

Labile sex chromosomes and a novel candidate sex-determination gene in the Australian freshwater fish family Percichthyidae

Alexandra Pavlova¹, Katherine Harrisson², Rustam Turakulov³, Yin Peng Lee⁴, Brett Ingram⁵, Dean Gilligan⁶, Paul Sunnucks¹, and Han Ming Gan⁷

¹Monash University

²La Trobe University

³Australian National University

⁴Deakin University - Geelong Waterfront Campus

⁵Victorian Fisheries Authority

⁶Department of Primary Industries

⁷GeneSEQ Sdn Bhd

July 22, 2021

Abstract

Sex-specific ecology has management implications, but rapid sex-chromosome turnover in fishes hinders development of markers to sex monomorphic species. Here, we use annotated genomes and reduced-representation sequencing data for two Australian percichthyids, the Macquarie perch *Macquaria australasica* and the golden perch *M. ambigua*, and whole genome resequencing data for 50 Macquarie perch of each sex, to detect sex-linked loci, identify a candidate sex-determining gene and develop an affordable sexing assay. In-silico pool-seq tests of 1,492,004 Macquarie perch SNP loci revealed that a 275-Kb scaffold, containing the transcription factor *SOX1b* gene, was enriched for gametologous loci. Within this scaffold, 22 loci were sex-linked in a predominantly XY system, with females being homozygous at all 22, and males being heterozygous at two or more. Seven XY-gametologous loci were within a 146-bp region. Being ~38 Kb upstream of *SOX1b*, it might act as an enhancer controlling *SOX1b* transcription in the bipotential gonad that drives gonad differentiation. A PCR-RFLP sexing assay, targeting one of the Y-linked SNPs, tested in 66 known-sex Macquarie perch and two individuals of each sex of three confamilial species, and amplicon sequencing of 400 bp encompassing the 146-bp region, revealed that the few sex-linked positions differ between species and between Macquarie perch populations. This indicates sex-chromosome lability in Percichthyidae, also supported by non-homologous scaffolds containing sex-linked loci for Macquarie- and golden perches. The resources developed here will facilitate genomic research in Percichthyidae. Sex-linked markers will be useful for determining genetic sex in some populations and studying sex chromosome turnover.

Introduction

Studies of sex-specific differences in ecology, physiology and behaviour enhance our understanding of processes driving evolution and maintenance of biodiversity, and assist in developing conservation management (Amos et al., 2014; Pavlova et al., 2013). However, lack of sex-specific genetic markers hampers such studies in many monomorphic fish species. Sex-determining systems in teleost fish are highly variable, including sequential hermaphroditism, genotypic sex determination (with X0, XY or complex XY being more prevalent than Z0, ZW or complex ZW), environmental dependency and environmental sex determination (Bachtrog et al., 2014; Mank, Promislow, & Avise, 2006). In fishes with genotypic sex determination, sex chromosomes are often homomorphic, where X/Y or Z/W pairs are similar or nearly identical in gene content and size (Bachtrog, 2013; Bachtrog et al., 2014). Various genes (including *amhy*, *amhr2*, *bcar1*, *dmrt1/dmY/dm-*

W, *gsdf*, *sdY* and members of the *SOX: SRY*-like *HMG*-box-containing gene family), single nucleotide polymorphisms (SNPs), inversions, or multiple loci, can all contribute to sex determination in fish (Bao et al., 2019; Graves & Peichel, 2010; Martínez et al., 2014; Natri, Merilä, & Shikano, 2019). Environmental influence on genotypic sex determination in fish (Devlin & Nagahama, 2002; Penman & Piferrer, 2008) results in occurrence of environmental sex reversal, which can trigger transitions between genotypic and environmental sex and facilitate evolution of new genotypic sex-determination systems (Holleley et al., 2015; Schwanz, Georges, Holleley, & Sarre, 2020). Due to rapid and frequent sex chromosome turnover in fish, high variation in sex-determination genes and systems is observed among closely-related species and across populations of the same species (Darolti et al., 2019; Kottler et al., 2020; Natri et al., 2019; Takehana et al., 2014). Thus, identification of reliable, generally applicable sex-linked markers in fish can be challenging.

When present, homomorphic sex chromosomes occur where gametologs have recombined relatively recently, and may result from recent switches in the chromosome pair used for sex determination or transition from temperature-dependent to genotypic sex determination (Bachtrog et al., 2014). In contrast, heteromorphic sex chromosomes, where X and Y—or Z and W—are diverged and highly distinct, evolve from homomorphic sex chromosomes through suppression of recombination between sex chromosomes and subsequent degradation of Y or W, which accumulate mutations and repeats and lose genes that similarly affect the sexes (Charlesworth, Charlesworth, & Marais, 2005). This inhibition of recombination starts from a sex-determining locus and occurs progressively and stepwise along chromosomes, resulting in genomic regions with different levels of sequence divergence between homologs (i.e. evolutionary strata), corresponding to different times since cessation of recombination (Charlesworth et al., 2005).

Several genomic approaches can be used to detect sex chromosomes (reviewed in Palmer, Rogers, Dean, & Wright, 2019). Three are particularly applicable to wildlife, because they use DNA sequence obtainable from non-lethally-collected samples and do not require captive breeding to build linkage maps. The first approach uses different *ploidy* of sex chromosomes diverged through Y or W degeneration: the homogametic sex (e.g. XX females or ZZ males) will have two copies of the same sex chromosome, whereas heterogametic sex (e.g. XY males or ZW females) will have one copy of each sex chromosome. This pattern will be reflected in read-depth coverage: for example, old strata in which recombination between homologs has long ceased will show half as many reads on X or Z in the heterogametic sex as the homogametic one, and Y- or W- loci will be absent in the homogametic sex (Darolti et al., 2019; Gan et al., 2019; Vicoso, Emerson, Zektser, Mahajan, & Bachtrog, 2013). But in young strata or homomorphic sex chromosomes, having high similarity between X and Y or Z and W, the read depth will be similar between sexes and to that of autosomal regions. For these younger regions of sex chromosomes, an approach based on differences in *sex-specific SNP density* across genomic regions is more appropriate (Darolti et al., 2019; Wright et al., 2017). With Y or W accumulating mutations faster than X or Z due to reduced recombination and weaker purifying selection, higher SNP density in young strata is expected in the heterogametic than homogametic sex. In contrast, in older strata with substantial Y or W degeneration, X- and Z-linked loci will be effectively hemizygous in the heterogametic sex (halved read depth), and higher SNP density is expected in the homogametic sex. Finally, older strata of Y or W chromosome could be detected using a k-mer approach, by matching short reads of the homogametic sex to the genome of the heterogametic sex (Carvalho & Clark, 2013), or using *in-silico* whole genome subtraction (Dissanayake et al., 2020). A combination of different approaches might be needed to detect sex-linked loci when the age of sex chromosomes is unknown.

Sex determination is not well understood in fish of the family Percichthyidae. Species from this family dominate the Australian freshwater fish fauna, and three genera occur in eastern Asia (*Coreoperca* and *Siniperca*), and South America (*Percichthys*). Many Australian species are threatened, including the endangered Macquarie perch *Macquaria australasica* and trout cod *Maccullochella macquariensis*. Although sex of an adult can be ascribed when it produces gametes, lack of sexual dimorphism and ways to determine the sex of individuals non-invasively outside breeding seasons hinders better understanding of species biology and more efficient conservation. Shams et al. (2019) examined karyotypes for two percichthyids, Murray cod *Maccullochella peelii* and golden perch *Macquaria ambigua*, and for both reported male heterogametic sex-chromosome systems (XX females/XY males) with diploid chromosome number $2n = 48$. Heteromorphism

was detected in sex chromosomes of Murray cod, but homomorphism in golden perch (Shams et al., 2019). In Macquarie perch, the sister-species of golden perch (Lavoué et al., 2014), a set of >1200 genome-wide SNPs did not reveal markers consistent with Y-linkage (i.e. present in males only) or strict X/Y homology, i.e. always heterozygous in males and always homozygous in females (Lutz et al., 2021), suggesting that it may also have young sex chromosomes. Infrequent (0.17-3.7%) synchronous hermaphroditism was reported in captive and wild Murray cod (Gooley, Anderson, & Appleford, 1995; Ingram, Ho, Turchini, & Holland, 2012) but not in Macquarie perch (Appleford, Anderson, & Gooley, 1998). There are no reports of sex change in Percichthyidae. Hatchery work suggests that environmental conditions may influence sex determination at an early stage of development in some percichthyids (Ingram et al., 2012). Lyon et al. (2012) reported 2.5 times as many females as males in a population of trout cod stocked from hatchery-bred fish. Because sex-ratio biases in stocked fish have implications for recovery of populations through stocking, being able to determine the genetic sex of fish, and the extent to which it can be overridden by environmental variation, would benefit conservation management of threatened percichthyid species.

With the overarching goal of uncovering the genetic basis of sex determination in two percichthyid perches, we aimed to (i) sequence and assemble the genomes of Macquarie perch and golden perch, (ii) annotate both genomes using respective transcriptome sequences, (iii) identify sex-linked loci using reduced-representation sequencing (DARtseq) data for both species and whole-genome resequencing (WGS) data for 100 known-sex Macquarie perch, (iv) identify candidate sex-determining genes using annotations, (v) develop a non-lethal, affordable and rapid molecular assay to determine the genetic sex of Macquarie perch and test its validity across four percichthyids (Macquarie and golden perches, Murray and trout cods), and (vi) test whether the candidate sex-determining region of Macquarie perch contains sex-specific variation across the four percichthyids using amplicon sequencing. Whereas the reduced-representation approach yielded few sex-linked loci, WGS data for Macquarie perch data revealed a small genomic region inherited in a predominantly XY fashion on a *SOX1b* -containing scaffold. This suggests a XY sex-determining system in Macquarie perch, with *SOX1b* as a novel candidate sex-determining gene, and the sex-linked region as its candidate regulatory locus. A test of a molecular sexing assay targeting a SNP with a male-specific allele in the sex-linked region, and amplicon sequencing data for four percichthyid species indicated that the Macquarie perch sexing region is species-specific and either specific to populations related to those in which it was detected or can be influenced by environment. The resources developed here will facilitate evolutionary research on sex chromosome turnover in fish, as well as conservation and ecological research in Percichthyidae, a dominant component of the freshwater Australian fish fauna. The workflow described here could be used for developing molecular tests to determine genetic sex of other fish species with monomorphic sex chromosomes.

Materials and Methods

Sample collection, sex identification

We used samples of 261 Macquarie perch, 68 golden perch, 4 Murray cod and 4 trout cod (Supplementary Material S1). Of 261 Macquarie perch individuals, 173 were Snobs Creek hatchery broodstock from Yarra and Dartmouth populations (see Lutz et al., 2021 for details of the breeding program), 25 were Narrandera hatchery broodstock from the Abercrombie River, one was a Snobs Creek hatchery-produced juvenile, and the remaining 62 were adults sampled from four inland and five coastal populations. The inland lineage (including Yarra, Dartmouth and Abercrombie populations) diverged from the coastal one in the Pleistocene; the two are morphologically distinguishable (Pavlova et al., 2017). Sex was identified for hatchery broodstock and 21 sampled-and-released fish from inland King Parrot Creek and Holland's Creek by observation of mature gametes during the breeding season. For all but two individuals (exceptions in next section), fin-clip samples were collected non-lethally and preserved in absolute ethanol at -20°C.

Of the 68 golden perch, one was an adult of unknown sex sampled non-lethally, one an adult from another population sampled lethally, and 66 were adults from three more populations sampled lethally for another project. Sex of lethally-sampled fish was identified during dissection based on presence of maturing or mature gonads. Fin clips from two males and two females were also used for each of Murray cod and trout cod, with their sex determined at Snobs Creek hatchery (details of sampling and sexing in Supplementary Material

S1). All procedures employed were approved by relevant animal care and ethics committees. Field sample collection was approved by NSW ACEC 14/10, DELWP AEC 14/04 and AEC 15/02, Scientific Collection Permit P01/0059(A)-3.0 and Victorian Fisheries Research Permit RP827. Hatchery work and sampling were conducted under approval of VIC DPI Fish AEC Jul09 0067 and NSW Fisheries ACEC committee 05/06.

Sequencing and assembling Macquarie perch and golden perch genomes

For Macquarie perch, we used fin, tail bone and muscle tissues of a 2-month-old hatchery-produced juvenile of unknown sex of Yarra River origin, born in November 2012 (sample ID MP_SCH12). For the golden perch genome, we used fin tissue collected non-lethally from adult of unknown sex (aged as 3+ years based on size), captured in the Broken River, Victoria, in May 2017 (sample ID GOP001). For both species, DNA samples were preserved in ethanol and kept at -20°C.

DNA was extracted using Qiagen DNeasy Blood & Tissue kits. For short-read sequencing, 100 ng of gDNA was fragmented to 350 bp using QSonica and processed with a NEB Ultra Illumina Library Preparation Kit. The libraries were pooled with libraries for other projects and sequenced on all four lanes of S4 flowcell of a Novaseq6000 at the Deakin Genomics Centre using 2×151 bp run configuration, with the aim of obtaining 100 Gb of data per sample (Appendix A). To obtain long-read data, 1 µg of gDNA was fragmented to 8 kb using a Covaris G-Tube and processed with a LSK108 library preparation kit according to the manufacturer's instructions (Oxford Nanopore, UK). The library was subsequently sequenced on a Nanopore R9.4 flowcell. Base-calling of the Nanopore signal used Albacore v2.0.1 (Oxford Nanopore, UK).

Illumina reads, adapter-trimmed using fastp v0.19.5 (Chen, Zhou, Chen, & Gu, 2018), and Nanopore long reads were hybrid-assembled *de novo* using MaSuRCA v3.2.4 (Zimin et al., 2017). The short Illumina reads were first error-corrected with QuORUM as implemented in the MaSuRCA pipeline and subsequently used to construct contigs by the de Bruijn graph approach. These contigs were used to error-correct the Nanopore long reads, generating “mega read” contigs for Overlap-Layout-Consensus assembly. Genome completeness was assessed using BUSCO v4 (Seppey, Manni, & Zdobnov, 2019) with default setting, based on the actinopterygii_odb10 database.

Transcriptome sequencing and annotation of Macquarie perch and golden perch genomes

To facilitate genome annotations, we performed mRNA sequencing for both species. We used an adult female Macquarie perch (sample ID MP_527), captured November 2017 from Lake Dartmouth. Samples from liver, ovary, brain, kidney and muscle were collected. We used an adult golden perch (sample ID GPTT01, aged as 3 years by otolith analysis), sexed in the field as putatively male, captured in the Ovens River, Victoria, in April 2018. Samples of brain, gills, heart, gonads, kidney, liver and cheek muscle were collected. Immediately after the fish were humanely killed, RNA samples were collected and preserved in DNA/RNA Shield (Zymo Research) and stored at -80°C (Macquarie perch) or -20°C (golden perch). Total RNA was extracted from individual tissue samples using Quick-RNA Kits (Zymo Research), quantified using a TapeStation (Agilent) and 440 ng pooled per tissue. This was enriched for mRNA via poly-T beads using NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB). The enriched mRNA was processed using Universal Plus mRNA-Seq Library Preparation Kits. The libraries were pooled with libraries for other projects and sequenced on one of four lanes of S4 flowcell of an Illumina NovaSeq6000, with the aim of obtaining 20 Gb of data per sample.

A repeat library was constructed *de novo* for the assembled genome with RepeatModeler2 (Flynn et al., 2020), and used to repeat-mask (soft-mask) the genome with RepeatMasker v 4.0.9 (Smit, Hubley, & Green, 2013-2015). Transcriptome reads were aligned to the repeat-masked genome using STAR v2.7.1a (Dobin et al., 2013). The transcriptome alignment (single-species bam file) and repeat-masked genome were used as the input for protein-coding gene prediction in BRAKER v2.1.2 (Bruna, Hoff, Stanke, Lomsadze, & Borodovsky, 2020). Functional annotation of the predicted proteomes was completed using InterProScan 5 (Jones et al., 2014).

Identifying sex-linked DArT SNPs for Macquarie perch and golden perch

We used DArTseq (Kilian et al., 2012), a reduced-representation sequencing method, to genotype 93 female and 78 male Macquarie perch from the Dartmouth and Yarra populations, and 41 female and 25 male golden perch from Macquarie, Murray and Murrumbidgee populations (samples in Supplementary Material S1). DArTseq approach is similar to double-digest restriction-associated sequencing, except it ensures high SNP quality by including ~25% of technical replicates from independent libraries in each sequencing lane and rejects SNPs with low reproducibility between technical replicates. Sequencing libraries were prepared at Diversity Arrays Technology Pty Ltd (Canberra, Australia) following Kilian et al. (2012). DNA samples were digested using a combination of restriction enzymes *Pst* I and *Sph* I that target low-copy genomic regions. Pooled libraries were sequenced using single-read technology on an Illumina HiSeq2500 (94 samples per lane; details in Appendix A). SNP discovery and genotyping were performed using DArT P/L's proprietary analytical pipeline (detailed in Nguyen, Premachandra, Kilian, & Knibb, 2018). Briefly, the primary pipeline removed poor-quality sequences, applying more stringent criteria to the barcode region than the rest of the sequence and corrected low-quality bases from singleton tags using collapsed tags with multiple members as a template. Then the secondary pipeline (DArTsoft14) parsed clusters, comprising 69-bp sequenced tags differing by no more than 3 bases, into separate SNP loci, while ensuring the balance of read counts for the allelic pairs: loci with a 5-fold or higher difference in read counts for each allele were rejected. SNPs with a reproducibility <95% were removed, but no other filtering was performed. DArT loci for each perch species were aligned to their respective newly-assembled reference genomes (Table 1) using BLAST, with e-value [?] $5e-5$ and sequence identity [?] 90% .

We tested each DArT SNP locus for belonging to one of four types: (i) *Y-linked* : homozygous in males, and absent in females; (ii) *XY-gametologs*: homozygous in females, heterozygous in all males; (iii) *loci with male-specific allele* : homozygous in females and heterozygous in >10% of males; and (iv) *loci with female-specific allele* : heterozygous in >10% of females and homozygous in males. The tests were performed using the *gl.sexlinkage* function of the dartR package (Gruber, Unmack, Berry, & Georges, 2018) in R (R Core Team, 2020) (details in Appendix B). To reduce the number of false positives due to small sample size, only loci successfully scored in >75% (>58) of male Macquarie perch and >95% (>23) male golden perch were considered, male sample sizes being smaller than for females in both species.

Identifying sex-linked loci using whole genome resequencing data for Macquarie perch

Whole genome resequencing data (WGS) was obtained for 100 known-sex Macquarie perch from Dartmouth and Yarra (25 individuals of each sex per population, all of which were also DArT-sequenced). DNA was extracted and an Illumina Novaseq library was constructed for each individual, as explained above. Libraries were sequenced at Deakin Genome Centre to at least 10X depth (>7 Gb of data per library; list of samples and individual read-depth coverage in Supplementary Material S1).

First, we searched the Macquarie perch genome for loci with sex-specific alleles (*ii-iv* above) using four *in-silico* pools: Dartmouth females, Dartmouth males, Yarra females, and Yarra males (25 individuals of each sex per population). Each pool combined 25 million paired-end reads randomly subsampled from each individual's WGS data. Reads from the four pools were mapped onto the newly-assembled Macquarie perch reference genome (Table 1) using bwa-mem v0.7.17 (Li & Durbin, 2009) followed by filtering for mapping quality of >20 using samtools (Li et al., 2009) (samtools view -q 20). A Cochran-Mantel-Haenszel (CMH) test implemented in *PoPoolation2* (Kofler, Pandey, & Schlotterer, 2011) was used to detect significant and consistent allele-frequency differences (based on read counts) between sexes in two biological replicates. After examining the genomic spread of loci with sex-specific alleles detected under $p < 1e-5$ significance threshold, we tested whether loci highly significantly ($p < 1e-20$) differentiated between sexes clustered in the same genomic region.

One Macquarie perch scaffold contained a region enriched with loci highly differentiated between sexes, and using individual WGS genotypes we tested it for presence of XY-gametologs that could act as sex-determining loci. WGS genotypes were obtained by aligning the trimmed paired-end reads for each individual to the reference genome using bwa-mem v0.7.17-r1188, followed by variant-calling on the merged BAM files from all individuals, using strelka v2.9.10 (Kim et al., 2018). Individual alignments for the putative sexing region

in males were visualized in Tablet (Milne et al., 2012) to investigate whether the mate-pair architecture of the sequenced libraries resolved female- and male-specific haplotypes (similar approach in Kijas et al., 2018). We then investigated read-depth coverage for each base of the sexing region, to test for sex-difference in ploidy. Read depth was collected for each individual using the *CollectAllelicCounts* function of GATK v4.1.0 (DePristo et al., 2011) from individual alignment files, normalized by that individual’s genome-wide average read depth (the number of all mapped reads multiplied by 151 bp (read length) and divided by genome size), and averaged across the four sex-by-population samples. We also calculated read depth for reference and alternative SNP alleles of the sexing region separately.

Identifying candidate sex-determining gene

To identify candidate sex-determining genes for both perches, we tested whether the scaffolds containing the sex-determining genes known from teleost fish also harbour sex-linked DArT SNPs or the Macquarie perch sexing region detected using WGS data. We searched our InterProScan5 output for *amh* , anti-Mullerian hormone, *amhr*, anti-Mullerian hormone receptor, *bcar1* , breast cancer anti-resistance, *brca2* , breast cancer type 2 susceptibility protein, *dmrt1* , *doublesex*- and *mab-3* -related transcription factor 1, *dmY*, *dm-W*, *dmW*, *gsdf*, gonadal soma derived growth factor, *sdY* and *SOX* (Bao et al., 2019; Graves, 2013; Rodriguez-Mari et al., 2011) and Gene Ontology term GO:0007530 (biological process: sex determination) and its child terms (<https://www.ebi.ac.uk/QuickGO/term/GO:0007530>). To confirm gene identity of the candidate sex-determining genes, we searched for similar sequences using Blastn (Zhang, Schwartz, Wagner, & Miller, 2000) in NCBI nucleotide database, retaining hits with annotations, an e-value $<10^{-10}$, query coverage $>99\%$ and percent identity $>90\%$. Finally, we tested the Macquarie perch sexing region for two genomic signatures of a cis-regulatory element (i.e. a DNA sequence that regulates tissue- and time-specific candidate gene expression): evolutionary conservation of non-coding regions, and the presence of multiple transcription factor binding sites (i.e. ~ 10 bp-long degenerate sequences recognized by transcription factors) (Hardison & Taylor, 2012). We queried the 600-bp region containing the sexing region (GenBank accession SEMN01000633, bases 93001-93600) against the NCBI nucleotide database (using Blastn, retaining all hits with e-value $<10^{-10}$). We also used CiiDER (Gearing et al., 2019) and the JASPAR database of transcription factor binding profiles (Fornes et al., 2020) to search for potential binding-site locations in ~ 246 -bp-long X- and Y- specific sequences (haplotypes) encompassing the sexing region. We investigated whether some transcription factor binding sites are specific to Macquarie perch X- or Y- haplotypes and if so, whether they are present in golden perch and Murray cod sequence (outgroups). A potential gain of a binding site on Y-haplotype was inferred if the site was present on Y- but absent on X-haplotype and both outgroups, whereas a potential loss of a binding site was inferred if it was absent on Y-haplotype but present on X-haplotype and in both outgroups.

Developing PCR-RFLP molecular sexing assay for Macquarie perch and testing it in four percichthyids

To develop the rapid and affordable PCR-RFLP (polymerase chain reaction - restriction fragment length polymorphism) molecular sexing assay, we targeted a male-specific allele of an *XY-gametologous* SNP in the Macquarie perch sexing region. Conserved PCR primers (F1, F2, R2 and R1; Table C1) were designed based on the Geneious Pro alignment of the 600-bp region of Macquarie perch scaffold 633 containing the sexing region (above) and the homologous region of the golden perch genome (scaffold VMKM01003747 bases 16365-16988; Appendix C, Figure C1). SMS online tool (https://www.bioinformatics.org/sms2/rest_map.html) was used to select *Ase* I restriction endonuclease, which recognizes a male-specific sequence ATTAAT and cuts it at Y-linked SNP 93229. After digestion of PCR products with *Ase* I, this assay was expected to yield, during visualization on an agarose gel, one long band (of length dependent on primer combination) for XX females (always homozygous) and three bands (including a long X band and two shorter Y ones cut by the enzyme) for XY males. After initial tests on two individuals of each sex for four percichthyids, the final assay using F1-R1 primers was tested on 45 hatchery-sexed Macquarie perch from Dartmouth, Yarra and Abercrombie, 21 field-sexed Macquarie perch from King Parrot Creek and Holland’s Creek, and 41 Macquarie perch of unknown sex from eight other populations (two inland and six coastal; Appendix C; list of samples in Supplementary Material S1).

Testing for sex-specific variation at the Macquarie perch sex-linked region in other percichthyids using amplicon-based sequencing

To investigate whether the Macquarie perch sexing region has sex-linked variation in three other species of percichthyids, we amplified and sequenced ~400 bp encompassing the region, using primers conserved across three percichthyid genomes: Macquarie perch, golden perch and Murray cod (primers F2 and R3, Table C1). Amplicon sequencing was performed on an Illumina ISeq100 at GeneSEQ Sdn Bhd (Malaysia) using the unidirectional read configuration of 1 x 300 bp (details in Appendix A). Two males and two females of each of the four species were sequenced, along with two coastal Macquarie perch individuals of unknown sex but appearing to be one of each sex based on the PCR-RFLP sexing assay (list of samples in Supplementary Material S1). Reads in the fastq files were trimmed of poly-G sequences (a sequencing artefact in the newer NextSeq500/NovaSeq6000/ISeq100 Illumina system) using *cutadapt* (Martin, 2011) via the option `-nextseq-trim`. Each individual was then genotyped for the 400-bp amplicon region using containerized best-practice GATK v4.1.0 protocol (Van der Auwera et al., 2013), available on the docker repository under tag: `trust1/gatk:version0.3.1`. The ~600 bp region used for primer design was also used as a reference (LKNJ01002546 bases 82756- 83361 for Murray cod and trout cod, VMKM01003747 bases 16365-16988 for golden perch and scaffold 633 bases 93001-93600 for Macquarie perch).

Results

Macquarie perch and golden perch genome assemblies

The Macquarie perch genome assembly had a size of 676 Mb comprising 2,962 scaffolds (N50=845 Kb; summary in Table 1). The golden perch assembly of 661 Mb was less contiguous, comprising 7,165 scaffolds (N50=252 Kb; Table 1).

DArT-derived sex-linked SNPs for Macquarie perch and golden perch

Despite more biallelic SNP loci being yielded by DArTseq for Macquarie perch (6,939) compared to golden perch (5,997), Macquarie perch had an order of magnitude fewer sex-linked loci (4 vs 48) with lower proportion of individuals sharing a sex-specific allele (20% vs 30%; Table 2). Although this could just reflect lower genome-wide diversity of Macquarie perch, it may also suggest that sex chromosomes in Macquarie perch are younger than those in golden perch. Both species had loci with male-specific and female-specific alleles, with similar number of both types in Macquarie perch (2 and 2) but more female-specific alleles in golden perch (14 and 34, respectively). All sex-linked DArT SNPs were on different scaffolds, except two golden perch SNPs were on the same scaffold (Supplementary Material S2). Neither dataset revealed loci consistent with Y- linkage or XY- (or ZW-) gametologs, suggesting that sex chromosomes in both perch species evolved recently.

Tests for sex-linked loci in Macquarie perch WGS-derived SNPs using population-by-sex pools

Cochran-Mantel-Haenszel (CMH) tests applied to four *in-silico* Macquarie perch population-by-sex pools tested 1,492,004 SNP loci, of which 4.1% were differentiated $p < 1e-5$ between male and female pools (Supplementary Material S3) and 0.25% had sex-specific alleles in both populations, with a similar number of male- and female-specific alleles (Table 2). The 274,905-bp long scaffold 633 had the most loci with male-specific alleles (40 of 1831), two thirds of which clustered within a ~6.2 Kb region (87120-93327; Appendix B; Table B2). Of 30 SNP loci most strongly differentiated between male and female pools (CMH $p < 1e-20$), nine were loci with male-specific alleles located within a ~3.8-Kb region 90152-94017 of scaffold 633 (Table B1). Of these nine loci, five clustered within a short 146-bp fragment (93182-93327; hereafter, '*the 146-bp sexing region*'). Fifteen additional scaffolds had smaller clusters of 3-6 loci with male-specific alleles (CMH $p < 1e-5$), and eight scaffolds had clusters of 3-4 loci with female-specific alleles (Appendix B, Table B3).

Individual Macquarie perch genotypes for sex-linked WGS loci on scaffold 633

Examination of individual genotypes for 50 females and 50 males for the 5 Kb region of scaffold 633 (90000-95000; Supplementary Material S4) revealed 22 loci with male-specific alleles, including eight XY gametol-

ogous ones. All females were homozygous at the X-linked alleles and all males were heterozygous at two or more of the 22 loci (Fig. 1). For the first 13 loci (region 90152-92914), 16% of males were homozygous, whilst the remaining males were heterozygous for two (52%), ten (14%) or four loci (2%). Alternative alleles for these loci were present in four distinct linkage patterns, suggesting that multiple Y- haplotypes are segregating in Macquarie perch (Fig. 1). The next seven loci (93182-93327) comprised the 146-bp sexing region, which had a male-specific deletion and insertion in addition to five SNPs detected with the CMH test, all inherited in a fashion consistent with XY-gametology (Table 3). All females were homozygous for X-alleles and 35/50 males were heterozygous for these and Y-alleles at all seven loci (Fig. 1). Of the remaining males, 14 were either homozygous for the Y-allele or heterozygous at most of these loci, and one (MP_CBR72) was homozygous for the X-allele for all seven loci. The 146-bp sexing region was followed by a SNP with male-specific allele (93879; numbered 21 on Fig. 1) occurring in 37 heterozygous and one homozygous males, and by another XY-gametologous SNP (94017; numbered 22) for which all males were heterozygous, except one homozygous for the Y-allele. Notably, the male MP_CBR72 noted above as displaying a female-like genotype for the 146-bp sexing region, was heterozygous at this locus, as were most other males, suggesting it was not simply a mis-sexed individual or sample mixed-up. The pattern of male heterozygosity for this region suggests a working hypothesis of sex determination in which presence of Y-alleles at one or more of these 22 loci triggers the development of a male phenotype. A sexing assay screening all these 22 loci is expected to achieve ~100% reliability in the Dartmouth and Yarra populations. Occurrence of homozygotes for X- alleles for up to seven of eight XY-gametologous loci in four (of 50) males suggests rare past or ongoing recombination between X and Y haplotypes.

Visual examination of mapped mate-pair reads for the 146-bp sexing region in males confirmed physical linkage of male-specific SNPs (a Y-haplotype). It also revealed a strong drop of read depth compared to the genome average in this region, in both sexes, including zero coverage around position 93400 for 59% of individuals regardless of their overall sequence depth (Fig. 2). This lower depth might be linked to difficulty of sequencing through a low-complexity region, including the (GT)_n microsatellite that ends around 93193 (Figure C1). The drop in depth did not result from failed mapping due to repetitive regions, because neither the 146-bp sexing region nor the 41-bp region 93380-93420 spanning the lowest depth were highly similar to any other region in the Macquarie perch genome. Despite this drop in depth, comparison of male and female read depth for loci with male-specific alleles suggested presence of segregating deletions on X chromosomes. In region 90152-92914 average depth was the same in both sexes, approaching genome-wide average (not shown), but sex-differences were apparent in the region containing the 146-bp sexing region (Fig. 2). Average depth for females was consistently lower than that for males from ~93250 to ~93600, containing the majority of the 146-bp sexing region, suggesting that some or most females could be hemizygous for this region (through deletion on one of the X chromosomes, thus the region being X0). Lower depth of reference alleles compared to alternative alleles in males for sex-linked SNPs 93229, 93299, 93315 and 93327 supports the region being haploid in some males too (i.e. Y0; Fig. 3). In contrast, the SNP locus 93879 with a male-specific allele was associated with a drop of male depth compared to that of females (Fig. 2), with male read depth for the alternative allele being just a half of that for the reference allele, suggesting deletion of the Y-allele in some males (Fig. 3). The read depth for the last XY-gametologous SNP locus in this region, 94017, was close to the genome-wide average for both sexes.

Proximity of the Macquarie perch and golden perch sex-linked SNPs to known sex-determining genes

Forty-four Macquarie perch genes (Appendix D, Table D1) and 45 golden perch genes (Table D2) found by BRAKER2 were annotated to genes or gene families that act as sex-determining genes in other fish species. Of these, one Macquarie perch gene (g10286 on scaffold 633, bases 131251-132300) and three golden perch genes (g1522, g23888 and g30513; Table D5), all annotated to the *SOX* gene family, were on the scaffolds containing sex-linked DArT SNPs and the 146-bp Macquarie perch sexing region. Three of these *SOX* genes- Macquarie perch gene g10286 and golden perch genes g1522 and g23888- were annotated to a function of regulating DNA-templated transcription (GO:0006355) and had best Blastn hits to *SOX1b*, *SOX1a* and *SOX14* of other bony fishes, respectively (Appendix E). This suggests a role of the *SOX* transcription factor gene family regulating the fate of stem and progenitor cells during development (Kiefer, 2007), in sex

determination of both perch species, with *SOX1* gene being the strongest candidate.

Potential role of the Macquarie perch 146-bp sexing region in regulation of the candidate sex-determining gene SOX1b

The 146-bp Macquarie perch sexing region was not annotated to any of the eight genes on scaffold 633 (Table D3), but was located ~38 Kb upstream of the candidate Macquarie perch sex-determining gene *SOX1b* (g10286). This intergenic region had both genomic characteristics of a cis-regulatory region: evolutionary conservation, and presence of putative transcription factor binding sites (Hardison & Taylor, 2012). Despite being a non-coding DNA, the 600-bp fragment containing the sexing region aligned to unannotated genomic regions of over a dozen other bony fishes from five taxonomic orders (Appendix E). In contrast, a 200-bp section of scaffold 633 containing the other XY gametologous SNP (94017) did not align to fish other than two percichthyids- golden perch and Murray cod. Transcription factor binding motifs for 145 transcription factors were predicted within 246 base pairs containing the 146-bp sexing region (Appendix F). Of these, 14 transcription factors had binding motifs present on Y-haplotype, but not on X-haplotype nor golden perch nor Murray cod sequences, implying potential gain of the binding sites due to mutations on Y-haplotypes (Table F1). In addition, binding motifs for three transcription factors were absent on Y-haplotype, but present on X-haplotype and the two outgroups, implying potential loss of the binding sites on Y-haplotype. The latter included *SOX14* - a negative regulator of transcription, involved in the regulation of embryonic development and in the determination of cell fate.

The 146-bp sexing region was also ~41 Kb downstream of gene g10285 that encodes for a protein of the *Dbl* homology (DH) domain superfamily, associated with guanyl-nucleotide exchange factor activity (GO:0005085) and protein-binding function (interacting selectively with any protein or protein complex; GO:0005515; Appendix D, Table D3). The scaffold 633 sex-linked DArT SNP was located within intron 5 of this gene.

Molecular sexing in Macquarie perch using PCR-RFLP

The sexing assay (PCR product bounded by F1-R1 primers digested with *Ase* I at the Y-specific scaffold 633 allele T at base 93229) tested in two males and two females with known genotypes showed a single band in females, and three bands in males, as expected (Appendix C; Figure C2). A larger sample of 10 females and 10 males from Dartmouth/Yarra also showed a single band in females, but only two short cut fragments of Y in males, consistent with Y0 genotype (Figure C3). The lack of the uncut X-specific band in five of these males was unexpected, as their WGS-based genotypes showed them as heterozygotes for the cut-site. The same assay tested in known-sex Macquarie perch from King Parrot Creek/Holland's Creek and Abercrombie populations yielded 2- or 3-band pattern in all males, but also in 20% of King Parrot Creek females and 60% of Abercrombie ones; the remaining females showed a single band, as expected (Figure C4). Because phenotypic females with a male-specific allele could potentially represent sex-reversed genotypic males (Baroiller, D'Cotta, & Saillant, 2009), we explored fertilization rates data for ten phenotypic Abercrombie females that yielded bands on the PCR-RFLP assay. Females with a male-specific allele (i.e. potentially feminized males) appear to have had eggs with lower fertilization rate (Supplementary Material 1). Of four females that produced a single female-like band, two produced eggs that were successfully fertilized in the hatchery, and two had eggs that failed to be fertilized. Of six females that produced two bands (male-like), four produced eggs that were not fertilized, one showed an egg-fertility rate of <1%, and one had eggs that were successfully fertilized. Further tests of the sexing assay in Macquarie perch of unknown sex from six additional populations yielded a single band in 34% of individuals, and two or three bands in 66% of individuals (Figure C5). Therefore, the PCR-RFLP sexing assay targeting Y-allele at 93299 is ~100% accurate only for the populations used in assay development, Dartmouth and Yarra (except it could missex ~2% of males as females, as one male out of 50 showed a female-like genotype; Fig. 1), but is not expected to work well in other populations of Macquarie perch due to presence of the Y-allele in phenotypic females.

Sex-specific sequence polymorphism in Macquarie perch, but not its three relatives, revealed by amplicon sequencing

For Macquarie perch, genotypes for the ~400-bp region of scaffold 633 captured by amplicon sequencing using F2-R3 primers were consistent with WGS genotypes for the two females and two males from Dartmouth (Appendix G, Table G1). Amplicon sequences for the two individuals from coastal populations (Table G1) showed that only variation at 93229 (the site chosen for the PCR-RFLP sexing assay) might be consistently sex-linked across sequenced individuals. Unique variation at the 146-bp sexing region in two coastal individuals of unknown sex further suggests that sex-linked variation in Macquarie perch could be population-specific. Genotypes for the amplicon-sequenced fragment homologous to the Macquarie perch sex-linked region did not reveal clear sex-linked variation in golden perch (Table G2), Murray cod or trout cod (Table G3).

Discussion

Percichthyids as a novel model for studying sex chromosome turnover

Sex-determining mechanisms are extremely diverse and may vary even in closely-related species or among populations of the same species (Kijas et al., 2018; Lubieniecki et al., 2015; Rodrigues, Merila, Patrelle, & Perin, 2014). In species with genotypic sex determination, an ancestral sex-determining locus can be replaced by another locus on the same or a different chromosome, resulting in chromosome turnover. Transitions to a different chromosome are more likely when sex chromosomes exhibit little differentiation, so that WW or YY genotypes are less likely to be lethal (Bachtrog et al., 2014). Therefore, species with homomorphic chromosomes and intraspecific variation of sex-determining loci provide a window into sex chromosome evolution in real time. Macquarie perch, with an order of magnitude fewer sex-linked loci than its sister-species golden perch, appears to have more recently-evolved sex chromosomes. Its XY-gametologous locus, putatively sex-determining in at least the two populations used in assay development, was less correlated with phenotypic sex in other populations and not sex-linked in other percichthyids. Whereas population-specific variation in genetic factors determining phenotypic sex has been observed in other fish species (Faggion, Vandeputte, Chatain, Gagnaire, & Allal, 2019; Ferrareso et al., 2021), it is also possible that genotypic sex determination in different Macquarie perch populations is variably influenced by environmental variation (Hill, Burrige, Ezaz, & Wapstra, 2018). In contrast to sex chromosomes of perches, heteromorphic chromosomes of the Murray cod (Shams et al., 2019) must be the oldest among the three percichthyids with sequenced genomes. The interspecific variation in candidate sex-determining mechanisms makes Percichthyidae a good system for studying sex chromosome turnover and dosage compensation. Comparative studies that demonstrate sex chromosome turnover within the same genus or families are still relatively rare (e.g. *Gambusia* (Kottler et al., 2020), *Oreochromis* (Tao et al., 2021), *Oryzias* (Tanaka, Takehana, Naruse, Hamaguchi, & Sakaizumi, 2007), Poeciliidae (Darolti et al., 2019; Schultheis, Bohne, Schartl, Volf, & Galiana-Arnoux, 2009), Gasterosteidae (Ross, Urton, Boland, Shapiro, & Peichel, 2009)).

SOX1b as a novel candidate sex-determination gene

Sex chromosome turnovers are triggered by a different gene becoming the master sex determiner. A gene can become sex-determining if it has alleles that are beneficial to one sex but harmful to the other, if it is physically linked to such a gene with sexually-antagonistic variation, or if a new mutation changes the timing and/or tissue of its expression (Bachtrog et al., 2014; Graves, 2013). Some gene families repeatedly take on the role of a sex-determining gene in diverse independently evolved systems, and genes of the *SOX* family are often recruited (Graves, 2013). *SOX* genes govern major vertebrate developmental mechanisms, including testis development, central nervous system neurogenesis, oligodendrocyte development, chondrogenesis, and neural crest cell development, and are thus expressed in many tissues (Kiefer, 2007; Sarkar & Hochedlinger, 2013). Three genes in particular, *SOX9*, *SOX8* (Koopman, 2005) and *SOX3*, a progenitor of the *SRY* gene in mammals (Graves, 2013; Takehana et al., 2014), have a role in male sex determination in unrelated species, their expression in the early gonad inducing testes development. In mammals, brief expression of *SRY* in a small number of undifferentiated gonadal cells stimulates differentiation of Sertoli cells and up-regulates *SOX9* expression, triggering testis development (Koopman, 2005). *SRY* mediates *SOX9* upregulation by binding to several *SOX9* enhancers, some of which cause XX or XY sex reversal in humans when duplicated or deleted, respectively (Croft et al., 2018). In the Indian ricefish *Oryzias dancena*, a Y-chromosome-specific cis-regulatory element of *SOX3* acts as a sex-determining locus, driving early gonadal *SOX3* expression in

XY individuals, which initiates testicular differentiation by upregulating expression of *gsdf* gene (Takehana et al., 2014). In medaka *Oryzias latipes*, *SOX5* has a dual role in sex-determination: it acts as an evolutionarily conserved regulator of germ-cell number, and as a de novo regulator of expression of another transcription factor, *dmrt1*, during primary sex determination (Schartl et al., 2018). Several other *SOX* genes also play important roles in sex differentiation in fish (Hu, Wang, & Du, 2021).

Although *SOX1* has not been reported to play a role in sex determination or sex differentiation, it belongs to the same B1 subfamily of transcription activators as *SOX3*, and thus could have acquired a novel sex-determining function in Percichthyidae. *SOX1a* and *SOX1b* are fish-specific co-orthologs of tetrapod *SOX1*, produced as a result of a teleost-specific genome duplication early in fish evolution, enabling diversification of gene function (Taylor, Braasch, Frickey, Meyer, & Van de Peer, 2003; Zhang et al., 2018). We hypothesize that in Macquarie perch, gonadal development is controlled by male-specific expression of *SOX1b* gene in a bipotential gonad during embryo development, modulated by the variation in a distal enhancer (the 146-bp sexing region). A somewhat similar mechanism of switching the master sex-determining gene exists in *Oryzias luzonensis*, where a mutation in a cis-regulatory region of the sex-determining gene *gsdf* drives high expression in males during sex differentiation (Myosho et al., 2012). So far, male-specific regulation of *SOX1b* appears to be unique to Macquarie perch, and potentially to only some populations. In the golden perch genome, a locus with a female-specific allele is co-located with *SOX1a*, suggesting the potential involvement of this gene in sex determination or sex differentiation in that species. In other fishes, *SOX1b* and *SOX1a* genes play roles in the early processes of neural development, defining the neural subdivisions and the workings of the central nervous system (Gao et al., 2015; Lekk et al., 2019; Okuda et al., 2006).

The 146-bp sexing region as a candidate distal enhancer of SOX1b

Non-coding DNA under evolutionary constraint and clustering of transcription-factor binding-site motifs are the two genomic features used to predict cis-regulatory elements, i.e. DNA sequences required to regulate tissue- and time-specific gene expression (Hardison & Taylor, 2012). Accordingly, the Macquarie perch 146-bp sexing region, occurring ~38 Kb upstream of *SOX1b* in a non-coding region of the genome, could act as distal enhancer of *SOX1b*: it shows evolutionary conservation across multiple orders of fish and multiple matches to known transcription factor binding motifs, with mutations on Y-chromosomes resulting in potential gains and losses of transcription factor binding sites (Appendix F). The loss on Y-haplotype of the binding site for *SOX14*, a negative regulator of transcription, provides a plausible candidate mechanism of male-specific transcription of *SOX1b*, with additional potential changes in regulation suggested by gains and losses of binding sites for homeodomain transcription factors. Future functional experiments, such as those performed by Takehana et al. (2014) and/or investigation of 3D organization of chromatin (Shlyueva, Stampfel, & Stark, 2014; Zheng & Xie, 2019) will be required to clarify the role of *SOX1b*, *SOX1a*, and the putative *SOX1b* enhancer in gonad differentiation and the sex-determination pathway in Percichthyidae. Changes in gene regulation are a quick and efficient ways by which genes can acquire novel functions during sex chromosome turnover (Schartl et al., 2018).

Enhancers can be located a million bases away from their target genes (Hardison & Taylor, 2012), but studies of cis-regulatory regions of *SOX* genes in other organisms generally report smaller distances. For example, in chicken, time-specific cranial and vagal/trunk *SOX10* expression in embryo is regulated by two clusters of ~160 bp and ~267 bp located within a 3.5-Kb genomic fragment ~1-Kb downstream of *SOX10* (Betancur, Bronner-Fraser, & Sauka-Spengler, 2010). In zebrafish (*Danio rerio*) a *SOX10* cis-regulatory region is located within a 4.9-Kb region upstream of *SOX10* (Carney et al., 2006). In Japanese flounder *Paralichthys olivaceus*, many putative transcription factor binding sites were identified within 2 Kb upstream of *SOX1a* and *SOX1b*, including those for factors regulating neurogenesis, pluripotent or stem-cell properties and diverse cellular processes (Gao et al., 2015).

Environmental sex reversal or selection for sex chromosome lability?

Sex determination in fishes appears to be regulated along a continuum of genetic and environmental factors (Heule, Salzburger, & Bohne, 2014) and experiments involving hormone treatments or temperature shock can

induce sex-reversal in some species (Devlin & Nagahama, 2002). Thus, environmental factors stimulating production of steroid hormones could influence genetic sex (Nakamura, 2010). Hatchery data for some percichthyids suggest a possibility of temperature- and/or hormone-driven sex-ratio biases (Ingram et al., 2012; Lyon et al., 2012). Our data support sex-reversal in Macquarie perch: a homozygous YY genotype for the XY-gametologous region in one male (male 2 on Fig. 1) suggests the existence of XY females, also supported by phenotypic females with Y-alleles and putatively lower fertilization rates. In addition, the XX genotype in another male for the sexing region, but not the other XY-gametologous SNP (male 19 on Fig. 1) could be explained by recombination between X and Y chromosomes in a XY female ancestor (simultaneous reverse mutation at seven loci being unlikely). Environmental variation overriding genetic sex can promote sex chromosome turnover (Baroiller et al., 2009) and induce shifts between XY and ZW systems (Holleley et al., 2015). Indeed, the prevalence in golden perch of SNPs with female-specific alleles, despite XY sex determination inferred from karyology (Shams et al., 2019), suggests ancestral transitions between ZW and XY systems, or even possible co-existence of XY and ZW systems in this species, as inferred for other fishes (Nguyen et al., 2021).

Alternatively, intraspecific polymorphism for sex-determining loci could be present in Macquarie perch, as observed in some tilapia species (Tao et al., 2021). Unstable environments, where one sex can become locally extinct by drift or maladaptation, can promote sex-chromosome turnover. Different sex-determination systems can evolve rapidly in populations of the same species in response to loss of sex chromosomes by drift. For example, natural populations of zebrafish have a monogenic ZW sex-determination system, whereas polygenic sex-determination systems of different laboratory strains have been inferred (Anderson et al., 2012; Bradley et al., 2011; Liew et al., 2012), thought to have evolved de novo after loss of W-alleles by unmasking latent pre-existing environmental sex-determination mechanisms or polygenic sex-determination (Wilson et al., 2014). In Macquarie perch, which breeds in shallow running waters and may experience extreme conditions during prolonged droughts or after catastrophic fires, frequent sex-chromosome turnover could prevent local extinctions. Weaker selection on sex chromosome lability in the golden perch and Murray cod, which occur in more stable and connected lowland river systems, can partially explain the apparently older age of the sex chromosomes of these percichthyids.

Conservation implications

Most remaining Macquarie perch population are small and isolated and thus vulnerable to extinction through environmental stochasticity, loss of adaptive potential, and inbreeding. Accordingly, genetic admixture among populations has been recommended (Pavlova et al., 2017) and can have positive fitness effects during restoration of previously extinct populations (Lutz et al., 2021). Potential presence of intraspecific polymorphism for the putative sex-determining locus in Macquarie perch raises the question about the consequences of admixture between populations that have different sex-determination systems. Examples from other fish suggest that one of the systems will dominate over the other. In interspecific crosses between two sister species with different independently evolved sex chromosome systems, western mosquitofish *Gambusia affinis* with ZW, and eastern mosquitofish *G. holbrooki* with XY, the Y chromosome is dominant over the W chromosome, and X is dominant over Z, such that YW fish develop as males (Kottler et al., 2020). In southern platyfish *Xiphophorus maculatus*, which has three well-defined sex chromosomes, YW fish develop as females, indicating that the W chromosome is dominant over the Y chromosome (Schultheis et al., 2009). Successful admixture of coastal and inland Macquarie perch lineages in the Cataract River for several generations (Pavlova et al., 2017) suggests that admixture will not cause severe issues for sex-ratios even if diverse sex-determination mechanisms exist.

Lessons learned and remaining uncertainties

The lack of XY-gametologous loci in DArT SNPs in either perch species and other fish with homomorphic sex chromosomes (Nguyen et al., 2021) suggests that approaches other than reduced-representation sequencing may be required to identify sex-linked loci in systems with young sex chromosomes. Methylation-sensitive approaches, such as DArTseq, will also have low power in detecting cis-regulatory regions that play a role of sex-determining loci, because these regions might be methylated in adults. Successful detection of a short

XY-gametologous locus in Macquarie perch using WGS data also implies that lack of sex-linked markers in reduced-representation data cannot be used as evidence for lack of sex chromosomes in a species. Acquiring WGS data for known-sex samples from other populations will be crucial for improving our understanding of sex determination in this and other percichthyid species. The set of 22 Macquarie perch loci with male-specific alleles (including eight XY-gametologous loci; Fig. 1) may not have the power to predict the genotypic or phenotypic sex in some populations of Macquarie perch, due to fast sex chromosome turnover and/or presence of environmental sex-reversal. At least some of these loci appear to be autosomally-inherited in golden perch, and in more distant relatives from the family Percichthyidae (Murray cod and trout cod), suggesting independent evolution of the Macquarie perch sex-determining locus. However, these properties make these markers a great resource for studying sex chromosome turnover across populations and species.

Whilst *SOX1b* appears to be a promising novel candidate sex-determining gene, it is possible that we missed other important sex-linked regions due to the highly repetitive nature of the fish genomes precluding efficient short-read mapping, or due to our draft genome missing relevant loci. Alternatively, polygenic sex-determination in Macquarie perch is possible, given the many sex-linked loci detected and some clustering of sex-specific alleles in scaffolds other than 633. In sea bass *Dicentrarchus labrax* at least four loci are associated with sex determination (Palaiokostas et al., 2015). Also, both XY and ZW sex-determining systems are present on different chromosomes in some *Metriaclima* cichlids, with epistasis between genotypes at two loci controlling sex (Ser, Roberts, & Kocher, 2010). Three alleles (X, Y, and W) segregate at the sex-determination locus and interact in platyfish *Xiphophorus maculatus* (Schultheis et al., 2009). Structural genomic variation can also play a major role in adaptation, speciation and sex determination (Merot, Oomen, Tigano, & Wellenreuther, 2020; Natri et al., 2019), and the role of sex-linked deletions and other structural variants in sex determination of Macquarie perch and other percichthyids remains to be clarified (Appendix H).

Conclusions

The Macquarie perch and golden perch genomes reported here triple the number of genomes available for Percichthyidae, the only other available genome being that of the Murray cod (Austin et al., 2017). These resources will facilitate population genomic and functional genomic research in Percichthyidae and development of reliable sexing markers for different populations, thus contributing to species management and conservation. Percichthyidae present a promising model for studying sex chromosome turnover, and the role of sex chromosomes in adaptation and speciation. Our study provides a novel candidate sex-determination gene and a distal enhancer for a key developmental gene for future studies. This makes Macquarie perch one of few fish species with a localized candidate sex-determining mechanism. Involvement of *SOX1b* in sex determination broadens the roles of this major developmental gene.

Acknowledgements

We thank Marty Asmus, Renae Ayres, Mark Babbs, Matt Beitzel, Tom Butterfield, Jonathon Doyle, Graeme Hackett, Dean Hartwell, Mark Lintermans, Alasdair McDonald, Duncan McLay, Justin O'Mahony, Luke Pearce, Rohan Rehwinkel, Joanne Sharley, Chris Smith, Daniel Stoessel, Jason Thiem, Zeb Tonkin and Ian Wooden for sample collection, and Maiko Lutz and Anna Poleskiy for lab assistance. This work was supported by Australian Research Council Linkage Grants LP110200017 to Monash University, Flinders University and University of Canberra, with Partner Organizations University of Montana, ACTEW Corporation (now Icon Water), Department of Sustainability and Environment (Victoria) (now within Department of Environment, Land, Water & Planning; DELWP), Fisheries Victoria (now within Department of Economic Development, Jobs, Transport and Resources) and Melbourne Water, and LP160100482 to Monash University and La Trobe University, with Partner Organizations University of Canberra, DELWP, Diversity Arrays Technology, Zoos Victoria, Environment, Planning & Sustainable Development Directorate (ACT Government), and Department of Biodiversity, Conservation and Attractions (Western Australia). Sample collection was supported by funding from the Commonwealth Environmental Water Office, Australian Government. Computationally intensive analyses were run on virtual servers hosted on NeCTAR Research Cloud, courtesy of Monash eResearch. We are grateful to Jennifer Graves, Sridevi Sureshkumar, Sureshku-

mar Balasubramanian, Craig Dent, Jamie Gearing, Kirill Tsyganov, Robert Bryson-Richardson and Alexei Tulin for helpful discussions and Andrew DeWoody and three anonymous reviewers for their thoughtful comments which helped us to improve the manuscript.

References

- Amos, J. N., Harrisson, K. A., Radford, J. Q., White, M., Newell, G., Mac Nally, R., . . . Pavlova, A. (2014). Species- and sex-specific connectivity effects of habitat fragmentation in a suite of woodland birds. *Ecology*, *95* (6), 1556–1568. doi:<http://dx.doi.org/10.1890/13-1328.1>
- Anderson, J. L., Mari, A. R., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., & Postlethwait, J. H. (2012). Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS ONE*, *7* (7).
- Appleford, P., Anderson, T., & Gooley, G. (1998). Reproductive cycle and gonadal development of Macquarie perch, *Macquaria australasica* Cuvier (Percichthyidae), in Lake Dartmouth and tributaries of the Murray–Darling Basin, Victoria, Australia. *Marine and Freshwater Research*, *49* (2), 163-169.
- Austin, C. M., Tan, M. H., Harrisson, K. A., Lee, Y. P., Croft, L. J., Sunnucks, P., . . . Gan, H. M. (2017). De novo genome assembly and annotation of Australia’s largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *GigaScience*, *6* (8), 1–6.
- Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14* (2), 113.
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., . . . Ming, R. (2014). Sex determination: why so many ways of doing it? *PLoS Biol*, *12* (7), e1001899.
- Bao, L., Tian, C., Liu, S., Zhang, Y., Elasad, A., Yuan, Z., . . . Zhou, T. (2019). The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. *Bmc Biology*, *17* (1), 6.
- Baroiller, J.-F., D’Cotta, H., & Saillant, E. (2009). Environmental effects on fish sex determination and differentiation. *Sexual Development*, *3* (2-3), 118-135.
- Betancur, P., Bronner-Fraser, M., & Sauka-Spengler, T. (2010). Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *Proceedings of the National Academy of Sciences*, *107* (8), 3570-3575.
- Bradley, K. M., Breyer, J. P., Melville, D. B., Broman, K. W., Knapik, E. W., & Smith, J. R. (2011). An SNP-based linkage map for zebrafish reveals sex determination loci. *G3: Genes, Genomes, Genetics*, *1* (1), 3-9.
- Bruna, T., Hoff, K., Stanke, M., Lomsadze, A., & Borodovsky, M. (2020). BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *bioRxiv*, doi:<https://doi.org/10.1101/2020.08.10.245134>
- Carney, T. J., Dutton, K. A., Greenhill, E., Delfino-Machin, M., Dufourcq, P., Blader, P., & Kelsh, R. N. (2006). A direct role for Sox10 in specification of neural crest-derived sensory neurons. *Development*, *133* (23), 4619-4630.
- Carvalho, A. B., & Clark, A. G. (2013). Efficient identification of Y chromosome sequences in the human and Drosophila genomes. *Genome Research*, *23* (11), 1894-1907.
- Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, *95* (2), 118-128.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34* (17), i884-i890.

- Croft, B., Ohnesorg, T., Hewitt, J., Bowles, J., Quinn, A., Tan, J., . . . Sreenivasan, R. (2018). Human sex reversal is caused by duplication or deletion of core enhancers upstream of SOX9. *Nature Communications*, *9* (1), 1-10.
- Darolti, I., Wright, A. E., Sandkam, B. A., Morris, J., Bloch, N. I., Farre, M., . . . Breden, F. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proceedings of the National Academy of Sciences*, *116* (38), 19031-19036.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Hanna, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43* (5), 491.
- Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture*, *208* (3-4), 191-364.
- Dissanayake, D. S., Holleley, C., Hill, L., O'Meally, D., Deakin, J., & Georges, A. (2020). Identification of Y chromosome markers in the eastern three-lined skink (*Bassiana duperreyi*) using in silico whole genome subtraction. *BMC Genomics*, *21* (1), 667.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29* (1), 15-21.
- Faggion, S., Vandeputte, M., Chatain, B., Gagnaire, P.-A., & Allal, F. (2019). Population-specific variations of the genetic architecture of sex determination in wild European sea bass *Dicentrarchus labrax* L. *Heredity*, *122* (5), 612-621.
- Ferraresso, S., Bargelloni, L., Babbucci, M., Cannas, R., Follesa, M. C., Carugati, L., . . . Sapounidis, A. (2021). fshr: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations. *Iscience*, *24* (1), 101886.
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). Repeat-Modeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117* (17), 9451-9457.
- Fornes, O., Castro-Mondragon, J. A., Khan, A., Van der Lee, R., Zhang, X., Richmond, P. A., . . . Baranašić, D. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, *48* (D1), D87-D92.
- Gan, H. M., Falk, S., Morales, H. E., Austin, C. M., Sunnucks, P., & Pavlova, A. (2019). Genomic evidence of neo-sex chromosomes in the Eastern Yellow Robin. *GigaScience*, *8* (9), 1-10. doi:<https://doi.org/10.1093/gigascience/giz111>
- Gao, J., Zhang, W., Li, P., Liu, J., Song, H., Wang, X., & Zhang, Q. (2015). Identification, molecular characterization and gene expression analysis of sox1a and sox1b genes in Japanese flounder, *Paralichthys olivaceus*. *Gene*, *574* (2), 225-234.
- Gearing, L. J., Cumming, H. E., Chapman, R., Finkel, A. M., Woodhouse, I. B., Luu, K., . . . Hertzog, P. J. (2019). CiiiDER: A tool for predicting and analysing transcription factor binding sites. *PLoS ONE*, *14* (9), e0215495.
- Gooley, G., Anderson, T., & Appleford, P. (1995). Aspects of the reproductive cycle and gonadal development of Murray cod, *Maccullochella peelii peelii* (Mitchell) (Percichthidae), in Lake Charlegrark and adjacent farm ponds, Victoria, Australia. *Marine and Freshwater Research*, *46* (4), 723-728.
- Graves, J. A. M. (2013). How to evolve new vertebrate sex determining genes. *Developmental Dynamics*, *242* (4), 354-359.
- Graves, J. A. M., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology*, *11* (4), 205.

- Gruber, B., Unmack, P., Berry, O., & Georges, A. (2018). Introduction to dartR.
- Hardison, R. C., & Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nature Reviews Genetics*, *13* (7), 469-483.
- Heule, C., Salzburger, W., & Böhne, A. (2014). Genetics of sexual development: an evolutionary playground for fish. *Genetics*, *196* (3), 579-591.
- Hill, P. L., Burrige, C. P., Ezaz, T., & Wapstra, E. (2018). Conservation of sex-linked markers among conspecific populations of a viviparous skink, *Niveoscincus ocellatus*, exhibiting genetic and temperature-dependent sex determination. *Genome Biology and Evolution*, *10* (4), 1079-1087.
- Holleley, C. E., O'Meally, D., Sarre, S. D., Graves, J. A. M., Ezaz, T., Matsubara, K., . . . Georges, A. (2015). Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. *Nature*, *523* (7558), 79-82.
- Hu, Y., Wang, B., & Du, H. (2021). A review on sox genes in fish. *Reviews in Aquaculture*, <https://doi.org/10.1111/raq.12554> . doi:<https://doi.org/10.1111/raq.12554>
- Ingram, B. A., Ho, H. K., Turchini, G. M., & Holland, M. (2012). *Gamete quality and spawning in captive Murray cod broodstock* . Fisheries Victoria Research Report Series No. 58.: Department of Primary Industries.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., . . . Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30* (9), 1236-1240.
- Kiefer, J. C. (2007). Back to basics: Sox genes. *Developmental dynamics: an official publication of the American Association of Anatomists*, *236* (8), 2356-2366.
- Kijas, J., McWilliam, S., Sanchez, M. N., Kube, P., King, H., Evans, B., . . . Verbyla, K. (2018). Evolution of Sex Determination Loci in Atlantic Salmon. *Scientific Reports*, *8* (1), 5664.
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., . . . Hopper, C. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Data Production and Analysis in Population Genomics: Methods and Protocols* , 67-89.
- Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Källberg, M., . . . Krusche, P. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods*, *15* (8), 591.
- Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27* (24), 3435-3436.
- Koopman, P. (2005). Sex determination: a tale of two Sox genes. *Trends in Genetics*, *21* (7), 367-370.
- Kottler, V. A., Feron, R., Nanda, I., Klopp, C., Du, K., Kneitz, S., . . . Lluch, J. (2020). Independent origin of XY and ZW sex determination mechanisms in mosquitofish sister species. *Genetics*, *214* (1), 193-209.
- Lavoué, S., Nakayama, K., Jerry, D. R., Yamanoue, Y., Yagishita, N., Suzuki, N., . . . Miya, M. (2014). Mitogenomic phylogeny of the Percichthyidae and Centrarchiformes (Percomorphaceae): comparison with recent nuclear gene-based studies and simultaneous analysis. *Gene*, *549* (1), 46-57.
- Lekk, I., Duboc, V., Faro, A., Nicolaou, S., Blader, P., & Wilson, S. W. (2019). Sox1a mediates the ability of the parapineal to impart habenular left-right asymmetry. *eLife*, *8* , e47376.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25* (14), 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). 1000 Genome Project Data Processing Subgroup: The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, *25* , 2078 - 2079.

- Liew, W. C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K. R., & Orban, L. (2012). Polygenic sex determination system in zebrafish. *PLoS ONE*, 7 (4).
- Lubieniecki, K. P., Lin, S., Cabana, E. I., Li, J., Lai, Y. Y., & Davidson, W. S. (2015). Genomic instability of the sex-determining locus in Atlantic salmon (*Salmo salar*). *G3: Genes, Genomes, Genetics*, 5 (11), 2513-2522.
- Lutz, M., Tonkin, Z., Yen, J. D., Johnson, G., Ingram, B., Sharley, J., . . . Pavlova, A. (2021). Using multiple sources during reintroduction of a locally extinct population benefits survival and reproduction of an endangered freshwater fish. *Evolutionary Applications*, 14 (4), 950-964. doi:10.1111/eva.13173
- Lyon, J. P., Todd, C., Nicol, S. J., MacDonald, A., Stoessel, D., Ingram, B. A., . . . Bradshaw, C. J. A. (2012). Reintroduction success of threatened Australian trout cod (*Maccullochella macquariensis*) based on growth and reproduction. *Marine and Freshwater Research*, 63 (7), 598. doi:10.1071/mf12034
- Mank, J. E., Promislow, D. E., & Avise, J. C. (2006). Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society*, 87 (1), 83-93.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17 (1), 10-12.
- Martínez, P., Viñas, A. M., Sánchez, L., Díaz, N., Ribas, L., & Piferrer, F. (2014). Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Frontiers in Genetics*, 5 (340). doi:10.3389/fgene.2014.00340
- Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, 35 (7), 561-572.
- Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., . . . Marshall, D. (2012). Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, 14 (2), 193-202. doi:10.1093/bib/bbs012
- 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.
- Nakamura, M. (2010). The mechanism of sex determination in vertebrates—are sex steroids the key-factor? *Journal of experimental zoology part A: Ecological genetics and physiology*, 313 (7), 381-398.
- Natri, H. M., Merila, J., & Shikano, T. (2019). The evolution of sex determination associated with a chromosomal inversion. *Nature Communications*, 10 (1), 1-13.
- Nguyen, D. H. M., Panthum, T., Ponjarat, J., Laopichienpong, N., Kraichak, E., Singchat, W., . . . Srikulnath, K. (2021). An investigation of ZZ/ZW and XX/XY sex determination systems in North African catfish (*Clarias gariepinus*, Burchell, 1822). *Frontiers in Genetics*, 11 (1719). doi:10.3389/fgene.2020.562856
- Nguyen, N. H., Premachandra, H., Kilian, A., & Knibb, W. (2018). Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi*. *BMC Genomics*, 19 (1), 107.
- Okuda, Y., Yoda, H., Uchikawa, M., Furutani-Seiki, M., Takeda, H., Kondoh, H., & Kamachi, Y. (2006). Comparative genomic and expression analysis of group B1 sox genes in zebrafish indicates their diversification during vertebrate evolution. *Developmental dynamics: an official publication of the American Association of Anatomists*, 235 (3), 811-825.
- Palaiokostas, C., Bekaert, M., Taggart, J. B., Gharbi, K., McAndrew, B. J., Chatain, B., . . . Vandeputte, M. (2015). A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genetics Selection Evolution*, 47 (1), 68.

- Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology* .
- Pavlova, A., Amos, J. N., Joseph, L., Loynes, K., Austin, J. J., Keogh, J. S., . . . Sunnucks, P. (2013). Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution*, *67* (12), 3412–3428. doi:10.1111/evo.12107; DRYAD doi:10.5061/dryad.f5q32
- Pavlova, A., Beheregaray, L. B., Coleman, R., Gilligan, D., Harrisson, K. A., Ingram, B. A., . . . Sunnucks, P. (2017). Severe consequences of habitat fragmentation on genetic diversity of an endangered Australian freshwater fish: a call for assisted gene flow. *Evolutionary Applications*, *10* , 531–550.
- Penman, D. J., & Piferrer, F. (2008). Fish gonadogenesis. Part I: genetic and environmental mechanisms of sex determination. *Reviews in Fisheries Science*, *16* (S1), 16-34.
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Rodrigues, N., Merila, J., Patrelle, C., & Perrin, N. (2014). Geographic variation in sex-chromosome differentiation in the common frog (*Rana temporaria*). *Molecular Ecology*, *23* (14), 3409-3418.
- Rodriguez-Mari, A., Wilson, C., Titus, T. A., Canestro, C., BreMiller, R. A., Yan, Y.-L., . . . Gray, E. M. (2011). Roles of brca2 (fancd1) in oocyte nuclear architecture, gametogenesis, gonad tumors, and genome stability in zebrafish. *Plos Genetics*, *7* (3), e1001357.
- Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., & Peichel, C. L. (2009). Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *Plos Genetics*, *5* (2), e1000391.
- Sarkar, A., & Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell stem cell*, *12* (1), 15-30.
- Schartl, M., Schories, S., Wakamatsu, Y., Nagao, Y., Hashimoto, H., Bertin, C., . . . Centanin, L. (2018). Sox5 is involved in germ-cell regulation and sex determination in medaka following co-option of nested transposable elements. *Bmc Biology*, *16* (1), 1-17.
- Schultheis, C., Bohne, A., Schartl, M., Volf, J., & Galiana-Arnoux, D. (2009). Sex determination diversity and sex chromosome evolution in poeciliid fish. *Sexual Development*, *3* (2-3), 68-77.
- Schwanz, L. E., Georges, A., Holleley, C. E., & Sarre, S. (2020). Climate change, sex reversal and lability of sex determining systems. *Journal of Evolutionary Biology* .
- Seppy, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In M. Kollmar (Ed.), *Gene Prediction. Methods in Molecular Biology, vol 1962* . New York, NY: Humana.
- Ser, J. R., Roberts, R. B., & Kocher, T. D. (2010). Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution: International Journal of Organic Evolution*, *64* (2), 486-501.
- Shams, F., Dyer, F., Thompson, R., Duncan, R. P., Thiem, J. D., Majtanova, Z., & Ezaz, T. (2019). Karyotypes and Sex Chromosomes in Two Australian Native Freshwater Fishes, Golden Perch (*Macquaria ambigua*) and Murray Cod (*Maccullochella peelii*)(Percichthyidae). *International Journal of Molecular Sciences*, *20* (17), 4244.
- Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, *15* (4), 272-286.
- Smit, A., Hubley, R., & Green, P. (2013-2015). RepeatMasker Open-4.0 <<http://www.repeatmasker.org>>.

- Takehana, Y., Matsuda, M., Myosho, T., Suster, M. L., Kawakami, K., Shin, T., . . . Fujiyama, A. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature Communications*, *5* (1), 1-10.
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2007). Evidence for different origins of sex chromosomes in closely related *Oryzias* fishes: substitution of the master sex-determining gene. *Genetics*, *177* (4), 2075-2081.
- Tao, W., Xu, L., Zhao, L., Zhu, Z., Wu, X., Min, Q., . . . Zhou, Q. (2021). High-quality chromosome-level genomes of two tilapia species reveal their evolution of repeat sequences and sex chromosomes. *Molecular Ecology Resources*, *21* (2), 543-560.
- Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., & Van de Peer, Y. (2003). Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Research*, *13* (3), 382-390.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., . . . Thibault, J. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*, *43* (1), 11.10. 11-11.10. 33.
- Vicoso, B., Emerson, J., Zektser, Y., Mahajan, S., & Bachtrog, D. (2013). Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol*, *11* (8), e1001643.
- Wilson, C. A., High, S. K., McCluskey, B. M., Amores, A., Yan, Y.-l., Titus, T. A., . . . Schartl, M. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics*, *198* (3), 1291-1308.
- Wright, A. E., Darolti, I., Bloch, N. I., Oostra, V., Sandkam, B., Buechel, S. D., . . . Mank, J. E. (2017). Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nature Communications*, *8* (1), 1-10.
- Zhang, S., Chen, X., Wang, M., Zhang, W., Pan, J., Qin, Q., . . . Jiang, H. (2018). Genome-wide identification, phylogeny and expressional profile of the Sox gene family in channel catfish (*Ictalurus punctatus*). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *28* , 17-26.
- Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7* (1-2), 203-214.
- Zheng, H., & Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology*, *20* (9), 535-550.
- Zimin, A. V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marcais, G., . . . Salzberg, S. L. (2017). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Research*, *27* (5), 787-792.

Data Accessibility

The raw reads and genome assemblies have been deposited in NCBI under the BioprojectID PRJNA516983 for Macquarie perch and PRJNA556086 for golden perch. Genome annotation data (repeat annotation, predicted protein-coding genes, repeat-masked genomes and InterProScan output) have been deposited in Zenodo repository (<https://doi.org/10.5281/zenodo.4315338>). Raw reads generated from whole genome resequencing (WGS) of the 100 Macquarie perch have also been deposited in NCBI under the BioProjectID PRJNA598965. DArT genotypes are available in Bridges Monash Research Repository <https://doi.org/10.26180/5ea1736cc3553> for Macquarie perch and <https://doi.org/10.26180/13551713> for golden perch.

Author Contributions

AP, HMG, PS and KH conceived the manuscript, BI and DG collected hatchery data, HMG sequenced, assembled and annotated genome, YPL prepared libraries and obtained WGS data, MHG, AP, RT and KH analysed data, AP, PS and KH designed and tested sexing assays, AP drafted the manuscript, all authors contributed to writing.

Tables and Figures

Table 1 Genome assembly and annotation metrics and accession for Macquarie perch and golden perch reference genomes, DArTseq and WGS data.

	Macquarie perch	golden perch
NCBI BioProjectID	PRJNA516983	PRJNA55608
DNaseq:		
Sample ID genome	MP_SCH12	GOP001
BioSample DNaseq	SAMN10817714	SAMN123406
Illumina DNaseq data	237.1 Gb	264.5 Gb
NCBI-SRA accession DNaseq	SRX7207073	SRX6595905
Nanopore 1D data	7.6 Gb	3.1 Gb
NCBI-SRA accession Nanopore	SRX7207074	SRX6595904
Genome assembly:		
Assembled genome size	675.98 Mb	661.43 Mb
Scaffold N50	845 Kb	252 Kb
Number of scaffolds	2,962 scaffolds	7,165 scaffolds
Contig N ₅₀	679 Kb	244 Kb
Number of contigs	3,369 contigs	7,352 contigs
GenBank assembly accession	GCA_005408345.1	GCA_0083609
WGS Project accession	SEMNO1	VMKM01
Whole Genome Shotgun accession	SEMNO0000000.1	VMKM00000
BUSCO3 genome completeness	96.9%	95.9%
Genome repetitiveness	27.7%	25.8%
RNaseq:		
Sample ID RNaseq	MP_527	GPTT01
BioSample RNaseq	SAMN13380527	SAMN149435
Illumina RNaseq, G bases	15.1 Gb	11.6 Gb
NCBI-SRA accession RNaseq	SRX7207075	SRX8356284
Genome annotation:		
Number of predicted protein-coding genes	33,422	36,108
Number of functionally annotated protein-coding genes	29,940	31,895
Number of genes with GO annotations	16,264	16,792
Zenodo DOI for annotations by RepeatModeler2, BRAKER2 and InterProScan5	10.5281/zenodo.4315338	10.5281/zenodo.4315338
DArTseq:		
Bridges DOI for genotypes and covariate files	10.26180/5ea1736cc3553	10.26180/135
Whole Genome Resequencing:		
NCBI BioProjectID	PRJNA598965	NA

Table 2. Summary of number of loci that are significantly differentiated between two sexes in two populations on CMH test and have sex-specific alleles (details in Supplementary Material S2, Tables C1 and C2).

	Loci with male-specific allele	Loci with male-specific allele	Loci with female-specific allele	Loci with female-specific allele
Data type	#scaffolds	# loci	#scaffolds	# loci
Golden perch DArT	13	14	30	34
Macquarie perch DArT	2	2	2	2
Macquarie perch WGS (CMH test p<1e-5)	669	1831	707	1864
Macquarie perch scaffold 633 (CMH test p<1e-5)	-	40	-	2
Macquarie perch WGS (CMH test p<1e-20)	1+	6+	0	0
Macquarie perch scaffold 633 (CMH test p<1e-20)	-	6+	-	0

+In addition to the six loci presented in the table (all with a single allele in females and located within 91773-93327 of Macquarie perch scaffold 633), three loci on scaffold 633 and one on scaffold 251 were strongly (p<1e-20) differentiated between male and female pools but had a single count of a second allele in one of the female pools (Table B2).

Table 3. Summary of Macquarie perch individual-based WGS genotypes for the 146-bp sexing region of scaffold 633 (in bold) plus four additional SNPs with CMH test p<1e-20). F ref hom/M ref hom = number of females/males homozygous for reference allele, F het/M het = number of females/males that are heterozygous, F alt hom/M alt hom = number of females/males that are homozygous for the alternative SNP allele. When more than one alternative allele was present, M alt hom is the count of homozygotes at the more common alternative allele. Male-specific - locus with male-specific allele, XY- XY-gametologs.

Scaffold 633 position	Ref allele	Alt alleles	F ref hom	F het	F alt hom	M ref hom	M het	M alt hom
90152	C	G	50	0	0	16	34	0
91773	G	A, T	50	0	0	23	27	0
93182	T	A	50	0	0	2	47	1
93184	TGTGTGTGTGA	T, TGTGTGT	50	0	0	2	44	4
93229+	C	T	50	0	0	1	46	3
93278	C	CTA	50	0	0	2	42	6
93299	T	C	50	0	0	1	42	7
93315	G	T	48++	0	0	2	40	7
93327	G	A	48++	0	0	1	41	7
93879	G	A	50	0	0	12	37	1
94017	A	G	50	0	0	0	49	1

+Locus used for PCR-RFLP sexing assay

++Two females had low read count for these SNPs, but reference allele was present

§One male had low read count for these SNPs but alternative allele was present

Figure 1. Visual summary of individual genotypes for sex-linked loci for 50 individual male Macquarie perch (numbered columns) for scaffold 633 region 90152-94017 (data in Supplementary Material S4). Loci with male-specific alleles: homozygote for the reference allele light blue, heterozygote- intermediate blue, homozygotes for alternative allele- dark blue. XY-gametologs: homozygotes for the reference allele- light yellow, heterozygotes- light orange, homozygotes for alternative allele- dark orange. Loci numbered in red text are XY-gametologous indels, black text-SNPs. All females were homozygous for the reference alleles at these loci (as per colour-coded genotype in the last column). Y-linked alleles for loci 1-13 appear in four distinct linkage patterns: (i) two loci: 1 and 9 (26 males); (ii) ten loci: 2, 3, 5-8, 10-13 (8 males); (iii) 10 loci: 1, 4-8, 10-13 (7 males); and (iv) 4 loci: 1, 7-9 (1 male).

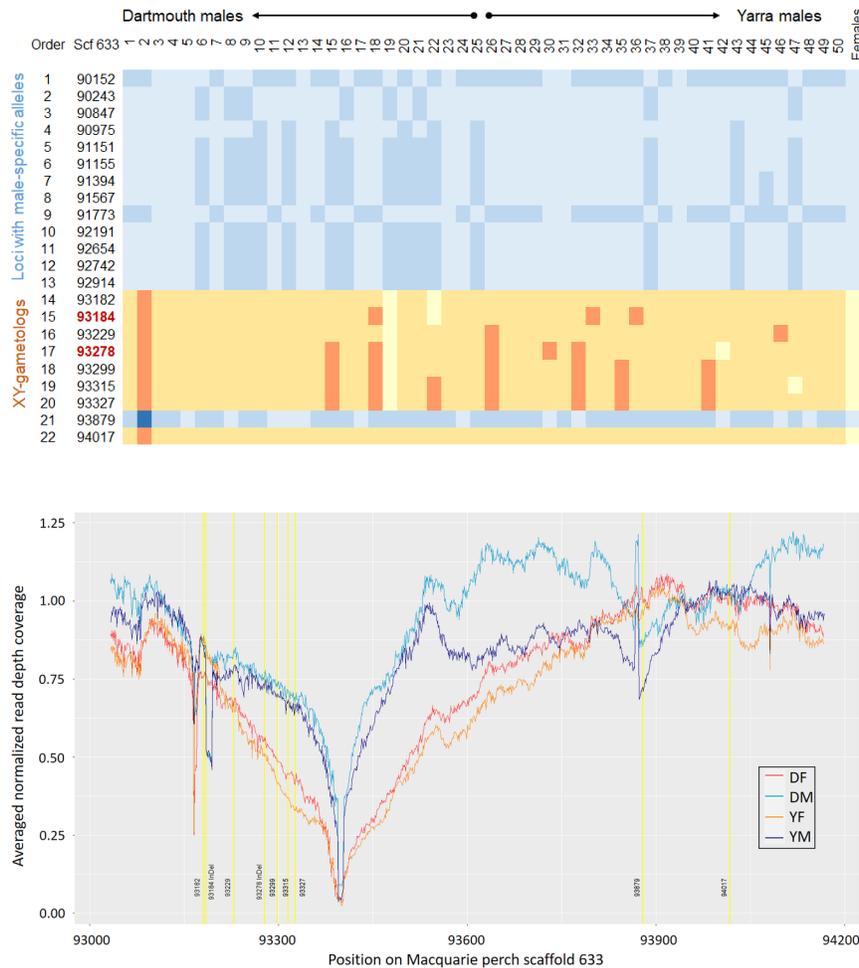


Figure 2 . Normalized WGS read depth coverage for four population-by-sex samples of individuals (Y-axis) along the sex-linked region of Macquarie perch scaffold 633 (SEMN0100063, bases 93050-94150 (X-axis). Dartmouth females- DF (red), Yarra females- YF (orange), Dartmouth males- DM (light blue), Yarra males- YM (dark blue). Vertical lines indicate location of nine sex-linked loci located in this region (see Table 3). Mean female:male read depth ratio for these loci was 1.15, 1.13, 0.92, 0.83, 0.72, 0.68, 0.65, 1.49 and 1.07, respectively.

Figure 3 . Read-depth coverage for reference and alternative alleles at nine consecutive sex-linked loci on

Macquarie perch scaffold 633 region 93182-94017 for four population-by-sex samples: Dartmouth females- DF, Dartmouth males- DM, Yarra females- YF and Yarra males- YM. Dashed horizontal line indicates genome-wide average. Box colours: pink- female, blue- male, reference allele (Ref)- lighter shade, alternative allele (Alt)- darker shade. All loci are XY-gametologs, except 93879 is a locus with a male-specific allele.

