Genomic data shed light on sex-determination in Australian freshwater Percichthyid fish species: Many ways to be a male

Alexandra Pavlova[1], Katherine Harrisson[2], Rustam Turakulov[3], Yin Peng Lee[4], Brett Ingram[5], Dean Gilligan[6], Paul Sunnucks[1], and Han Ming Gan[7]

[1]Monash University
[2]La Trobe University
[3]Australian National University
[4]Deakin University - Geelong Waterfront Campus
[5]Victorian Fisheries Authority
[6]Department of Primary Industries
[7]GeneSEQ Sdn Bhd

February 23, 2021

Abstract

Understanding sex-specific biology can aid conservation management. But understanding genomic sex differences of monomorphic fish species and developing molecular sexing assays is challenged by their diverse sex-determination systems. To facilitate research on Percichthyid fish, predominant in the Australian freshwater biota, we report whole genome sequences and annotations of the endangered Macquarie perch Macquaria australasica and its sister species, the golden perch M. ambigua. To identify sex-linked loci, we conducted whole genome resequencing on 100 known-sex Macquarie perch. In-silico pool-seq comparisons revealed few sex differences, but a 275-Kb SOX-containing scaffold was enriched for gametologous loci- homozygous in females, heterozygous in males. Within this scaffold we reconstructed X- and Y-linked 146-bp haplotypes containing 5 sex-linked SNPs, ~38 Kb upstream of SOX, and developed a PCR-RFLP sexing assay targeting the Y-linked allele of one SNP. We tested this assay in a panel of known-sex Macquarie perch, and smaller panels of three other confamilial species. Amplicon sequencing of 400 bp encompassing the 146-bp region revealed that the few sex-linked positions differ interspecifically, and within Macquarie perch such that its sexing test approached 100% reliability only for the populations used in assay development. Similarly, Macquarie- and golden perch genome-wide DArTseq SNPs revealed different sex-linked loci across non-homologous scaffolds. Overall, we identified 22 sex-linked SNPs in Macquarie perch in a predominantly XX/XY system in which females are homozygous at all 22, and males are heterozygous at 2 or more. The resources here will facilitate multi-locus sexing assays for both species and research on Percichthyid biology.

Introduction

Studies of sex-specific biology 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) and environmental dependency (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*, *gsdfY, 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). Moreover, new sex-determination systems can evolve relatively rapidly and frequently, resulting in intra-specific variation in sex-determination (Natri et al., 2019). Thus, identification of reliable, generally applicable sex-linked markers in fish can be challenging.

Homomorphic sex-chromosomes occur where gametologs have recombined relatively recently, and may result from recent switches in the chromosome pair used for 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 (Charlesworth, Charlesworth, & Marais, 2005). This suppression often occurs progressively and stepwise along chromosomes, starting from a sex-determining gene, resulting in strata of different ages, i.e. regions of the chromosome where recombination stopped at different evolutionary times (Charlesworth et al., 2005).

Several genomic approaches can be used to detect sex chromosomes (reviewed in Palmer, Rogers, Dean, & Wright, 2019). Two 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 on X or Z will show half as many reads in the heterogametic sex as the homogametic one, and Y- or W- loci will be absent in the homogametic sex. 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. 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 (Palmer et al., 2019). 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. Accordingly, combining approaches based on read depth and sex-specific variability can assist in detecting sex-linked loci.

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 Percichthyid species, golden perch *Macq. ambigua* and Murray cod *Macc. peelii*, 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). Consistent with low morphological differentiation of golden perch sex chromosomes, in its sister-species Macquarie perch (Lavoué et al., 2014), a set of >1200 genome-wide reduced-representation SNP loci 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., 2020). This suggests that in Macquarie perch sex-chromosomes may have young strata.

Sex determination in many fishes can be influenced by environmental factors, especially temperature (Devlin & Nagahama, 2002; Penman & Piferrer, 2008). Hatchery work suggests that environmental conditions may influence sex determination at an early stage of development in some Percichthyids (Ingram, Ho, Turchini,

& Holland, 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 genetic sex of fish would benefit conservation management of threatened Percichthyid species.

Here, we sequenced, assembled and annotated the genomes of two closely related sexually-monomorphic Percichthyids, Macquarie perch and golden perch. Then, using genomic datasets targeted for different resolution, we explored sex-linkage in these species. In particular, we aimed to test whether XY sex determination fully explains patterns of sex-linked genetic variation in these species, and to develop a non-lethal, affordable and rapid molecular test to determine the genetic sex of individuals. We tested for sex-linkage of SNPs obtained using reduced genomic representation DArTseq (Kilian et al., 2012) for Macquarie perch and golden perch of known sex, and examined whether putative sex-linked loci are clustered on some of the newly assembled scaffolds of respective reference genomes. Next, we obtained whole-genome resequencing (WGS) data for 50 female and 50 male Macquarie perch equally representing two relatively large populations used as sources for conservation interventions. We searched for SNPs significantly differentiated between sexes, using genotypes of four population-by-sex pools. This enabled identification of a small genomic region, showing predominantly XY patterns (one allele in females, two in males) on Macquarie perch scaffold 633. We explored individual sequence variation, identified X- and Y-linked haplotypes for this region, and developed a PCR-RFLP (polymerase chain reaction - restriction fragment length polymorphism) molecular sexing assay targeting a SNP with a Y-linked (male-specific) allele. We tested this assay in a panel of known-sex Macquarie perch individuals, and on smaller equivalent panels of golden perch, Murray cod and trout cod. We used amplicon sequencing to further explore sequence variation in this region for the four species. Our results revealed that the XY-homologous genomic region in Macquarie perch is species-specific and furthermore could be specific to populations related to those in which the test was developed. Overall, sequence variation across one small scaffold 633 region in two Macquarie perch populations generally supports XY sex-determination, but deletions in sexing region on both X and Y chromosomes, variation across species, and potentially across other genomic regions, calls for more complex sex-determination scenarios. The results and resources developed here will facilitate research on Percichthyidae, including the evolution of sex-determination. More broadly, the workflow described here could be used for developing molecular tests to sex 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 the Yarra River and Lake Dartmouth (see Lutz et al., 2020 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, Beheregaray, 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 from three populations, 66 were sampled lethally for another project, and their sex 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.

Genome and transcriptome sequencing for Macquarie perch and golden perch

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

DNA was extracted using Qiagen DNeasy Blood & Tissue kits. For Illumina sequencing, 100 ng of gDNA was fragmented to 350 bp using QSonica, processed with a NEB Ultra Illumina Library Preparation Kit and sequenced on a Novaseq6000 at the Deakin Genomics Centre using 2 × 151 bp run configuration (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).

To facilitate genome annotations, we performed mRNA sequencing for both species. We used adult female Macquarie perch MP_527, captured November 2017 from Lake Dartmouth. Samples from liver, ovary, brain, kidney and muscle were collected. We used adult golden perch 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 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 and sequenced on an Illumina NovaSeq6000.

Assembly and annotation of genomes for Macquarie perch and golden perch

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 and used to construct contigs by the de Bruijn graph approach, which then 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.

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).

Searching for sex-linked loci in reduced-representation DArT genotypes for Macquarie perch and golden perch

For 171 Macquarie perch and 66 golden perch of known sex we obtained genome-wide SNP markers using DArTseq, a reduced representation sequencing method similar to double-digest restriction-associated sequencing (list of samples in Supplementary Material S1). We genotyped 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.

4

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 (details in Appendix A). For quality control, each lane included ~25% of technical replicates from independent libraries. SNP discovery and genotyping were performed using DArT P/L's proprietary analytical pipeline (Jaccoud, Peng, Feinstein, & Kilian, 2001; Kilian et al., 2012). The method assembles short 69-bp DArT loci *de novo* . The DArT pipeline removed poor-quality sequences, applying more stringent criteria to the barcode region than the rest of the sequence, corrected low-quality bases from singleton tags using collapsed tags with multiple members as a template, and used a secondary pipeline (DArTsoft14) for SNP calling. Clusters, comprising tags differing by no more than 3 bases, were parsed 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. Reproducibility of SNP calls was then assessed based on score consistency between technical replicates, and SNPs with a reproducibility <95% were removed. No other filtering was performed at this stage, to maximize the chance of finding sex-linked loci. DArT loci were aligned to their respective newly-assembled reference genomes (NCBI WGS Project accessions: SEMN01 for Macquarie perch and VMKM01 for golden perch) using BLAST, with e-value [?]5e-5 and sequence identity [?]90%.

We tested each SNP locus for belonging to one of four types of strata consistent with XY sex-determination (Shams et al., 2019):

(*i*) Y-only sequences on old strata (hereafter, "*Y-linked* "): always apparently homozygous (actually hemizygous) in males, and absent in females.

(*ii*) Different alleles for non-recombining X- and Y- gametologs on moderately old strata (hereafter "*XY-gametologs* "): homozygous in all females and heterozygous in all males.

(*iii*) Y-limited variation on young strata (hereafter "*loci bearing recent-Y-specific polymorphism* "): homozygous in all females and either homozygous or heterozygous in males. Variation on Y-gametologs can accumulate faster than on X-gametologs due to lower X-Y than X-X recombination.

(*iv*) X-limited variation on old strata for loci missing on Y-gametologs (hereafter "*old-X-linked* "): homozygous or heterozygous in females but always hemizygous in males (males have half the read depth of females).

For these tests we used the *gl.sexlinkage* function of the dartR package (Gruber, Unmack, Berry, & Georges, 2018) in R (R Core Team, 2020). Under default settings- t.het = 0, t.hom = 0, system = 'xy'- the function looks for patterns consistent with (*i*) and (*ii*); here, the tolerance parameter t.het is the proportion of individuals of the sex expected to be homozygous (XX females) allowed to be heterozygous, and t.hom is the proportion of individuals of the sex expected to be heterozygous (XY males) allowed to be homozygous. To find (*iii*) and (*iv*), defined here as loci heterozygous in >10% of males or females, t.het =0 and t.hom = 0.9 were used under 'xy' and 'zw' systems respectively; these settings allowed up to 90% of individuals in the sex expected to have heterozygotes to be homozygous. None of the values of t.hom<0.8 in Macquarie perch, and t.hom<0.7 in golden perch returned sex-linked loci, indicating very low variability for sex-linked loci. 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.

Searching for sex-linked loci in whole genome sequencing (WGS) data for Macquarie perch

To explore sex-linkage across the draft genome, we obtained whole genome resequencing data (WGS) for 100 Macquarie perch from Dartmouth and Yarra; 25 individuals of each sex per population (all of these were also DArT-sequenced). DNA was extracted as above. For each individual, an Illumina Novaseq library was constructed, as 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).

A total of 25 million paired-end reads were randomly subsampled from each individual's WGS data and

combined into four pools (25 individuals of each sex per population): Dartmouth females, DF, Dartmouth males, DM, Yarra females, YF, and Yarra males, YM. Reads from the four pools were mapped onto the Macquarie perch reference genome (NCBI WGS Project accessions: SEMN01) using bwa-mem v0.7.17 (Li & Durbin, 2009). A Cochran-Mantel-Haenszel (CMH) test implemented in *PoPoolation2* (Kofler, Pandey, & Schlotterer, 2011) was used to detect significant and consistent allele frequency differences between sexes in two biological replicates. SNP loci with allele frequencies highly significantly (p<1e-20) differentiated between sexes were examined for being clustered in the same genomic region. Because one scaffold (scaffold 633, 274905 bp length, GenBank accession SEMN01000633) contained a set of the most differentiated loci, we also examined whether it had other loci with lower CMH test significance levels (p<1e-5) and whether some loci were polymorphic in one sex but monomorphic in the other.

Individual genotyping using all WGS read data was also performed, 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 genotypes for loci inferred to be sex-linked by CMH tests were examined for consistency with being XY-gametologs, old-X-linked loci or bearing recent-Y-specific polymorphism (*ii-iv* above). For the set of most strongly sex-differentiated SNP loci on scaffold 633, we visualized individual alignments for males in Tablet (Milne et al., 2012) to investigate whether the mate-pair architecture of the sequenced libraries resolved female- and male-specific haplotypes via manual inspection (similar approach in Kijas et al., 2018b). We also explored ploidy of these loci by investigating read-depth coverage for each base of the scaffold 633 region enriched with sex-linked loci. Read depth was collected using the *CollectAllelicCounts* function of GATK v4.1.0 (DePristo et al., 2011) from individual alignment files, output files were merged and analysed in R. Read depth for each individual was normalized by its genome-wide average read depth, and averaged across the four sex-by-population samples. The genome-wide average for each individual was calculated as the number of all mapped reads multiplied by 151 bp (read length) and divided by genome size. We calculated read depth for reference and alternate alleles separately.

Finally, we investigated whether scaffold 633 enriched with sex-linked loci has more indels in males than females, and whether these indels are clustered, as would be expected from accumulation of Y-specific polymorphisms due to recent reduction in recombination. Manta (Chen et al., 2016) was used to score structural variants (insertions, deletions, inversions and repeats) based on supporting paired and split-read evidence, using default parameters. Fisher and beta-binomial tests implemented in R libraries *R-core* and *ibb* (Pham & Jimenez, 2012) respectively were used to compare counts of different classes of structural variants between sexes, and sex-by-population samples.

Exploring annotations of sex-linked SNPs and seeking candidate sex-determining genes in reference genomes

We used genome organization and functional annotations for Macquarie perch and golden perch to test whether sex-linked SNPs were located in gene regions, and whether Macquarie perch scaffold 633 has annotated genes that might play a role in sex-determination. We also looked for the presence and location in the perch genomes of sex-determining genes known from teleost fish, searching 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, gsdfY,* 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).

PCR-RFLP molecular sexing assay for Macquarie perch

To develop the molecular sexing assay, we targeted a 146-bp region on Macquarie perch scaffold 633 (bases 93182-93327), which contained 5 sex-linked SNP loci homozygous for the (X-linked) reference allele in all sequenced females, with the alternative (Y-linked) allele present in most sequenced males. To develop conservative PCR primers intended to amplify in all Macquarie perch populations and possibly other species, we first used Blastn to search the golden perch genome for a region homologous to 600 bp of Macquarie perch

6

sequence encompassing the 146-bp region of scaffold 633 (SEMN01000633, bases 93001-93600). This search yielded a 624-bp region on golden perch scaffold VMKM01003747 (bases 16365-16988), which we aligned to putative X- and Y-linked Macquarie perch haplotypes in Geneious Pro (Appendix B, Fig. B1). We designed two forward and two reverse primers in regions identical in the two species (Table B1, primers F1, F2, R2 and R1, with the 5'-end starting on scaffold 633 bases 93027, 93107, 93380, and 93474, respectively) and then used the SMS online tool (https://www.bioinformatics.org/sms2/rest_map.html) to find positions of restriction endonuclease cut sites targeting Macquarie perch SNPs with Y-specific alleles. *Ase* I was chosen because it recognizes ATTAAT that occurs only in the male-specific sequence (T base at 93229).

All four primer combinations were trialled initially in two female and two male Macquarie perch (see Appendix B for details). This assay was expected to produce 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. For the final assay on a larger sample, we ran PCRs with F1-R1 primers followed by *Ase* I digestion (Appendix B; list of samples in Supplementary Material S1). This assay was tested using 45 Macquarie perch hatchery-sexed individuals from Dartmouth, Yarra and Abercrombie, 21 field-sexed individuals from King Parrot Creek and Holland's Creek, and 41 Macquarie perch of unknown sex from eight other populations (two inland and six coastal).

Amplicon-based sequencing to test for sex-specific variation at the Macquarie perch sex-linked region in other species

We investigated whether the Macquarie perch sex-linked region is also sex-linked in three other species of Percichthyids. The 600-bp sex-linked fragment of Macquarie perch scaffold 633 (SEMN01000633; 93182-93327) had 91% identity (6% gaps) with a corresponding part of golden perch scaffold 3747 (VMKM01003747; 16365-16989) and 88% identity (7% gaps) with a section of Murray cod scaffold 2546 (LKNJ01002546; 82756-83361). An alignment of scaffolds of all three species revealed that one of the Macquarie perch sex-linked SNPs (scaffold 633 base 93181) was located in a microsatellite, part of which was missing from the Murray cod sequence (Fig. B1). Positions corresponding to the other four putatively sex-linked Macquarie perch SNPs were present in all three genomes (Fig. B1).

Initial tests of the PCR-RFLP sexing assay (above) on two females and two males of the golden perch, Murray cod and trout cod did not reveal any sex-specific variation in length of the amplified fragments or presence of a male-specific cut site (Appendix B, Fig. B7). We then amplified a ~400 bp region using primers conservative across the three genomes (F2 and R3, Table B1) and sequenced it on an Illumina ISeq100 system by 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 best-practice GATK v4.1.0 protocol (Van der Auwera et al., 2013), containerized and saved on the docker repository with tag:*trust1/gatk:version0.3.1* . The ~600 bp region was used as a reference (LKNJ01002546 bases 82756- 83361 for Murray cod and trout cod, VMKM01003747 bases 16365-16988 for golden perch and SEMN01000633 bases 93001-93600 for Macquarie perch). Raw reads were also examined in Geneious Pro for the presence of sex-specific haplotypes.

Results

Genome assembly and annotation for Macquarie perch and golden perch

The assembled genome for Macquarie perch (NCBI WGS Project accession SEMN01) had a size of 675.98 Mb comprising 2,962 scaffolds (scaffold N50=845 Kb; Table 1) and was estimated to be 96.9% complete, with only 74 (2.0%) out of 3,640 BUSCO groups missing within the Actinoperygii odb10 database. The Macquarie perch genome was found to be 27.67% repetitive with 33,422 predicted protein-coding genes of which 29,940 were functionally annotated by InterProScan5 with 16,264 having GO annotation.

The 661.43 Mb assembled genome for golden perch (NCBI WGS Project accession VMKM01) was less contiguous, comprising 7,165 scaffolds (scaffold N50=252 Kb; Table 1) and was estimated to be 95.9% complete, missing 85 (2.3%) out of the BUSCO groups above. The golden perch genome was found to be 25.83% repetitive with 36,108 predicted protein-coding genes of which 31,895 were functionally annotated by InterProScan5 with 16,792 having GO annotation.

Sex-linked homozygosity/heterozygosity patterns in DArT-derived SNPs for Macquarie perch and golden perch

We tested 6,939 biallelic DArT-derived SNP loci scored for 93 female and 78 male Macquarie perch, and 5,997 loci scored for 41 female and 25 male golden perch for sex-linked patterns of heterozygosity/homozygosity. Neither species had loci consistent with Y- linkage or XY- (or ZW-) gametologs as defined in Materials and Methods. An order of magnitude more sex-linked loci were found in golden perch compared to Macquarie perch, as follows. Two loci in Macquarie perch and 34 in golden perch were homozygous in males and heterozygous in 10-30% of females, consistent with being old-X-linked, and two loci in Macquarie perch and 14 in golden perch were homozygous in females and heterozygous in 10-30% of males, consistent with bearing recent-Y-specific polymorphism (Table 2, Supplementary Material S2). All putative sex-linked loci aligned to different scaffolds of their respective reference genomes, except two golden perch loci were on the same scaffold VMKM01000203, 226 Kb from each other. Heterozygotes of sex-linked SNPs were present in one sex in multiple populations, except two SNPs per species (indicated in Table 2) had heterozygotes of only one sex but only in a single population.

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

Median WGS depth coverage for 50 female and 50 male Macquarie perch was 16.3X per sample (range 11.3X-93.0X; Supplementary Material S1). For the majority (75) of samples, coverage ranged from 11X to 20X; 25 libraries for Dartmouth individuals were re-sequenced due to insufficient yield during the first run, these had a range of coverage from 32X to 57X, with one outlier of 92X coverage.

Cochran-Mantel-Haenszel (CMH) tests applied to four pooled population-by-sex samples (each comprising 25 million paired-end reads in each of 25 individuals) detected 30 SNP loci strongly (p<1e-20) differentiated between males and females in all four population-by-sex comparisons (Table 3). Of these, 10 SNPs (nine on scaffold 633, SEMN01000633, and one on SEMN01000251) had two alleles in males but only one in females (with the exception of a single count of a second allele at one SNP), and the other 20 SNPs had two alleles in each sex but with different allele frequencies between sexes.

The eleven SNP loci with the lowest CMH test p-values (p<1e-24) were located on Macquarie perch scaffold 633, within a ˜6.6-Kb region 90152-96798 (Table 3). For nine of these SNPs, all located within a ˜3.8-Kb region 90152-94017, both female pools had a single allele present (the same allele in both populations), and both male pools had two alleles (the same two in both populations; Table 3). Five of these SNPs were located within a short 146-bp fragment 93182-93327 (hereafter, the '*146-bp sexing region* ').

When the CMH test threshold was lowered to p<1e-5, there were 61,452 loci differentiated between male and female Macquarie perch pools (results in Supplementary Material S3). Of these, 1,711 loci in 644 scaffolds had one allele in female- and two in male pools, and 1,715 loci in 678 scaffolds had one allele in male- and two in female pools. Of the former set of scaffolds, scaffold 633 (length 274,905 bases), had the most loci (36) (Appendix C). Of these 36, together with three strongly (p<1e-20) differentiated loci that did not make this list because they had a single count for an alternative allele in females, 27 loci clustered within the ˜6.3 Kb region 87668-94017 of scaffold 633 (Table C1). Two loci on scaffold 633 (p<1e-5) had one allele in male- and two in female pools (Table C1).

Individual Macquarie perch genotypes for sex-linked WGS-derived loci on scaffold 633 region 90152-94017

a. Complex patterns of inheritance of Y-specific polymorphisms suggests multiple Y-linked haplotypes

We explored individual strelka genotypes for 50 females and 50 males for a 5 Kb region of scaffold 633

8

(90,000-95,000; Supplementary Material S4), focussing on loci with sex-specific alleles in the region 90152-94017 containing the nine most significantly differentiated loci (CMH test p<1e-20; Fig. C1). The region 90152-92914 contained 13 SNP-loci bearing recent-Y-specific polymorphism (Fig. C1). Of these, two (90152 and 91773) were heterozygous in >54% of males (34 and 27, respectively) and 11 were heterozygous in 14-32% of males, with the two sets of loci generally present in different sets of males (Fig. C1). The other males and all females were homozygous for the reference allele at these loci. The 146-bp sexing region 93182-93327 had five SNPs and two indels (a male-specific deletion at 93184 and insertion at 93278) inherited in a fashion consistent with XY-gametology (Table 4; Supplementary Material S4; Fig. C1). All females were homozygous for reference (X-) alleles and 35/50 males were heterozygous for this and alternative (Y-) alleles at all seven loci. The remaining 15 males included 11 homozygous for the Y-allele for one or more locus and heterozygous for the remaining ones, three homozygous for the Y-allele at two loci and homozygous for the X-allele for one locus and heterozygous for three loci, and one (MP_CBR72) homozygous for the X allele for all seven loci. For the SNP-locus 93879 bearing recent-Y-specific polymorphism, 37 males were heterozygous, one homozygous for the alternative allele, and the other 12 males and all females were homozygous for the reference allele. Lastly, SNP-locus 94017 was consistent with being XY-gametologous: all females were homozygous for the reference X-allele and all males were heterozygous, except one male (MP_CBR101) homozygous for the Y-allele. Notably, 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. Overall, females were homozygous for the reference allele at all 22 loci with sex-specific alleles, and males were heterozygous at two or more of these, indicating presence of multiple Y-haplotypes.

b. Depressed read-depth coverage indicates short-read Illumina sequencing issues

Examination of mapped mate-pair reads for the 146-bp sexing region of scaffold 633 (93182- 93327) in heterozygous 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 (Fig. 1). Near position 93250, average read-depth coverage (calculated for each base of the region 93000-94200 and normalized for each individual by average genome-wide depth before being averaged across the four sex-by-population samples) was close to the genome average, but dropped to almost zero at ~93400 (zero reads in 31 females and 28 males), then gradually returned to the genome-wide average by ~93800, except that the depth recovered earlier (at ~93650) for Dartmouth males.

c. Sex-differences in read depth coverage suggests segregating sex-linked deletions

For the loci bearing recent-Y-specific polymorphism 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. 1). Average depth for females was consistently lower than that for males from ~93250 to ~93600, containing 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. 2). In contrast, the SNP locus 93879 (bearing a recent-Y-specific polymorphism) was associated with a drop of male depth compared to that of females (Fig. 1), with male read depth for the alternative allele being just a half of that for the reference allele, suggesting deletion of Y-allele in some males (Fig. 2). The last XY-gametologous SNP locus in this region, 94017, was in a region where depth was close to the genome-wide average for both sexes (Table 4; Fig. 1).

Structural variation on Macquarie perch scaffold 633

Manta detected a total of 221 structural variants (42 insertions and 179 deletions) in 90 individuals (47 males, 43 females) spread relatively evenly through 28 positions on scaffold 633 (from 8693 to 274267; Appendix D). Mean detected indel size was 98 bp (range 50-425 bp, despite minimum default size cutoff of 8 bp). None of the structural variants were strongly sex-linked (Fig. D1). In both populations, males had more indels

than did females (Dartmouth: 82 vs 58, Yarra: 46 vs 35, respectively) but the difference between the sexes was not significant (per population Fisher test P=0.38, beta-binomial test for pooled populations: P>0.62). Overall, Manta did not show large regions that are haploid in a large proportion of individuals of one sex but not the other.

Exploring Macquarie perch and golden perch genome annotations

Forty-four Macquarie perch genes found by BRAKER2 were annotated by InterProScan5 to genes/gene families that act as sex-determining genes in other fish species, including *amh* (anti-Mullerian hormone; g1995 on SEMN01000046), *amhr2 (* anti-Mullerian hormone receptor 2 (g4926 on SEMN01000461 and g28883 on SEMN01000098), *bcar1* (breast cancer anti-resistance gene; g8496 on SEMN01000123), *brca2* (breast cancer type 2 susceptibility protein; g21038 on SEMN01000023 and g32091 on SEMN01000434)*, dmrt1* (*doublesex-* and *mab-3* -related transcription factor 1; g4593 on SEMN01000160), and *SOX* (*Sry* -related HMG-box family; 37 genes on 34 scaffolds; Appendix E, Table E1).

Macquarie perch scaffold 633 was found to contain eight genes (Table E3), one of which (g10286; bases 131251- 132300) was annotated to the gene transcription factor *SOX* (IPR022097), which has the function of regulating DNA-templated transcription (GO:0006355). This gene was located ˜38 Kb downstream of the scaffold 633 region containing 11 strongly sex-linked loci (including the 146-bp sexing region 93182- 93327; Table 3). About 41 Kb upstream of the 146-bp sexing region was a gene (g10285; bases 35783- 52002) with several domains annotated with protein binding function (GO:0005515) and guanyl-nucleotide exchange factor activity (GO:0005085). The scaffold 633 sex-linked SNP found by DArT (base 43200; Table 2) was located within intron 5 of this gene. The scaffold 633 region 52003-131250, containing the 146-bp sexing region, was not annotated to a gene. Of the other genome-wide Macquarie perch sex-linked SNPs (Tables 2 and 3), one was in a coding region (of the protein *DUF3504* of unknown function; gene g30509 on SEMN01000396), and seven were in introns of annotated genes other than known sex-determining ones (Table E4); the other SNPs were not in annotated regions.

Forty-five putative sex-determining genes were observed in golden perch annotations: *amh:* g25528 on VMKM01002284, *amhr:* g34189 on VMKM0100047038 and g6203 on VMKM01004095, *bcar1:* g9675 on VMKM01002618, *brca2:* g12532 on VMKM01000091 and g16441 on VMKM01000026*, dmrt1:* g13446 on VMKM01000340, and *SOX* genes on 36 scaffolds; Table E2). Three of 44 golden perch sex-linked DArT-SNPs mapped to the genome were in the scaffolds containing *SOX* genes, at a distance of 67-635 Kb from these genes (Table E5).

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 (C/C homozygote in females, C/T heterozygote in males based on WGS data) showed a single band in females, consistent with XX genotype, and three bands in males, consistent with uncut X allele and two fragments of Y-allele cut at the Y-specific SNP, as expected (Appendix B; Fig. B2). 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 (Fig. B3). The lack of the uncut X-specific band in digested PCR product for some of these males was unexpected, because five males had WGS data and of these three showed C/T heterozygotes for the cut-site (although all five displayed Y0 genotype for >3 of other male-specific loci in the 146 bp sexing region). The PCR-RFLP assay also revealed length variation for all bands, likely reflecting variation at microsatellite repeat number. Despite this variation, and WGS showing a single male out of 50 being homozygous for a putative X-allele at 93229, the sexing assay appears promising in these two populations.

The same assay tested in known-sex Macquarie perch from King Parrot Creek/Holland's Ck and Abercrombie populations showed that males consistently yielded 2- or 3-band patterns, but so did 20% of King Parrot Creek females and 60% of Abercrombie ones, indicating the presence of the T allele at 93229. The remaining females showed a single band, expected from homozygotes for the X-linked C allele at 93229 (Appendix B,

10

Fig. B4). Abercrombie 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 on the PCR-RFLP assay (out of ten that yielded bands), two produced eggs that were successfully fertilized in 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 eggs fertility rate of <1%, and one had eggs which were successfully fertilized. The same PCR-RFLP assay tested 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 (Appendix B, Fig. B5).

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 F, Table F1). Amplicon sequences for the two individuals from coastal populations (Table F1) showed that only variation at 93229 (the site chosen for the PCR-RFLP sexing assay) might be consistently sex-linked across sequenced individuals. 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 F2), Murray cod or trout cod (Table F3).

Discussion

Using two newly assembled genomes, reduced-representation sequencing (DArTseq) and whole-genome re-sequencing (WGS) for panels of individuals of known sex, and amplicon sequencing, we aimed to identify sex-linked genomic regions and develop and test molecular sexing markers in two *Macquaria* perch sister-species along with two confamilial species. One of these species, golden perch, is inferred by karyology to have XY sex-determination and homomorphic sex chromosomes, and the confamilial Murray cod to have XY and heteromorphic chromosomes (Shams et al., 2019). We did not uncover DArTseq SNPs consistent with Y-linkage or XY-gametology in either golden perch or Macquarie perch, but found loci polymorphic only in one sex (sometimes males, sometimes females), scattered through different scaffolds. This suggests that regions with old strata (with long-term lack of recombination), if present on sex-chromosomes of these species, are short. Finding an order of magnitude more sex-linked DArT loci in golden perch than in Macquarie perch, none of which were on the golden perch scaffold VMKM01003747 homologous to Macquarie perch scaffold 633, indicates that these species have independently-evolved sex-determination systems, with the Macquarie perch one being more recent. Alternatively, Macquarie perch could have lost sex-linked variation during the Pleistocene bottleneck involved in colonization of the Murray Darling Basin from the coastal range or during recent range contraction and population isolation (Pavlova, Gan, et al., 2017). DArTseq has been used to find sex-linked loci and infer sex-determination systems, including in catfish *Clarias gariepinus* with homomorphic sex-chromosomes, where both XY and ZW sex-determination were inferred based on the presence of moderately male- and moderately female-linked loci (Nguyen et al., 2021).

Exploring male and female population pools of WGS data for Macquarie perch with CMH tests guided us to focus on scaffold 633 region 90152-94017, enriched with sex-linked loci, including XY-gametologous region 93182-94017, encompassing five SNPs and two indels in the 146-bp sexing region 93182-93327, and an additional SNP at 94017. Female pools had a single allele and male pools had two approximately equally represented alleles for these six SNPs. Individual WGS genotypes showed all females to be homozygous and 70% of males heterozygous at these loci. Rare past or ongoing recombination within this XY-gametologous region was suggested by 30% of males being homozygous for X- or Y- alleles for up to seven of these eight loci. Comparisons of male and female read depth for sex-linked loci and separately for sex-linked alleles (Figs. 1, 2) suggested that some females and males might be haploid for part of this XY-gametologous region (i.e. through a deletion on one of the X-variants). Whereas similar read depth for males and females for the first and last loci in this region (93182 and 94017) implies diploidy of these SNP loci in both sexes, for intervening loci 93229, 93299, 93315 and 93327 most females appeared haploid (X0) for this region: females had consistently lower depth than males (Fig. 1), and similar depth for X-alleles as each of the X- and Y-

11

alleles in males (Fig. 2). Lower average X-allele depth in males than the Y-allele suggests that some males could be haploid for Y (i.e. Y0).

Upstream of the XY-gametologous region 93182-94017 was the region 90152-92914 containing 13 SNPs bearing recent-Y-specific polymorphism, which might be a more recently emerging region of sex-difference. Patterns of appearance of these recent-Y-specific polymorphisms in males (Fig. C1) suggest that multiple occasionally recombining Y-haplotypes are segregating in Macquarie perch. The XY-gametologous region itself included another recent-Y-specific SNP 93879, suggesting occasional recombination between X and Y haplotypes (Fig. C1). The lower read depth for 93879 in males compared to females suggested Y-specific deletions at this locus (Figs. 1 and 2). Manta analysis did not detect any structural variants that would support X- or Y- deletion polymorphism consistent with these patterns (Fig. D1), however our read depth coverage (~10-20x) was much lower than the ~50x used in original tests of Manta (Chen et al., 2016), potentially underpowering our analysis.

Overall, in the panel of 22 loci with sex-biased patterns of heterozygosity in scaffold 633 region 90152-94017 (8 XY-gametologs and 14 loci bearing recent-Y-specific polymorphism; Fig. C1), screened in 100 individuals from Dartmouth and Yarra, all females were homozygous for X-alleles, and Y-alleles were present in 100% of males for 94017, 98% for 93229, 93299 and 93327, 96% of for 93182, 93184, 93278 and 93315, and in 14%-76% for the other loci. This suggests a working hypothesis of sex-determination in which only individuals homozygous at the X-alleles at all 22 of these loci develop a female phenotype, whereas presence of Y-alleles at some of the loci leads to a male phenotype. However, presence of the Y-allele at 93229 in 36% of King Parrot Creek and Abercrombie females is not consistent with this mechanism based on the Dartmouth and Yarra populations (Fig. B4). This single deviation from the suggested model call for other mechanisms of producing a female phenotype.

Tests of our molecular sexing protocol, that targeted the Y-allele of 93229 in the 146-bp sexing region of Macquarie perch scaffold 633, in other populations of Macquarie perch, in golden perch, and in more distant relatives from the family Percichthyidae (Murray cod and trout cod) showed that this sexing region is Macquarie perch-specific, and could even be limited to a subset of the species. This Y-allele was present in all but one male out of 50 tested, hence the test should correctly identify Dartmouth and Yarra individuals as males (based on two or three bands on a gel) or females (based on one band), except ~2% of males may be missexed as females. The same test is not expected to work well in other populations of Macquarie perch, because 20% of putative females from King Parrot Creek and 60% from Abercrombie had the male-specific allele. Unless these individuals are feminized (egg-producing) males, female homozygosity for the reference allele at 93299 might be limited to some populations. Unique variation at these loci in amplicon sequencing for two coastal individuals of unknown sex further suggests that sex-specific variation in Macquarie perch could be specific to some populations. Intraspecific variation in chromosomal locations of sex-determining loci occurs in other vertebrates (Kijas et al., 2018a; Lubieniecki et al., 2015; Rodrigues, Merila, Patrelle, & Perrin, 2014). In the spotted snow skink*Niveoscincus ocellatus* , highland and lowland populations, which differ in their degree of influence of environmental temperature on sex, shared- and population-specific sex-linked loci were detected, suggesting a different degree of sex-chromosome differentiation between populations (Hill, Burridge, Ezaz, & Wapstra, 2018). 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 (*Danio rerio*) 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) which evolved de-novo after loss of W-alleles (Wilson et al., 2014).

Understanding the nature of homo/hemizygosity at the XY-gametologous region of scaffold 633 was obscured here by the drop in read depth for the sex-linked region compared to the genome-average, including zero coverage around position 93400 for 59% of individuals regardless of their overall sequence depth (Fig. 1). This lower depth might be linked to difficulty of sequencing through a low-complexity region, including the $(GT)_n$ microsatellite that ends around 93193 (Fig. B1). Accumulation of repetitive sequences was previously reported for golden perch and Murray cod sex-chromosomes (Shams et al., 2019). Given its high read-depth

coverage, amplicon sequencing may provide a more reliable way of genotyping genomic regions that contain low-complexity fragments. Despite the locally reduced read depth in WGS, WGS-based genotypes were identical to amplicon-based genotypes for the four individuals analysed with both techniques. Apparently real Y-allele homozygotes were observed even for males with reasonable WGS depth, including a male for which amplicon sequencing was done. Low depth also did not preclude detecting sex-specific and population-specific patterns of depth coverage (Fig. 2). Thus, an X0 or Y0 genotypes for some XY-homologous loci in some individuals likely represents a biological phenomenon. It is unlikely that the drop in depth for the scaffold 633 sexing region was due to failed mapping due to repetitive regions: BLAST search of the 146-bp sexing region did not find highly similar regions in other Macquarie perch scaffolds.

The scaffold 633 sex-linked region 90152-94017 was not annotated to a known gene, but was located ˜39 Kb upstream of gene of*SOX* transcription factor family, and 41 Kb downstream of a protein with *Dbl* homology (DH) domain superfamily, associated with guanyl-nucleotide exchange factor activity and protein binding (interacting selectively with any protein or protein complex).*SOX* genes regulate fate of stem and progenitor cells during development (Sarkar & Hochedlinger, 2013). Three genes in particular,*SOX9* , *SOX8* (Koopman, 2005) and *SOX3* , a progenitor of the *SRY* gene in mammals (Graves, 2013), have a role in male sex determination in different species, with their expression inducing testes development. Thus, it is possible that *SOX* gene g10286, modulated by the variation within the sex-linked region 90152-94017, controls gonadal development in Macquarie perch.

Alternatively, polygenic sex-determination in Macquarie perch is possible, given the many loci detected as putatively female- and male-linked by the CMH tests (at p<1e-5; Supplementary Material S3). 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, Bohne, Schartl, Volff, & Galiana-Arnoux, 2009). Future studies are needed to demonstrate whether sequence variation at the sex-linked region on Macquarie perch scaffold 633 has mechanistic links with development of female or male phenotypes. Acquiring WGS data for known-sex samples from other populations will be crucial for improving our understanding of sex-determination in this species. Sex-determination in *Macquaria*appears to be complex and dynamic, potentially involving multiple sites or genomic regions and different mechanisms in quite closely-related lineages, including within species.

Acknowledgements

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).

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., Elaswad, 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.

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*

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.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., . . . Saunders, C. T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics, 32* (8), 1220-1222.

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.

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.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences, 117* (17), 9451-9457.

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.

Hill, P. L., Burridge, 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.

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.

Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research, 29* (4), e25-e25.

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.

Kijas, J., McWilliam, S., Sanchez, M. N., Kube, P., King, H., Evans, B., . . . Verbyla, K. (2018a). Evolution of Sex Determination Loci in Atlantic Salmon. *Scientific Reports, 8* (1), 5664.

Kijas, J., McWilliam, S., Sanchez, M. N., Kube, P., King, H., Evans, B., . . . Verbyla, K. (2018b). Evolution of sex determination loci in Atlantic salmon. *Scientific Reports, 8* (1), 1-11.

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., Kallberg, M., . . . Krusche, P. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods, 15* (8), 591.

Kofler, R., Pandey, R. V., & Schlotterer, 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.

Lavoue, 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.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25* (14), 1754-1760.

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. (2020). Using multiple sources during reintroduction of a locally extinct population benefits survival and reproduction of an endangered freshwater fish. *Evolutionary Applications, 10.1111/eva.13173, published online 22 November 2020* . 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.

Martinez, P., Vinas, A. M., Sanchez, L., Diaz, 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

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

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

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.

Pavlova, A., Gan, H. M., Lee, Y. P., Austin, C. M., Gilligan, D., Lintermans, M., & Sunnucks, P. (2017). Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. *Heredity, 118* , 466–476. doi:doi:10.1038/hdy.2016.120

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.

Pham, T. V., & Jimenez, C. R. (2012). An accurate paired sample test for count data. *Bioinformatics, 28* (18), i596-i602.

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.

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.

Schultheis, C., Bohne, A., Schartl, M., Volff, J., & Galiana-Arnoux, D. (2009). Sex determination diversity and sex chromosome evolution in poeciliid fish. *Sexual Development, 3* (2-3), 68-77.

Seppey, 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.

Smit, A., Hubley, R., & Green, P. (2013-2015). RepeatMasker Open-4.0 <http://www.repeatmasker.org>.

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.

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.

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.

	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

	Macquarie perch	golden perch
Number of contigs	3,369 contigs	7,352 contigs
GenBank assembly accession	GCA_005408345.1	GCA_0083609
WGS Project accession	SEMN01	VMKM01
Whole Genome Shotgun accession	SEMN00000000.1	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/zeno

Table 2. DArT-derived SNP loci with sex-biased patterns of heterozygosity for Macquarie perch (MP; 93 females and 78 males) and golden perch (GP; 41 females and 25 males; see Supplementary Material S2 for more information). Old-X-linked loci- those homozygous in all males but heterozygous in 10% or more females; loci bearing recent-Y-specific polymorphism- homozygous in all females but heterozygous in 10% or more males. For each DArT locus mapped to a reference genome, species abbreviation, scaffold ID and coordinate are given. For all loci, count of homozygous and heterozygous individuals in female (F) and male (M) sample is given (see Supplementary Material S2 for details): F0 ref hom/M0 ref hom = number of females/males homozygous for a reference allele, F1 het/M1 het = number of females/males that are heterozygous, F2 alt hom/M2 alt hom = number of females/males that are homozygous for the alternative allele. MP scaffold 633 loci are in bold; the two GP scaffold 203 loci are in bold italic. Asterisks indicate loci for which heterozygotes were present in only one sex of a single population. Loci are sorted by scaffold ID (four golden perch DArT-tags did not align to the golden perch reference genome).

Dart locus ID	NCBI reference genome scaffold ID	Scaffold position of the locus se
Macquarie perch		
Old-X-linked loci	*Old-X-linked loci*	*Old-X-linked loci*
36356960-12-G/A	SEMN01000004	1395201
36356824-43-G/A	SEMN01000754	59288
Macquarie perch	Macquarie perch	Macquarie perch
Loci bearing recent-Y-specific polymorphism	*Loci bearing recent-Y-specific polymorphism*	*Loci bearing recent-Y-specific p*
28931412-11-C/T*	SEMN01000020	1691939
17035690-65-T/A*	**SEMN01000633**	**43200**
Golden perch		
Old-X-linked loci	*Old-X-linked loci*	*Old-X-linked loci*
28619580-14-C/T	VMKM01001065	173435
28618482-9-G/A	VMKM01001081	157581
33707197-25-A/G	VMKM01001101	21759
28618575-31-A/T	VMKM01001181	130180
28931872-6-A/G	VMKM01001374	135276
37096050-57-C/T	VMKM01001533	7572
28618539-6-C/T	VMKM01001928	18984
28619655-25-A/T	VMKM01001960	71667

18

Dart locus ID	NCBI reference genome scaffold ID	Scaffold position of the locus se
37096062-11-G/A	VMKM01000203	476876
37096215-12-T/A	VMKM01000203	251318
28619537-13-C/A	VMKM01002203	64254
33707200-37-T/C	VMKM01002345	52353
37096174-14-C/T*	VMKM01002499	24594
100010491-66-G/T	VMKM01002812	29291
37096320-25-C/A	VMKM01002906	27593
28618166-37-C/T	VMKM01000328	288497
20266862-57-T/A	VMKM01000035	589065
37096511-30-G/A	VMKM01000378	112501
17036577-54-A/G	VMKM01000394	15044
37095925-14-T/A	VMKM01004054	18850
28620041-44-A/G	VMKM01004724	1909
37095883-61-C/T	VMKM01000489	168594
37095970-57-G/T	VMKM01000526	174302
28619004-10-A/C	VMKM01000064	749908
28618065-41-T/G	VMKM01000738	28036
28618085-27-G/A	VMKM01000739	213128
17039139-40-T/C	VMKM01000770	200871
17038132-46-A/T	VMKM01000877	142832
37096210-58-C/T	VMKM01000097	312116
28619737-22-G/A	VMKM01000979	76143
28618970-31-T/C	VMKM01000987	3422
28619462-17-T/A		
20268010-19-T/A		
28618437-13-T/A		
Golden perch	Golden perch	Golden perch
Loci bearing recent-Y-specific polymorphism	*Loci bearing recent-Y-specific polymorphism*	*Loci bearing recent-Y-specific p*
37096522-21-G/A	VMKM01000001	139620
100008933-5-C/A*	VMKM01000010	1113651
37096713-34-G/T	VMKM01001354	147539
100025231-56-G/A	VMKM01000160	325681
20265649-22-C/T	VMKM01000017	301933
20267388-24-G/A	VMKM01000212	50324
28619198-20-C/A	VMKM01002530	26259
17038127-64-C/A	VMKM01003059	2935
100026247-65-T/A	VMKM01003205	30436
20267873-5-G/C	VMKM01000323	268476
28619579-52-G/A	VMKM01000445	132713
20268621-32-A/C	VMKM01000475	99375
33707406-50-T/C	VMKM01000991	41923
28619935-62-T/A		

Table 3. Allele counts for WGS-derived Macquarie perch loci strongly (P<1e-20) differentiated between male and female pools, detected by Cochran-Mantel-Haenszel test in *PoPoolation2* ; loci sorted by test p-value. Each pool contains 25 million reads from each of 25 Dartmouth females, 25 Dartmouth males, 25 Yarra females or 25 Yarra males. Loci with just one allele in female pools (one error allowed) but two alleles in male pools are in italic. Scaffold 633 loci are shaded grey, five SNP located within the 146-bp sexing region 93182-93327, are in bold.

19

Macquarie perch Scaffold ID	Scaffold coordinate	Dartmouth Females	Dartmouth Males	Yarra Females	Yarra Males	CMH p-value cutoff
		A:T:C:G:N:indel	A:T:C:G:N:indel	A:T:C:G:N:indel	A:T:C:G:N:indel	
SEMN01000633	*94017*	*90:0:0:0:0:0*	*33:0:0:24:0:0*	*92:0:1:0:0:0*	*47:0:0:34:0:0*	*2.1971E-41*
SEMN01000633	93229	0:0:46:0:0:0	0:34:28:0:0:0	0:0:71:0:0:0	0:35:28:0:0:0	5.156E-40
SEMN01000633	93299	0:42:0:0:0:0	0:22:36:0:0:0	0:47:0:0:0:0	0:15:33:0:0:0	5.5909E-40
SEMN01000633	93315	0:0:0:38:0:0	0:36:0:20:0:0	0:0:0:45:0:0	0:33:0:15:0:0	7.5488E-39
SEMN01000633	93327	0:0:0:33:0:0	36:0:0:20:0:0	0:0:0:41:0:0	35:0:0:14:0:0	5.1401E-37
SEMN01000633	*90152*	*0:0:74:0:0:0*	*0:0:45:22:0:0*	*1:0:88:0:0:0*	*0:0:40:36:0:0*	*2.9651E-36*
SEMN01000633	93182	0:64:0:0:0:0	32:29:0:0:0:0	0:82:0:0:0:0	20:42:0:0:0:0	1.0476E-34
SEMN01000633	93727	0:0:56:8:0:0	0:0:25:49:0:0	0:0:54:14:0:0	0:0:18:44:0:0	3.8445E-33
SEMN01000633	*91773*	*0:0:0:95:0:0*	*18:0:0:73:0:0*	*0:0:0:64:0:0*	*41:0:0:58:0:0*	*3.4415E-27*
SEMN01000633	96798	10:0:0:42:0:0	51:0:0:22:0:2	25:0:0:58:0:1	52:0:0:19:0:0	8.8577E-27
SEMN01000633	*93879*	*0:1:0:81:0:0*	*14:0:0:35:0:0*	*0:0:0:78:0:0*	*17:0:0:35:0:0*	*4.6027E-25*
SEMN01000508	32899	81:0:0:6:0:0	59:0:0:42:0:0	83:0:0:18:0:0	43:0:0:47:0:0	2.1701E-24
SEMN01000409	363800	68:0:22:0:0:0	28:0:59:0:0:0	37:0:42:0:0:0	15:0:79:0:0:0	3.1558E-24
SEMN01000110	836251	40:0:61:0:0:0	18:0:88:0:0:0	54:0:51:0:0:0	10:0:93:0:0:0	4.9314E-24
SEMN01000396	409936	0:86:6:0:0:0	0:41:28:0:0:0	0:80:9:0:0:0	0:42:33:0:0:0	8.3177E-24
SEMN01000251	*399756*	*0:0:0:87:0:0*	*11:0:0:65:0:0*	*1:0:0:81:0:0*	*33:0:0:50:0:0*	*1.5484E-23*
SEMN01000130	722926	21:1:0:89:0:0	47:0:0:36:0:0	42:0:0:40:0:0	70:0:0:13:0:0	5.5955E-23
SEMN01000282	465844	0:53:24:0:0:0	0:87:5:0:0:0	0:59:48:0:0:0	0:74:7:0:0:0	6.4122E-23
SEMN01000351	305726	7:0:0:75:0:0	14:0:0:60:0:0	12:0:0:93:0:0	56:0:0:33:0:0	7.2449E-23
SEMN01000248	99134	89:11:0:0:0:0	67:20:0:0:0:0	84:8:0:0:0:0	47:59:0:0:0:0	2.4506E-22
SEMN01000105	280957	77:0:0:8:0:0	64:0:0:32:0:0	85:0:0:5:0:0	52:0:0:39:0:0	5.3869E-22
SEMN01000020	1944824	69:0:0:30:0:0	19:0:0:77:0:0	9:0:0:69:0:0	4:0:0:89:0:0	6.5797E-22
SEMN01000338	493881	17:80:0:0:0:0	52:44:0:0:0:0	7:88:0:0:0:0	29:58:0:1:0:0	8.6443E-22
SEMN01000017	1931657	65:0:0:40:0:0	22:0:0:92:0:0	56:0:0:18:0:0	44:0:0:45:0:0	1.0292E-21
SEMN01000038	657608	72:0:0:32:0:0	33:1:0:62:0:0	73:0:0:19:0:0	42:0:0:52:0:0	1.0428E-21
SEMN01000055	877605	0:31:50:0:0:0	0:14:76:0:0:0	0:80:14:0:0:0	0:29:43:0:0:0	2.4059E-21
SEMN01000081	477061	23:0:0:81:0:0	1:0:0:68:0:0	34:0:0:46:0:0	5:0:0:89:0:0	2.6182E-21
SEMN01000713	12203	0:15:0:78:0:0	0:40:0:53:0:0	0:15:0:89:0:0	0:54:0:54:0:0	3.0293E-21
SEMN01000442	300259	0:63:34:0:0:0	0:26:66:0:0:0	0:54:30:0:0:0	0:29:65:0:0:0	4.7816E-21
SEMN01001079	97477	15:0:84:0:0:0	33:0:70:0:0:0	2:0:97:0:0:0	39:0:55:0:0:0	7.1893E-21

Table 4. Summary of Macquarie perch individual-based strelka genotypes for nine WGS-derived SNPs with male-specific alleles (CMH test p<1e-20), plus two sex-linked indels within the 146-bp sexing region of scaffold 633 (SEMN01000633) bases 93182-93327 (this region is shown in bold). F0 ref hom/M0 ref hom = number of females/males homozygous for reference allele, F1 het/M1 het = number of females/males that are heterozygous, F2 alt hom/M2 alt hom = number of females/males that are homozygous for the alternative SNP allele. When more than one alternative allele was present, M2 alt hom is the count of homozygotes at the more common alternative allele. Loci are sorted by scaffold number/position in a scaffold. Sex-linkage pattern: recent-Y- loci bearing recent-Y-specific polymorphism, XY- XY-gametologs, Y0- in males Y allele is present but X is not (due to low coverage or deletion).

Scaffold 633 position	Ref allele	Alt alleles	F0 ref hom	F1 het	F2 alt hom	M0 ref hom	M1 h
90152	C	G	50	0	0	16	34
91773	G	A, T	50	0	0	23	27
93182	**T**	**A**	**50**	**0**	**0**	**2**	**47**
93184	**TGTGTGTGTGA**	**T, TGTGTGT**	**50**	**0**	**0**	**2**	**44**

Scaffold 633 position	Ref allele	Alt alleles	F0 ref hom	F1 het	F2 alt hom	M0 ref hom	M1 h
93229+	**C**	**T**	**50**	**0**	**0**	**1**	**46**
93278	**C**	**CTA**	**50**	**0**	**0**	**2**	**42**
93299	**T**	**C**	**50**	**0**	**0**	**1**	**42**
93315	**G**	**T**	**48++**	**0**	**0**	**2**	**40**
93327	**G**	**A**	**48++**	**0**	**0**	**1**	**41**
93879	G	A	50	0	0	12	37
94017	A	G	50	0	0	0	49

+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
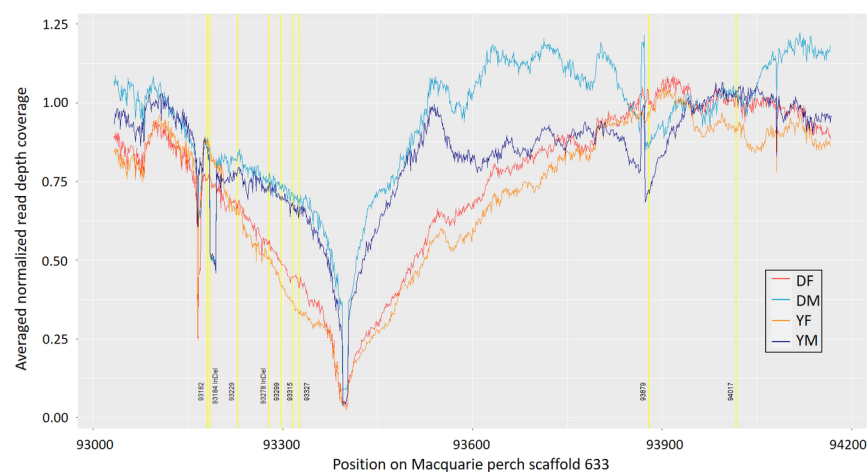
Figure 1 . Normalized WGS read depth coverage for four population-by-sex samples of individuals (Y-axis) along sex-linked region of Macquarie perch scaffold 633 (SEMN0100063, bases 93050-94150 (X-axis). Dartmouth females- DF (red), Yarra females- YF (orange), Dartmouth males- DM (light blue), Yarra males-YM (dark blue). Vertical lines indicate location of nine sex-linked loci located in this region (yellow- SNP, orange- indels; see Table 4). 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 2 . 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-homologs, except 93879 is a locus bearing recent-Y-specific polymorphism. PCR-RFLP sexing assay targets the alternative allele at 93229.