This complete sex-linkage result makes the yellow perch *amhr2by* a strong candidate as a sex determining gene. Interestingly, anti-Mullerian hormone (Amh) has been also characterized as a male-promoting gene in zebrafish (Yan et al., 2019) and as a master sex determining gene both in Patagonian pejerrey (Hattori et al., 2012), Nile tilapia (M. Li et al., 2015) and Northern pike (Pan et al., 2019). The anti-Mullerian hormone belongs to the transforming growth factor beta (TGF-ß) family that contains structurally related growth factors involved in many differentiation processes. But TGF-ß members involved in sex determination are not limited to the Amh pathway as additional TGF-ß family genes have also been characterized as master sex determining genes, including growth differentiation factor 6 (*gdf6*) in the turquoise killifish (Reichwald et al., 2015) and gonadal soma derived factor (*gsdf*) in the Luzon medaka and the sablefish (Myosho et al., 2012; Rondeau et al., 2013). Additional evidence, including loss of *amhr2by* function experiments in XY males and gain of *amhr2by* function experiments in XX females, is necessary to critically test the hypothesis that this male-specific *amhr2by* duplication really functions as a master sex determining gene in yellow perch. However, given the known importance of the Amh pathway in fish sex determination, and that no other gene in that small sex locus is known to play a role in sex differentiation, *amhr2by* is a prime candidate for the yellow perch master sex determining gene. This finding provides another example of the recurrent utilization of the TGF-ß pathway in fish sex determination, and thus supports the 'limited option' hypothesis (Marshall Graves & Peichel, 2010), which states that some genes are more likely than others to be selected as master sex determining genes. How this N terminal truncated Amhr2 could trigger its function as a master sex determining gene is as yet unknown, but our hypothesis is that this truncation constitutively activates the Amh receptor causing it to signal in the absence of Amh ligand.

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

REFERENCES

Asahida T, Kobayashi T, Saitoh K, & Nakayama I (1996). Tissue Preservation and Total DNA Extraction form Fish Stored at Ambient Temperature Using Buffers Containing High Concentration of Urea. Fisheries Science, 62, 727–730. doi: doi:10.2331/fishsci.62.727

Baldwin N, Saalfeld R, Dochoda M, Buettner H, & Eshenroder L (2009). Commercial Fish Production in the Great Lakes 1867–2009. Great Lakes Fishery Commission, Ann Arbor, MI.

Beçak ML, Beçak W, Roberts FL, Shoffner RN, & Volpe EP (1973). Perca flavescens (Mitchill) (Yellow perch) 2n = 48 In Beçak ML, Beçak W, Roberts FL, Shoffner RN, & Volpe EP (Eds.), Chromosome Atlas: Fish, Amphibians, Reptiles, and Birds: Volume 2 (pp. 5–7). doi: 10.1007/978-3-642-65751-1_2

Becker GC (1983). Fishes of Wisconsin. Retrieved from http://digicoll.library.wisc.edu/cgi-bin/ EcoNatRes/EcoNatRes-idx?id=EcoNatRes.FishesWI

Beletsky D, Mason DM, Schwab DJ, Rutherford ES, Janssen J, Clapp DF, & Dettmers JM (2007). Biophysical Model of Larval Yellow Perch Advection and Settlement in Lake Michigan. Journal of Great Lakes Research, 33(4), 842–866. doi: 10.3394/0380-1330(2007)33[842:BMOLYP]2.0.CO;2

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research, 27(2), 573–580. [PubMed: 9862982]

Bergek S, & Björklund M (2009). Genetic and morphological divergence reveals local subdivision of perch (Perca fluviatilis L.). Biological Journal of the Linnean Society, 96(4), 746–758. doi: 10.1111/j.1095-8312.2008.01149.x

Böhling P, & Lehtonen H (1984). Effect of environmental factors on migrations of perch (Perca fluviatilis L.) tagged in the coastal waters of Finland. Finnish Fisheries Research, 5, 31–40.

Brazo DC, Tack PI, & Liston CR (1975). Age, Growth, and Fecundity of Yellow Perch, Perca flavescens, in Lake Michigan near Ludington, Michigan. Transactions of the American Fisheries Society, 104(4), 726–730. doi: 10.1577/1548-8659(1975)104<726:AGAFOY>2.0.CO;2

Buchfink B, Xie C, & Huson DH (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59–60. doi: 10.1038/nmeth.3176 [PubMed: 25402007]

Clapp DF, & Dettmers JM (2004). Yellow Perch Research and Management in Lake Michigan. Fisheries, 29(11), 11–19. doi: 10.1577/1548-8446(2004)29[11:YPRAMI]2.0.CO;2

Couture P, Pyle G, & Pyle G (2015, 8 5). Evolutionary Relationships, Population Genetics, and Ecological and Genomic Adaptations of Perch (Perca). doi: 10.1201/b18806-5

Dettmers JM, Janssen J, Pientka B, Fulford RS, & Jude DJ (2005). Evidence across multiple scales for offshore transport of yellow perch (Perca flavescens) larvae in Lake Michigan. Canadian Journal of Fisheries and Aquatic Sciences, 62(12), 2683–2693. doi: 10.1139/f05-173

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, … Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England), 29(1), 15–21. doi: 10.1093/bioinformatics/bts635

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, … Aiden EL (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science (New York, N.Y.), 356(6333), 92–95. doi: 10.1126/science.aal3327

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, & Aiden EL (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems, 3(1), 95–98. doi: 10.1016/j.cels.2016.07.002 [PubMed: 27467249]

Eilbeck K, Moore B, Holt C, & Yandell M (2009). Quantitative measures for the management and comparison of annotated genomes. BMC Bioinformatics, 10, 67. doi: 10.1186/1471-2105-10-67 [PubMed: 19236712]

Evans MS (1986). Recent Major Declines in Zooplankton Populations in the Inshore Region of Lake Michigan: Probable Causes and Implications. Canadian Journal of Fisheries and Aquatic Sciences, 43(1), 154–159. doi: 10.1139/f86-017

Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, … Giuffra E (2019). Transcriptome and chromatin structure annotation of liver, CD4+ and CD8+ T cells from four livestock species. BioRxiv, 316091. doi: 10.1101/316091

Forney JL (1971). Development of Dominant Year Classes in a Yellow Perch Population. Transactions of the American Fisheries Society, 100(4), 739–749. doi: 10.1577/1548-8659(1971)100<739:DODYCI>2.0.CO;2

Frazer KA, Pachter L, Poliakov A, Rubin EM, & Dubchak I (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Research, 32(Web Server issue), W273–279. doi: 10.1093/nar/gkh458 [PubMed: 15215394]

Fuentes-Pardo AP, & Ruzzante DE (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. Molecular Ecology, 26(20), 5369–5406. doi: 10.1111/mec.14264 [PubMed: 28746784]

Gammerdinger WJ, Conte MA, Baroiller J-F, D'Cotta H, & Kocher TD (2016). Comparative analysis of a sex chromosome from the blackchin tilapia, Sarotherodon melanotheron. BMC Genomics, 17(1), 808. doi: 10.1186/s12864-016-3163-7 [PubMed: 27756226]

Glover DC, Dettmers JM, Wahl DH, & Clapp DF (2008). Yellow perch (Perca flavescens) stock structure in Lake Michigan: an analysis using mark–recapture data. Canadian Journal of Fisheries and Aquatic Sciences, 65(9), 1919–1930. doi: 10.1139/F08-100

Hammond SA, Warren RL, Vandervalk BP, Kucuk E, Khan H, Gibb EA, … Birol I (2017). The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nature Communications, 8. doi: 10.1038/s41467-017-01316-7

Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, … Strüssmann CA (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. Proceedings of the National Academy of Sciences of the United States of America, 109(8), 2955–2959. doi: 10.1073/pnas.1018392109 [PubMed: 22323585]

Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, & Carvalho GR (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). Proceedings of the National Academy of Sciences, 99(18), 11742–11747. doi: 10.1073/pnas.172242899

Heldin C-H, & Moustakas A (2016). Signaling Receptors for TGF-β Family Members. Cold Spring Harbor Perspectives in Biology, 8(8), a022053. doi: 10.1101/cshperspect.a022053 [PubMed: 27481709]

Higgins DG, & Sharp PM (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene, 73(1), 237–244. doi: 10.1016/0378-1119(88)90330-7 [PubMed: 3243435]

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, & Stanke M (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics (Oxford, England), 32(5), 767–769. doi: 10.1093/bioinformatics/btv661

Holt C, & Yandell M (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics, 12(1), 491. doi: 10.1186/1471-2105-12-491 [PubMed: 22192575]

Höök TO, McCormick MJ, Rutherford ES, Mason DM, & Carter GS (2006). Short-term Water Mass Movements in Lake Michigan: Implications for Larval Fish Transport. Journal of Great Lakes Research, 32(4), 728–737. doi: 10.3394/0380-1330(2006)32[728:SWMMIL]2.0.CO;2

Ieda R, Hosoya S, Tajima S, Atsumi K, Kamiya T, Nozawa A, … Kikuchi K (2018). Identification of the sex-determining locus in grass puffer (Takifugu niphobles) provides evidence for sex-chromosome turnover in a subset of Takifugu species. PLOS ONE, 13(1), e0190635. doi: 10.1371/journal.pone.0190635 [PubMed: 29293639]

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, … Loose M (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature Biotechnology, 36(4), 338–345. doi: 10.1038/nbt.4060

Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, … Kikuchi K (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (fugu). PLoS Genetics, 8(7), e1002798. doi: 10.1371/journal.pgen.1002798 [PubMed: 22807687]

Keast A (1977). Diet overlaps and feeding relationships between the year classes in the yellow perch (Perca flavescens). Environmental Biology of Fishes, 2(1), 53–70. doi: 10.1007/BF00001416

Kinnunen RE (2003). Great Lakes Commercial Fisheries. Michigan Sea Grant Extension, Marquette (p. 54). Retrieved from https://pdfs.semanticscholar.org/d243/74bc18e240e29a8418e96e766ea8e8be50ed.pdf

Kofler R, Pandey RV, & Schlötterer C (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics (Oxford, England), 27(24), 3435–3436. doi: 10.1093/bioinformatics/btr589

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, & Phillippy AM (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research, 27(5), 722–736. doi: 10.1101/gr.215087.116 [PubMed: 28298431]

Leis JM (2006). Are Larvae of Demersal Fishes Plankton or Nekton? In Advances in Marine Biology (Vol. 51, pp. 57–141). doi: 10.1016/S0065-2881(06)51002-8

Li C, Liu X, Liu B, Ma B, Liu F, Liu G, … Wang C (2018). Draft genome of the Peruvian scallop Argopecten purpuratus. GigaScience, 7(4). doi: 10.1093/gigascience/giy031

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-Bio]. Retrieved from http://arxiv.org/abs/1303.3997

Li H (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics (Oxford, England), 34(18), 3094–3100. doi: 10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, … Wang D (2015). A Tandem Duplicate of Anti-Müllerian Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, Oreochromis niloticus. PLOS Genetics, 11(11), e1005678. doi: 10.1371/journal.pgen.1005678 [PubMed: 26588702]

Louro B, De Moro G, Garcia C, Cox CJ, Veríssimo A, Sabatino SJ, … Canário AVM (2019). A haplotype-resolved draft genome of the European sardine (Sardina pilchardus). GigaScience, 8(5). doi: 10.1093/gigascience/giz059

Ludsin SA, DeVanna KM, & Smith REH (2014). Physical–biological coupling and the challenge of understanding fish recruitment in freshwater lakes. Canadian Journal of Fisheries and Aquatic Sciences, 71(5), 775–794. doi: 10.1139/cjfas-2013-0512

Malison JA, & Garcia-Abiado M. a. R. (1996). Sex control and ploidy manipulations in yellow perch (Perca flavescens) and walleye (Stizostedion vitreum). Journal of Applied Ichthyology, 12(3–4), 189–194. doi: 10.1111/j.1439-0426.1996.tb00088.x

Malison Jeffrey A. (2003). A white paper on the status and needs of yellow perch aquaculture in the north central region North Central Regional Aquaculture Center, Michigan State University, East Lansing Retrieved from https://www.ncrac.org/status-and-needs-yellow-perch-aquaculture-north-central-region

Malison Jeffrey A., & Held JA (1992). Effects of fish size at harvest, initial stocking density and tank lighting conditions on the habituation of pond-reared yellow perch (Perca flavescens) to intensive culture conditions. Aquaculture, 104(1), 67–78. doi: 10.1016/0044-8486(92)90138-B

Malison Jeffrey A., Kayes TB, Best CD, Amundson CH, & Wentworth BC (1986). Sexual Differentiation and Use of Hormones to Control Sex in Yellow Perch (Perca flavescens). Canadian Journal of Fisheries and Aquatic Sciences, 43(1), 26–35. doi: 10.1139/f86-004

Malison Jeffrey A., Kayes TB, Wentworth BC, & Amundson CH (1988). Growth and Feeding Responses of Male versus Female Yellow Perch (Perca flavescens) Treated with Estradiol-17β. Canadian Journal of Fisheries and Aquatic Sciences, 45(11), 1942–1948. doi: 10.1139/f88-226

Marçais G, & Kingsford C (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics (Oxford, England), 27(6), 764–770. doi: 10.1093/bioinformatics/btr011

Marsden JE, & Robillard SR (2004). Decline of Yellow Perch in Southwestern Lake Michigan, 1987–1997. North American Journal of Fisheries Management, 24(3), 952–966. doi: 10.1577/M02-195.1

Marshall Graves JA, & Peichel CL (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? Genome Biology, 11(4), 205. doi: 10.1186/gb-2010-11-4-205 [PubMed: 20441602]

Morgulis A, Gertz EM, Schäffer AA, & Agarwala R (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. Journal of Computational Biology: A Journal of Computational Molecular Cell Biology, 13(5), 1028–1040. doi: 10.1089/cmb.2006.13.1028 [PubMed: 16796549]

Morinaga C, Saito D, Nakamura S, Sasaki T, Asakawa S, Shimizu N, … Kondoh H (2007). The hotei mutation of medaka in the anti-Mullerian hormone receptor causes the dysregulation of germ cell and sexual development. Proceedings of the National Academy of Sciences of the United States of America, 104(23), 9691–9696. doi: 10.1073/pnas.0611379104 [PubMed: 17535919]

Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, … Sakaizumi M (2012). Tracing the emergence of a novel sex-determining gene in medaka, Oryzias luzonensis. Genetics, 191(1), 163–170. doi: 10.1534/genetics.111.137497 [PubMed: 22367037]

Ozerov MY, Ahmad F, Gross R, Pukk L, Kahar S, Kisand V, & Vasemägi A (2018). Highly Continuous Genome Assembly of Eurasian Perch (Perca fluviatilis) Using Linked-Read Sequencing. G3: Genes, Genomes, Genetics, 8(12), 3737–3743. doi: 10.1534/g3.118.200768 [PubMed: 30355765]

Pan Q, Feron R, Yano A, Guyomard R, Jouanno E, Vigouroux E, … Guiguen Y (2019). Identification of the master sex determining gene in Northern pike (Esox lucius) reveals restricted sex chromosome differentiation. PLOS Genetics, 15(8), e1008013. doi: 10.1371/journal.pgen.1008013 [PubMed: 31437150]

Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, … Bobe J (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics, 17, 368. doi: 10.1186/s12864-016-2709-z [PubMed: 27189481]

Pinsky ML, & Palumbi SR (2014). Meta-analysis reveals lower genetic diversity in overfished populations. Molecular Ecology, 23(1), 29–39. doi: 10.1111/mec.12509 [PubMed: 24372754]

Pritt JJ, Roseman EF, & O'Brien TP (2014). Mechanisms driving recruitment variability in fish: comparisons between the Laurentian Great Lakes and marine systems. ICES Journal of Marine Science, 71(8), 2252–2267. doi: 10.1093/icesjms/fsu080

Quinlan AR, & Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England), 26(6), 841–842. doi: 10.1093/bioinformatics/btq033

Ráb P, Roth P, & Mayr B (1987). Karyotype Study of Eight Species of European Percid Fishes (Pisces, Percidae). Caryologia, 40(4), 307–318. doi: 10.1080/00087114.1987.10797833

Reichwald K, Petzold A, Koch P, Downie BR, Hartmann N, Pietsch S, … Platzer M (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. Cell, 163(6), 1527–1538. doi: 10.1016/j.cell.2015.10.071 [PubMed: 26638077]

Robillard SR, & Marsden JE (2001). Spawning Substrate Preferences of Yellow Perch along a Sand–Cobble Shoreline in Southwestern Lake Michigan. North American Journal of Fisheries Management, 21(1), 208–215. doi: 10.1577/1548-8675(2001)021<0208:SSPOYP>2.0.CO;2

Rondeau EB, Messmer AM, Sanderson DS, Jantzen SG, von Schalburg KR, Minkley DR, … Koop BF (2013). Genomics of sablefish (Anoplopoma fimbria): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. BMC Genomics, 14(1), 452. doi: 10.1186/1471-2164-14-452 [PubMed: 23829495]

Rougeot C (2015). Sex and Ploidy Manipulation in Percid Fishes In Kestemont P, Dabrowski K, & Summerfelt RC (Eds.), Biology and Culture of Percid Fishes: Principles and Practices (pp. 625–634). doi: 10.1007/978-94-017-7227-3_23

Ruan J (2019). Ultra-fast de novo assembler using long noisy reads: ruanjue/smartdenovo [C]. Retrieved from https://github.com/ruanjue/smartdenovo (Original work published 2015)

Schneeberger PJ (2000). Population Dynamics of Contemporary Yellow Perch and Walleye Stocks in Michigan Waters of Green Bay, Lake Michigan, 1988–96

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210–3212. doi: 10.1093/bioinformatics/btv351 [PubMed: 26059717]

Slater GSC, & Birney E (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics, 6, 31. doi: 10.1186/1471-2105-6-31 [PubMed: 15713233]

Tarailo-Graovac M, & Chen N (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics, Chapter 4, Unit 4.10. doi: 10.1002/0471250953.bi0410s25

Therkildsen Nina O., Wilder AP, Conover DO, Munch SB, Baumann H, & Palumbi SR (2019). Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. Science, 365(6452), 487–490. doi: 10.1126/science.aaw7271 [PubMed: 31371613]

Therkildsen Nina Overgaard, & Palumbi SR (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Molecular Ecology Resources, 17(2), 194–208. doi: 10.1111/1755-0998.12593 [PubMed: 27496322]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, … Pachter L (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 28(5), 511–515. doi: 10.1038/nbt.1621

Vaser R, Sovi I, Nagarajan N, & Šiki M (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research, 27(5), 737–746. doi: 10.1101/gr.214270.116 [PubMed: 28100585]

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, & Schatz MC (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics (Oxford, England), 33(14), 2202–2204. doi: 10.1093/bioinformatics/btx153

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, … Earl AM (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS One, 9(11), e112963. doi: 10.1371/journal.pone.0112963 [PubMed: 25409509]

Wallat GK, Tiu LG, Wang HP, Rapp D, & Leighfield C (2005). The Effects of Size Grading on Production Efficiency and Growth Performance of Yellow Perch in Earthen Ponds. North American Journal of Aquaculture, 67(1), 34–41. doi: 10.1577/FA04-003.1

Whiteside MC, Swindoll CM, & Doolittle WL (1985). Factors affecting the early life history of yellow perch,Perca flavescens. Environmental Biology of Fishes, 12(1), 47–56. doi: 10.1007/BF00007709

Wick R (2019). adapter trimmer for Oxford Nanopore reads. Contribute to rrwick/Porechop development by creating an account on GitHub [C++]. Retrieved from https://github.com/rrwick/Porechop (Original work published 2017)

Wilberg MJ, Bence JR, Eggold BT, Makauskas D, & Clapp DF (2005). Yellow Perch Dynamics in Southwestern Lake Michigan during 1986–2002. North American Journal of Fisheries Management, 25(3), 1130–1152. doi: 10.1577/M04-193.1

Yan Y-L, Batzel P, Titus T, Sydes J, Desvignes T, Bremiller R, … Postlethwait JH (2019). The roles of Amh in zebrafish gonad development and sex determination. BioRxiv, 650218. doi: 10.1101/650218
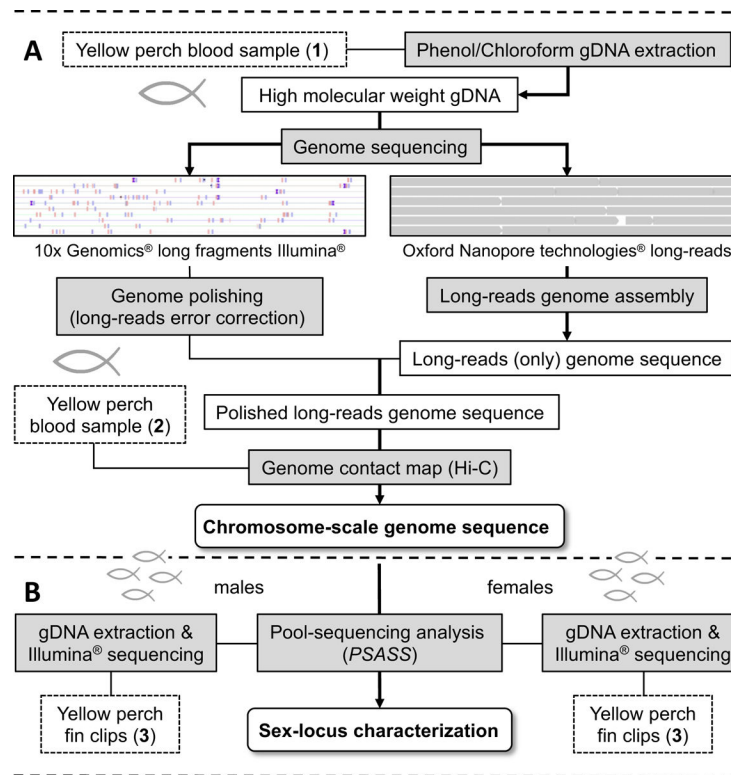
Figure 1: Flow chart of the genome sequencing, assembly strategies, and sex-locus characterization.

A. To build a high-continuity, chromosome-scale *de novo* genome assembly in the yellow perch we first extracted high molecular weight (HMW) genomic DNA (gDNA) from a blood sample taken on a yellow perch male [individual sample (1)]. This HMW gDNA was used as a template for whole genome sequencing using a combination of two sequencing approaches i.e., Oxford Nanopore Technologies (ONT) long-reads and 10X genomics Illumina short linked reads (PE150 chemistry). ONT long-reads were assembled separately to produce a ONT-based long-reads genome assembly. As classical long-reads sequencing technologies are known to be error prone (10–15 % errors), this long-reads only genome assembly was subsequently polished by some iterative corrections with the 10X genomics Illumina short linked reads resulting in polished long-reads genome assembly. Cryopreserved blood from a second yellow perch male [individual sample (2)] was then used to produce a chromosomal contact map using a Hi-C chromosome conformation capture method. This chromosomal contact map information was then integrated with our polished long-reads genome assembly in order to produce a final yellow perch chromosomal-scale genome assembly. **B**. To characterize the yellow perch sex locus region we performed whole genome sequencing of a pool of male gDNA and a pool of female gDNA (Pool-sequencing approach). These pool-sequencing datasets were analyzed with the PSASS (https://zenodo.org/record/2615936#.XTyIS3s6_AI) pipeline and these results were used to develop a sexual genotyping assay. See text for details.
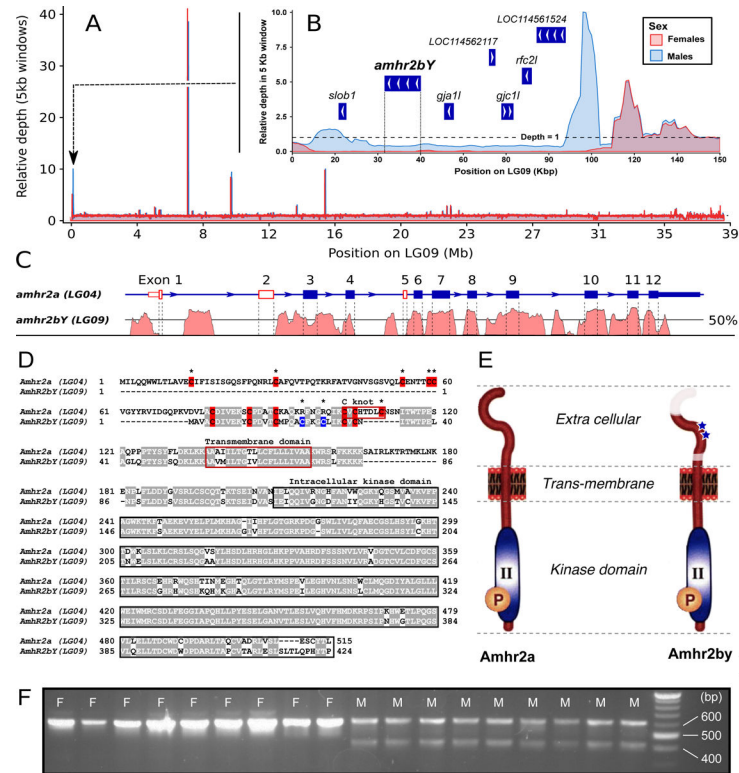
Figure 2: Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone receptor gene (*amhr2by*) in yellow perch.

A. Pool-seq data illustrating relative read depth across chromosome 9 (Chr09) for the male (blue line) and female (red line) pools showing a coverage difference between males (blue area) and females (red area) in the first 100 kb of Chr09. **B**. Zoom-in on the read depth difference between males and females in the first 150 kb of Chr09. Gene annotation is represented by blue boxes with arrows to indicate transcript orientation (NCBI *Perca flavescens* Annotation Release 100). Abbreviations: *slob1* (probable inactive serine/threonine-protein kinase slob1, *LOC114561790*), *amhr2by* (anti-Mullerian hormone type-2 receptor-like, *LOC114561927*), *gja1l* (gap junction alpha-1 protein-like, *LOC114562210*), *gjc1l* (gap junction gamma-1 protein-like, *LOC114562012*), *rfc2l* (replication factor C subunit 2-like, *LOC114561955*). **C**. Identity plot of the alignment of *amhr2by* gene sequence (Chr09, bottom panel) with the autosomal *amhr2a* gene sequence (Chr04, top panel). The structure of the *amhr2a* gene is depicted with blue boxes (exons, E1 to E12) and blue lines (introns) with arrows indicating the transcription orientation. The solid line on the identity plot (bottom panel) represents 50% nucleotide identity between the two sequences. **D**. ClustalW (Higgins & Sharp, 1988) alignment of Amhr2a and Amhr2by proteins. Identical amino-acids are shaded and the cysteines in the extracellular domain of Amhr2 are shown with bolded black asterisks. Additional cysteines specific to Amhr2by are highlighted in blue. The different domains of the receptor are boxed. **E**. Schematic representation of the two yellow perch Amhr2 proteins showing that the main differences impact the extracellular domain with parts missing in Amhr2by represented in white and the two additional cysteines represented by blue asterisks. **F**. Validation of *amhr2by* sex linkage in yellow perch. Agarose gel electrophoresis of multiplex PCR of *amhr2a* (higher size PCR fragment, 638 bp), and

amhr2by (lower size PCR fragment, 443 bp) in nine females (F, left side) and nine males (M, right side) genomic DNA. The complete analysis on 25 females and 25 males is available as supporting information (Figure S1).

Table 1.

Yellow perch assembly statistics and assembly completeness.

Genome assembly completeness was assessed with BUSCO before (Pre-Hi-C) and after (Post Hi-C) the scaffolding using the chromosome contact map produced with Hi-C. The slightly lower quality of the BUSCO's scores post Hi-C is not due to a local degradation of the quality of our assembly (see text for details).

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	267
Total size of the assembly	877,025,633	877,440,133
Longest fragment	18,280,501	44,580,961
Shortest fragment	160	200
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%)

Table 2.

Yellow perch annotation statistics.

Genome annotation metrics were taken from our own yellow perch genome annotation (current study) and from the National Center for Biotechnology Information (NCBI) genome annotation [NCBI *Perca flavescens* Annotation Release 100 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Perca_flavescens/100/)]. Genome annotation gene completeness was assessed using BUSCO based on the Actinopterygii database.

Gene annotation	Current study	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%)

Table 3.

Comparison of the metrics of the Yellow perch genome assembly (PLFA_1.0) with the Eurasian perch genome assembly (PFLUV_1.0).

Genome annotation metrics were calculated based our own yellow perch genome assembly (PFLA_1.0) and on the published Eurasian perch genome assembly (GCA_003412525.1, PFLUV_1.0).

Assembly metrics	PFLA_1.0	PFLUV_1.0
Number of scaffolds	267	31,105
Total size of the assembly	877,440,133	958,225,486
Longest fragment	44,580,961	29,260,448
Shortest fragment	200	1,000
N50 fragment length	37,412,490	6,260,519
L50 fragment count	11	35
Assembly completeness	**PFLA_1.0**	**PFLUV_1.0**
Complete BUSCOs	4,472 (97.6%)	4178 (91.1%[)
Complete and single-copy BUSCOS	4,363 (95.2%)	4058 (88.5%)
Complete and duplicated BUSCOS	109 (2.4%)	120 (2.6%)
Fragmented BUSCOs	41 (0.9%)	206 (4.5%)
Missing BUSCOs	71 (1.5%)	200 (4.4%)