Journal of Heredity, 2026, **00**, 1–10
https://doi.org/10.1093/jhered/esag054
Published: 10 July 2026
Genome Resources

Chromosome-level genome assemblies of Southern pygmy perch (*Nannoperca australis*) and Yarra pygmy perch (*N. obscura*) reveal several miniaturization genes

Jonathan Sandoval-Castillo*, Luciano B. Beheregaray

Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Bedford Park, SA, Australia

*Corresponding author: Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Bedford Park, 5042, SA, Australia.
Email: jonathan.sandoval-castillo@flinders.edu.au

Abstract

Freshwater biodiversity is declining at alarming rates, and the biological characteristics of small freshwater fishes make them particularly vulnerable to habitat degradation and fragmentation. Pygmy perches are small Australian fishes whose populations have recently undergone rapid declines due to human-induced pressures. They also exhibit paedomorphic miniaturization, an evolutionary phenomenon whose consequences for the persistence and diversification of freshwater fishes remain unclear. Genomic resources for pygmy perches would significantly enhance both conservation efforts and evolutionary research. Here, we describe chromosome-level genome assemblies for the Southern pygmy perch (*Nannoperca australis*) and the Yarra pygmy perch (*N. obscura*). We provide evidence for high synteny among percichthyid genomes and for positive selection on size- and growth-related genes in the pygmy perch lineage. The *N. australis* and *N. obscura* assemblies are 675.7 and 653.9 Mbp in length, with scaffold N50 values of 26.98 and 26.22 Mbp, respectively. Each assembly comprises 24 pseudo-chromosomes and 890 or 1,476 contigs, with repeat contents of 22.3% and 28.1%. We predicted 26,390 and 24,654 protein-coding genes, and BUSCO completeness scores for Teleostei genes were 96.9% and 96.5%, respectively, for *N. australis* and *N. obscura*. Their genomes are comparable in quality to the best currently available percichthyid assemblies. These genomic resources will support conservation management efforts and provide a foundation for comparative studies on the adaptive significance of miniaturization.

Keywords conservation genetics, genetic rescue, landscape genomics, Murray-Darling basin, Percichthyidae, teleost.

Introduction

Worldwide, freshwater ecosystems are increasingly threatened by human activities, and their biodiversity is declining at a much faster rate than that of other ecosystems (Vörösmarty et al. 2010; Albert et al. 2021). Despite growing awareness, public attention and conservation funding continue to focus on large, commercially important, or charismatic species (Darwall et al. 2011; Donaldson et al. 2016; Todd et al. 2017). This bias is concerning because most freshwater species are small, and their direct value to humans is often unknown. Small-bodied species also tend to be more prone to extinction due to limited dispersal, reduced migration potential, and heightened sensitivity to habitat fragmentation (Saddlier et al. 2013; Lintermans et al. 2020). This is particularly true in Australia, where 70% of native freshwater fishes reach >200 mm in length (Allen et al. 2002) and face multiple threats including artificial barriers (Coleman et al. 2018; Brauer and Beheregaray 2020), invasive species (Barrett et al. 2014), climate change (Lennox et al. 2019), and habitat disturbance (Tonkin et al. 2016; Thiem et al. 2020). Developing genomic and ecological resources for the conservation of small freshwater fishes is therefore essential.

Australian pygmy perches (genera *Nannoperca* and *Nannatherina*) are small (up to 85 mm) fishes inhabiting vegetated margins of slow-flowing streams and wetlands in temperate southern Australia. Unlike larger recreational or commercially important fishes, pygmy perches have historically attracted little attention and in the past were largely overlooked in conservation management (Saddlier et al. 2013). Their biological characteristics and ecological requirements make them particularly susceptible to extended droughts and habitat degradation (Hammer et al. 2013; Beheregaray et al. 2021). The last decade has seen an upsurge in ex-situ and in-situ conservation efforts for pygmy perch species, including captive-breeding, reintroduction and monitoring of threatened populations, particularly in Southeastern's Australia highly impacted Murray-Darling Basin (Hammer et al. 2013; Attard et al.

Received: 31 December 2025. **Revised:** 29 May 2026. **Accepted:** 7 July 2026

2016a; Attard et al. 2016b; Marshall et al. 2022; Buckley et al. 2024a). Seven described and at least two cryptic species of pygmy perches are currently recognized (Buckley et al. 2018; Buckley et al. 2024b). Although only three species are listed as endangered by the IUCN, six species are considered endangered at state or national level (Lintermans et al. 2020; Beheregaray et al. 2021), including the Southern pygmy perch (*Nannoperca australis*) and the Yarra pygmy perch (*N. obscura*).

Pygmy perches exhibit several reductive morphological specializations, including paedomorphic miniaturization, such as the absence of a subocular shelf, reduced numbers of precaudal rays and dorsal spines, and an interrupted lateral line (Johnson 1984; Hanken and Wake 1993). Miniaturization, the evolution of sexually mature individuals with small adult body size, often involves reductions in structures or organs and has significant ecological and evolutionary consequences (Hanken and Wake 1993; Blanckenhorn 2000; Biswas and Karanth 2025). It is widespread in animals, including multiple lineages of freshwater fishes (Esin et al. 2020), and frequently emerges under stressful environmental conditions such as limited resources (MacColl et al. 2013) or extreme water parameters (Cooke et al. 2012; Esin et al. 2020). However, the processes shaping the evolution of miniaturization and its genetic basis remain poorly understood (Blanckenhorn 2000; Troyer et al. 2025), limiting our ability to infer its role in lineage diversification and persistence.

Here, we describe pseudochromosome-level genome assemblies for *N. australis* and *N. obscura*. We generated hybrid assemblies by combining Illumina short reads, PacBio long reads, and a previously published transcriptome (Brauer et al. 2017). We also developed a multi-step bioinformatic pipeline for genome construction, gene prediction, and repeat annotation. These chromosome-level assemblies are expected to provide essential resources for conservation programs, such as captive breeding and genetic rescue, targeting declining species. They are also expected to help clarifying how human impacts affect evolutionary potential in small, isolated populations. Additionally, we conducted comparative phylogenetic analyses to identify signatures of positive selection in genes associated with miniaturization. These results provide candidate genes for future research on the evolution of reduced body size and diversification in freshwater fishes.

Materials and methods

Sample collection and sequencing

A male individual of each species was sampled from the captive breeding program at Flinders University (for details about the program, see Attard et al. 2016a; Marshall et al. 2022). We chose inbred individuals (i.e. more than eight generations in captivity) with the aim of reducing heterozygosity and simplifying genome assembly. The parents of the *N. australis* individual originated from the Ovens River, whereas the parents of the *N. obscura* individual came from the Lower Lakes, both within the Murray-Darling Basin. Individuals were humanely euthanized using an overdose of clove oil. Fresh muscle, skin, fin, gonad, liver, and brain tissues were collected, preserved in 100% ethanol, and rapidly stored at −80 °C. Total genomic DNA was extracted from muscle tissue using a salting-out protocol (Sunnucks and Hales 1996), and impurities were removed using AMPure XP magnetic beads at a 0.8 bead-to-DNA volume ratio. For *N. australis*, 1 μg of DNA was shipped on dry ice to Novogene, where a short-read library (350 bp) was prepared and sequenced (150 bp PE) on an Illumina HiSeq 4000 platform. Additionally, 500 ng of DNA were sent on dry ice to the Centre for Genome Research and Biocomputing at Oregon State University, where a long-read (20 kb) library was prepared and sequenced on one SMRT cell using the PacBio Sequel II platform. For *N. obscura*, 1 μg of DNA was shipped on dry ice to the Australian Genome Research Facility, where a long-read (20 kb) library was prepared and sequenced on a PacBio Revio platform in HiFi mode on a single-SMRT cell.

Genome assemblies

For the *N. australis* assembly, Illumina reads with low average quality ($Q < 20$), excess of "N" (>10%), and small length (>60 bp) were filtered out, and adapters were trimmed using TRIMMOMATIC 3.9 (Bolger et al. 2014). Because PacBio reads have relative high error rate, especially around indels, the high coverage short reads were used to correct sequencing errors in long reads with LORDEC 0.9 (Salmela and Rivals 2014). We then generated contig-level assembly from the Illumina short read using SOAPDENOVO2 2.4.2 (Luo et al. 2012) without gap closing or scaffolding steps. These contigs, together with the corrected PacBio long reads, were inputted into DBG2OLC (Ye et al. 2016) to create a hybrid assembly. This software uses accurate short read contigs to anchor and compress reads, performs multiple sequence alignments, removes reads with structural errors (chimeras), and constructs an assembly backbone. Finally, uncompressed long reads are aligned to the backbone, and the most likely consensus genome is generated. This consensus was improved using the paired-end RNA sequencing reads previously generated for *N. australis* (Brauer et al. 2017), and P_RNA_SCAFFOLDER (Zhu et al. 2018). In this step, paired RNA reads are first aligned to the consensus contigs, pair reads in which both reads map to the same contig or in which at least one read aligns to several contigs, are removed. The remaining RNA read pairs are used to guide the connection, ordering, and orientation of consensus contigs. A scaffolding graph is then created by linking optimal connections and inserting a number of "N" equal to the median intron size at each junction. Assembly polishing was performed in three steps. First closing gaps using corrected long reads and TGS-GAPCLOSER 1.1.1 (Xu et al. 2020), followed by two rounds of scaffold polishing with PILON 1.23 (Walker et al. 2014) and the cleaned short reads. For the *N. obscura* we did not have short Illumina reads, but we had HiFi long reads, which have substantially lower error rates than conventional long reads. Thus, we used the PacBio pipeline, Improved Phased Assembly HiFi Assembler (Sovic et al. 2022), to generate an initial assembly. Both species assemblies were further improved by aligning the scaffolds to the closest related chromosome-level genomes available: Macquarie perch (*Macquaria australasica*; Pavlova et al. 2022), golden perch (*Maquaria ambigua*; Pavlova et al. 2022), Murray cod (*Maccullochella peelii*; Austin et al. 2017), and Mandarin perch (*Siniperca chuatsi*; He et al. 2020) in RAGOUT 2.3 (Kolmogorov et al. 2018). This approach reconstructs the most parsimonious chromosome structure for the target genome by analyzing structural rearrangements (synteny breaks) among reference genomes. The resulting pseudo-chromosome-level assemblies were polished again using the three-step described above for the *N. australis* scaffold-level assembly.

Annotation of repetitive elements and gene prediction

For each separate assembly, a de novo identification of repeat elements was implemented using REPEATMODELER 2.0 (Flynn et al. 2020), and both genomes were annotated to their respective repeat library using REPEATMASKER 4.1 (Tarailo-Graovac and Chen 2009). For gene prediction and annotation, we applied a three-evidence based strategy using the MAKER 3.01.03 (Campbell et al. 2014) pipeline. As transcriptome evidence, we used the published *N. australis* assembly (Brauer et al. 2017), a newly generated assembly of *N. obscura*, and transcriptomes of two other closely related species (*M. australasica* and *M. ambigua*; Pavlova et al. 2022). As protein evidence, we used a Teleostei database from Uniprot (2024). Both transcript and protein evidence were aligned to the genome using EXONERATE (Slater and Birney 2005) and BLAST+ 2.10.1 (Camacho et al. 2009). An ab initio gene prediction was generated with SNAP (Korf 2004) and AUGUSTUS 3.4 (Stanke et al. 2008), both trained with all protein, transcriptome, and repeat alignments. MAKER then integrates all evidence to find the best gene model for the genome. The genes detected were functionally annotated, and gene ontology (GO) terms were assigned using the SwissProt database (2024).

Assembly quality control

The quality of the genome assemblies was assessed using several complementary approaches. Expected genome size and heterozygosity were estimated using the k-mer–based statistical method implemented in GENOMESCOPE2 (Ranallo-Benavidez et al. 2020). K-mer frequency histograms were generated from the raw short-read data using JELLYFISH 2.3 (Marçais and Kingsford 2011). Basic assembly statistics, including assembly size, number of contigs, longest scaffold, N50, L50, number of supper scaffolds (chromosomes), N's, and GC content, were calculated using the program QUAST 5.0.2 (Gurevich et al. 2013). Genome completeness was evaluated using BUSCO 5.2 (Manni et al. 2021) with Actinopterygii lineage-specific single-copy orthologues dataset (actinopterygii_odb10; 24 February 2021). These quality statistics were summarized in snail plots generated using ASSEMBLY-STATS 1.5 (Challis and Blaxter 2025). Synteny analyses were performed between both pygmy perch genomes and those of *M. australasica, M. ambigua, M. peelii*, and *S. chuatsi* using NTSYNT (Coombe et al. 2024). This software uses k-mers and an undirected minimizer graph to identify syntenic blocks. We use default parameters appropriate for genome divergences >10% and visualize synteny using NTSYNT-VIZ (Coombe et al. 2025). Finally, to assess the utility of our assemblies as genome references, low coverage whole genome sequences from five *N. australis* and five *N. obscura* individuals were mapped to both assemblies. For this, we used TRIMMOMATIC 3.9 (Bolger et al. 2014) to trim and filter reads, and BOWTIE2 (Langmead and Salzberg 2012) to perform alignments.

Positive selected miniaturization genes

To assess the role of positive selection over the miniaturization of pygmy perches, we used PRANK 170427 (Löytynoja and Goldman 2005) to probabilistically align codons of the Actinopterygii orthologue genes that were ≥80% complete and single copy in *N. australis, N. obscura, M. australasica, M. ambigua, M. peelii*, and *S. chuatsi* genomes. A phylogenetic tree was inferred with IQ-TREE 1.6.12 (Nguyen et al. 2015) under the standard codon model and 10,000 bootstrap replicates to assess branch support. We tested for positive selection along the *Nannoperca* lineages using PAML 4 (Yang 2007). Likelihood ratio tests were performed comparing branch-site model 2A as alternative hypothesis against both model 0 (neutral) and model 1 (nearly neutral) as null hypotheses, as suggested by Álvarez-Carretero et al. (2023). From the genes showing significance, we performed enrichment analyses of their GO and mammalian phenotype (MP) ontology terms. The GO enrichment analysis was conducted using the ZEBRAFISHMINE 1.0.4.1 (Ruzicka et al. 2019), and the MP enrichment analysis was conducted using the Mammal Phenotype Ontology database (Smith and Eppig 2009), the latter focusing only on terms associated with body size and growth reduction. Although the zebrafish database includes a phenotype ontology, it is currently less comprehensive than the MP ontology. We selected the MP ontology because it provides broader integration, richer phenotype annotations, and more extensive functional mappings. Importantly, the MP ontology also contains specific terms related to body size and growth reduction that could not be directly tested for enrichment on the zebrafish phenotype ontology.

Results

For *N. australis,* a total of 58 Gb Illumina short reads and 176Gb PacBio long reads were generated, representing 255× average depth coverage (Table 1). For *N. obscura*, 27.8 Gb PacBio HiFi long reads were generated, representing 39.7× average depth coverage (Table 1). Estimated genome sizes were 570.2 and 583.2 Mb, with heterozygosity rates of 0.2% and 0.13% for *N. australis* and *N. obscura,* respectively. The final genomes were ∼675.7 and ∼653.9 Mb in length and comprised 914 and 1,476 scaffolds with N50 values of ∼26.99 and ∼26.22 Mb (Table 1 and Fig. 1), of which 92% and 93% were clustered in 24 scaffolds (chromosomes). Overall GC content was ∼41.7%, with ∼0.15 and 0.92% Ns being inserted. Mapping rates were high, with 96.3% of short reads and 92.5% of long reads aligning back to the *N. australis* genome; and 97.3% of short reads and 93.7% of long reads aligning back to the *N. obscura* genome. The final genomes show strong synteny with all publicly available Australian perch species genomes (Fig. 2) despite more than 50 million yr of divergence (Buckley et al. 2018).

Repetitive elements and protein-coding gene

Just over 1 million repeat elements were identified (Table 1), comprising a total of 156.6 to 182.1 Mb (23.1% to 28% of the genome). Most repeats were interspersed (19.7% to 23.2%) (Table 1). We identified 26,390 and 24,654 protein-coding genes in *N. australis* and *N. obscura,* respectively (Table 1; Fig. 3). Among these, 26,032 (98.6%) and 24,344 (98.7%) were functionally annotated to a protein product. Gene length ranges from 128 to 223,085 bp, with means of 9,644 and 8,676 bp for *N. australis* and *N. obscura,* respectively. Average gene density was 42.5 genes/Mb (Fig. 3). One to 168 exons were detected per gene, with average of 10 exons per gene. BUSCO assessment showed that completeness reached 96.3% to 98.7%, with 0.2% to 3.2% duplicated, 1% to 1.5% fragmented,

Table 1 General summary statistics for the two pygmy perch species genome assemblies.

		N. australis	*N. obscura*
Assembly	Total length	675,704,800	653,929,830
	Number of contigs	914	1,476
	Longest contig	32,847,025	32,217,311
	N50	26,992,376	26,217,731
	L50	12	12
	Number of supper scaffolds (chromosomes)	24	24
	N (%)	0.15	0.92
	GC (%)	41.65	41.70
BUSCO v5 (actinopterygii_odb10)	Complete (%)	96.3	98.7
	Complete single copy (%)	93.1	98.5
	Complete duplicated (%)	3.2	0.2
	Fragmented (%)	1.5	1.0
	Missing (%)	2.2	0.3
Repetitive elements	Total	1,029,839	1,091,847
	Total (%)	23.1	28
	SINEs (%)	0.0	0.1
	LINEs (%)	3.1	4.6
	LTR elements (%)	0.9	1.4
	DNA transposons (%)	2.0	5.2
	Simple repeats/low complexity (%)	3.2	2.4
	Unclassified (%)	14.1	12.1
Annotation	Predicted genes	28,732	27,838
	Coding genes	26,390	24,654
	Functional annotated genes	26,032	24,344
	Mean gene length (bp)	9,644	8,679
	Mean exon length (bp)	169	172
	Mean intron length (bp)	884	891
	Mean exons per gene	10	9
	Mean introns per gene	9	8

and 0.3% to 2.2% genes missing (Table 1; Fig. 1). These results demonstrate the high reliability and completeness of our genome assemblies.

Miniaturization genes under positive selection

Around 3,196 orthologous genes were successfully aligned in all the genomes. From these, 855 genes showed evidence of positive selection along the pygmy perch lineage. Although no GO terms were significantly enriched, 795 positively selected genes (PSGs) were associated with 33 MP terms related to body size and/or growth reduction; six of these MP terms were significantly enriched (Table S2). Candidate genes associated with the MP terms were evenly distributed through our assembly, with a clear correlation between chromosome length and the number of candidates per chromosome (Fig. S2). We considered these 795 PSGs to be excellent candidates for exploring the genetic basis and evolution of miniaturization.

Discussion

The combination of hybrid sequencing (long and short reads) and multiple bioinformatic tools produced a high-quality assembly for the endangered pygmy perch *N. australis*. Similarly, the use of HiFi long reads, their dedicated assembly pipeline, and a synteny approach, generate a high-quality assembly for the critically endangered Yarra pygmy perch *N. obscura*. The total size of the new assemblies was slightly larger than the estimates from GENOMESCOPE2 (631.6 Mb vs. 675.7 Mb), and the BUSCO completeness was similar or better than that observed for high-quality chromosome-level percichthyid genome assemblies (Fig. S1). The pygmy perch genomes also are highly contiguous. Matching its low proportion of fragmented orthologous genes, our assemblies have a slightly smaller N50 (Fig. 1 and Table 1) than the Macquarie perch (28.78 Mb) and golden perch (28.07 Mb) genomes, but larger than the reported for the Murray cod genome (23.8 Mb). Although our assemblies can be further improved with additional resources, such as a high-density linkage map or Hi-C sequencing, their metrics, together with the high alignment rate for both long and short reads, support the overall completeness and quality of our assemblies.

Pygmy perch genomes as resources for conservation

Captive breeding, population reintroduction, and enhancement programs are increasingly important for biodiversity conservation

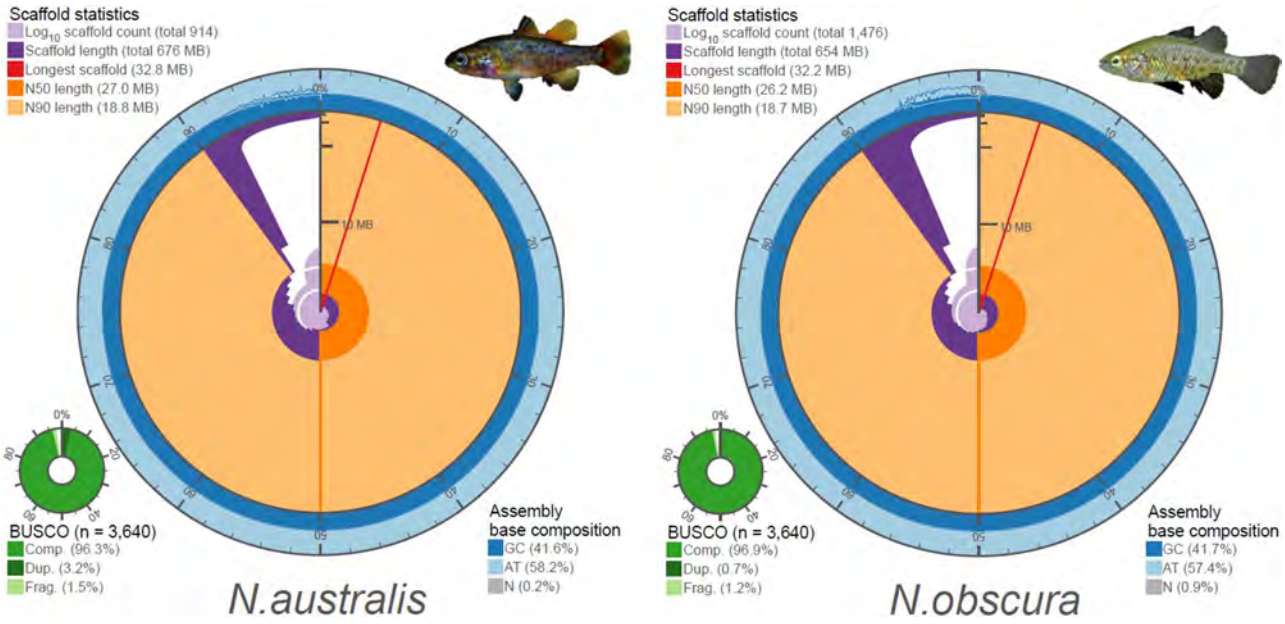

Fig. 1 Snail plots showing the basic statistics of Southern pygmy perch (*Nannoperca australis*) and Yarra pygmy perch (*N. obscura*) genome assemblies. Interactive version of the plots can be found in https://yuma248.github.io/assembly-statsNaus/?assembly=Naustralis and https://yuma248.github.io/assembly-statsNaus/?assembly=Nobscura.

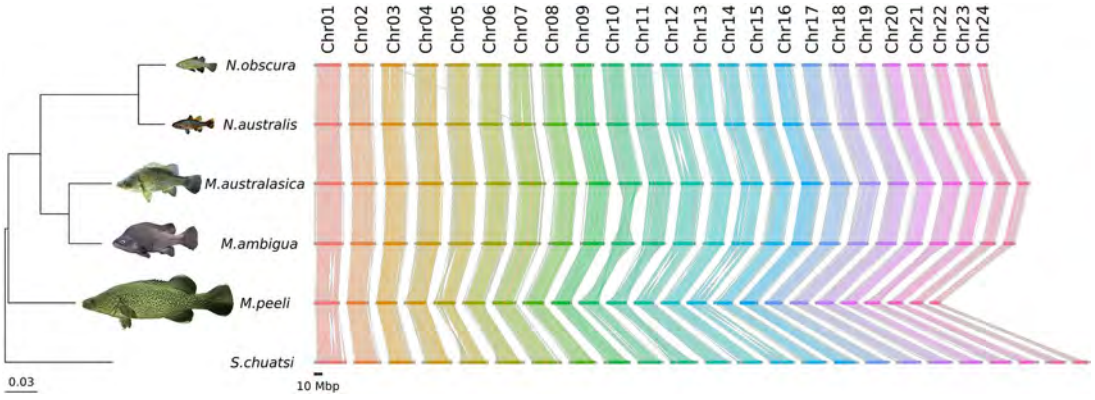

Fig. 2 Multi-genome synteny blocks from five Australian percichthyid species and an outgroup. Vertical order is based on a phylogenetic tree built with 3,588 orthologous genes present (each at least 80% complete) in the six genomes: Yarra pygmy perch (*Nannoperca obscura*), Southern pygmy perch (*N. australis*), Macquarie perch (*Macquaria australasica*), golden perch (*M. ambigua*), Murray cod (*Maccullochella peelii*), and mandarin perch (*Siniperca chuatsi*). This tree was used for the positive selection test in PAML.

as human activities accelerate climate change, habitat loss, and population extinctions (Attard et al. 2016a; Goergen and Gilliam 2018). The success of these programs is significantly improved when integrated with the right genomic data (Ralls et al. 2020; Hoffmann et al. 2021; Kardos et al. 2021), which usually requires a reference genome (Brandies et al. 2019; Formenti et al. 2022). Currently, chromosome-level genome assemblies for percichthyids are available for two *Maquaria* species (Pavlova et al. 2021). Although these genomes could be used to map reads of pygmy perches, their long phylogenetic distance decreases their effectiveness as references (DiBattista et al. 2017). This is evident when comparing the mapping success of low-coverage whole-genome sequences of pygmy perches to the *M. ambigua* genome (~26%) versus mapping to our assemblies (>96%). These results, the completeness, and the contiguousness of our assemblies demonstrate that our assemblies provide a robust resource for scanning the full spectrum of genomic diversity in pygmy perches, including SNPs, structural variation, copy number variance, and runs of homozygosity. In addition, our assemblies improve orthology assumptions and facilitate interpretation based on genome organization (Formenti et al. 2022); increasing the accuracy and power for individual assignment, relatedness, population structure, and demographic inferences. Such information is essential for assessing species adaptive potential to climatic changes (Mérot et al. 2020; Kardos et al. 2021; Wold et al. 2021) and for reconstructing population demographic history (Leitwein et al. 2020; Dussex et al. 2021), both critical for effective conservation planning.

The assemblies and their annotation also provide a foundation for identifying genes and gene interactions involved in local adaptation (Brauer et al. 2016) and inbreeding depression that could occur in small and fragmented pygmy perch populations (Cole et al. 2016; Brauer and Beheregaray 2020; Adams et al. 2022). This knowledge facilitated the identification of loci associated

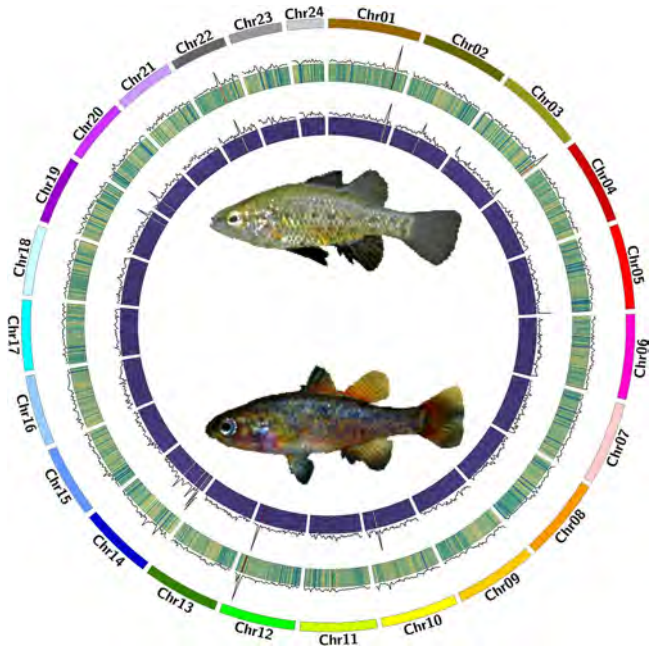

Fig. 3 Circos plot; outer circles represent the relative sizes of the 24 chromosomes of pygmy perches. The middle circle represents gene density per chromosome in Southern pygmy perch, and the inner circle represents gene density per chromosome in Yarra pygmy perch. Both densities are per 1 Mbp and are shown as a histogram (top) and a heatmap (bottom).

with phenotypic and fitness variation and links putatively causal genes to underlying biological processes. Such functional context is likely to improve existing captive breeding programs (Attard et al. 2016a; Attard et al. 2016b) and support the monitoring and maintenance of genomic variation in wild and reintroduced populations (Marshall et al. 2022). For instance, this genomic information combined with common garden experiments could enable genomic prediction of key traits for selecting breeders that would enhance the persistence of reintroduced pygmy perch populations (Laverdière et al. 2022; Sandoval-Castillo et al. 2022). Moreover, our reference genomes allow the detection of chromosomal or genic incompatibilities, disruption of coadapted genetic interactions, and the introduction of maladaptive derivatives from assisted migration/hybridization (Ralls et al. 2020; Hoffmann et al. 2021). For example, the high contiguity of our assemblies will permit the application of haplotype data in the identification of adaptive/maladaptive introgressions and monitoring their effects within reintroduced/restocked populations (Leitwein et al. 2020; Brauer et al. 2023; Gates et al. 2025). This will support genetic rescue approaches, enhancing the benefit of "mixing lineages," but minimizing risks of negative effects (Beheregaray et al. 2026). Overall, these high-quality pygmy perch reference genomes are valuable resources for future and ongoing conservation efforts for these species and other pygmy perches.

Strong synteny and candidate genes for miniaturization

The pygmy perch genomes will also facilitate comparative genomic studies designed to understand percichthyid evolution. Here, we compared our assemblies with the available genomes of Australian percichthyids to explore the genomic architecture underlying miniaturization evolution. Notably, the pygmy perch genomes exhibit strong synteny with all other perch genomes, despite more than 50 million yr of divergence (Buckley et al. 2018). Although chromosome rearrangements are common in fishes, most Perciformes species show limited chromosomal divergence (Galetti Jr et al. 2000). The similarly strong synteny to the Australian perches and the more phylogenetically distant Mandarin perch genomes (Fig. 2) suggest that our results reflect high genome architecture stability within the Percichthyidae family, rather than methodological artifacts. This architectural stability has implications for adaptation and speciation processes in the group, particularly for understanding mechanisms and rates of evolution (Flaxman et al. 2014; Kess et al. 2020).

Miniaturization is a key mechanism in the evolution and diversification of pygmy perches (Johnson 1984; Hanken and Wake 1993). Although miniaturization has evolved in multiple animal lineages (Lautenschlager et al. 2018; Xu et al. 2021), its genomic bases remain poorly understood. In Cypriniformes, large chromosomal rearrangements and genome reduction have been proposed as drivers of miniaturization (Liu et al. 2012), particularly via the loss of key Hox genes (Malmstrøm et al. 2018). In contrast, our pygmy perch assemblies exhibit strong synteny with other percichthyids and no evidence of chromosome number reduction or large-scale chromosomal loss. In gobies, the differential expression of 54 genes has been linked to repeated miniaturization across multiple lineages (Troyer et al. 2025), highlighting the potential for parallel evolutionary routes to small body size via a few key genes. Consistent with this idea, we identified 795 genes under divergent selection in pygmy perches that may underlie adaptation to ecological niches favoring reduced body size. These candidate genes are evenly distributed across the chromosomes (Fig. S2), indicating a genome-wide adaptive signal rather than a localized genomic driver. The use of only six genomes can reduce the power of the positive selection analysis. However, these taxa adequately represent the evolutionary diversity of extant Australian percichthyids and span the divergence range relevant to this study. In PAML branch-site analyses, phylogenetic representativeness, and appropriate evolutionary distances are often more important than simply maximizing taxon number, as the inclusion of highly divergent taxa can reduce alignment reliability and complicate evolutionary inference. Although functional validation is still required, these 795 PSGs represent strong candidates for dissecting the genomic basis of miniaturization in the pygmy perch lineage and its role in percichthyid diversification.

Conclusions

We assembled chromosome-level genomes for the endangered pygmy perches *N. australis* and *N. obscura*, providing an essential resource for conservation programs targeting pygmy perches in Australia. These high-quality assemblies will facilitate the identification of adaptive loci and the monitoring of their diversity in wild and captive breeding populations, thereby improving the success of in-situ and ex-situ conservation efforts. In addition, these assemblies contribute to the growth of comparative genomics and the possibility of exploring new perspectives on genomic architecture and function. More specifically, we identified a set of candidate genes as a foundation for studying the evolution

of miniaturization and its role in the diversification of freshwater fishes. We expect that these resources will also enhance our understanding of human impacts on the evolutionary potential of isolated and small populations.

Acknowledgements

We thank the Flinders University CSE Biodiversity Facility staff for their husbandry of pygmy perches, and Chris Brauer and Diana Vornicu for their assistance in processing the animals and DNA samples.

Author contributions

Jonathan Sandoval-Castillo (Conceptualization, Funding acquisition, Data curation, Formal analysis, Methodology, Software, Visualization, Writing—original draft, review & editing), Luciano B. Beheregaray (Conceptualization, Funding acquisition, Resources, Methodology, Visualization, Writing—original draft, review & editing)

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

This research was supported by an Australian Research Council grant (FT130101068) to L.B.B. We also thank Bioplatforms Australia, specifically the Threatened Species Initiative, which provided funds for some of the sequencing.

Data availability

Assemblies are available at NCBI under BioProject PRJNA1392779 (Southern pygmy perch: JBTGTW000000000; Yarra pygmy perch: JBTGTX000000000). Raw Illumina and PacBio reads for Southern pygmy perch are deposited in the NCBI Sequence Read Archive (SRA) under accessions SRR36611936 and SRR36611937, respectively. Raw PacBio HiFi reads for Yarra pygmy perch are available under SRA accession SRR36611935. The candidate gene list with inferred genomic positions and both genome annotations are accessible via figshare (DOI: 10.6084/m9.figshare.30970432).

References

Adams PE, Crist AB, Young EM, Willis JH, Phillips PC, Fierst JL. Slow recovery from inbreeding depression generated by the complex genetic architecture of segregating deleterious mutations. *Mol Biol Evol*. 2022;39:msab330.

Albert JS, Destouni G, Duke-Sylvester SM, Magurran AE, Oberdorff T, Reis RE, Winemiller KO, Ripple WJ. Scientists' warning to humanity on the freshwater biodiversity crisis. *Ambio*. 2021;50:85–94.

Allen GR, Midgley SH, Allen M. *Field guide to the freshwater fishes*. Australia: Western Australian Museum; 2002.

Álvarez-Carretero S, Kapli P, Yang Z. *Beginner's guide on the use of PAML to detect positive selection*. *Mol Biol Evol* 40. 2023:msad041.

Attard C, Möller L, Sasaki M, Hammer M, Bice C, Brauer C, Carvalho D, Harris J, Beheregaray L. A novel holistic framework for genetic-based captive-breeding and reintroduction programs. *Conserv Biol*. 2016a;30:1060–1069.

Attard CR, Brauer CJ, Van Zoelen JD, Sasaki M, Hammer MP, Morrison L, Harris JO, Möller LM, Beheregaray LB. Multigenerational evaluation of genetic diversity and parentage in captive southern pygmy perch (*Nannoperca australis*). *Conserv Genet*. 2016b;17:1469–1473.

Austin CM, Tan MH, Harrisson KA, Lee YP, Croft LJ, Sunnucks P, Pavlova A, Gan HM. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and nanopore sequencing read. *GigaScience*. 2017;6:gix063.

Barrett J, Bamford H, Jackson P. Management of alien fishes in the Murray-Darling Basin. *Ecol Manag Restor*. 2014;15:51–56.

Beheregaray L, Attard C, Brauer C, Whiterod N, Wedderburn S, Hammer M. Conservation breeding and reintroduction of pygmy perches in the lower Murray-Darling Basin, Australia: two similar species, two contrasting outcomes. In: Soora PS, editor. *IUCN global conservation translocation perspectives: 2021 case studies from around the globe: 2021*. IUCN SSC Conservation Translocation Specialist Group, Environment Agency-Abu Dhabi & Calgary Zoo, Canada; 2021:26–31.

Beheregaray LB, Fitzpatrick S, Whiteley A, Anderson J, Sunnucks P. Genetic Rescue: latest advances and applications. *Evol Appl*. 2026;19:e70225.

Biswas A, Karanth KP. Thinking small: miniaturization might have facilitated adaptation to new niches in tarantulas. *Zool J Linnean Soc*. 2025;203:zlae069.

Blanckenhorn WU. The evolution of body size: what keeps organisms small? *Q Rev Biol*. 2000;75:385–407.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120.

Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. *Genes*. 2019;10:846.

Brauer CJ, Beheregaray LB. Recent and rapid anthropogenic habitat fragmentation increases extinction risk for freshwater biodiversity. *Evol Appl*. 2020;13:2857–2869.

Brauer CJ, Hammer MP, Beheregaray LB. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Mol Ecol*. 2016;25:5093–5113.

Brauer CJ, Unmack PJ, Beheregaray LB. Comparative ecological transcriptomics and the contribution of gene expression to the evolutionary potential of a threatened fish. *Mol Ecol*. 2017;26:6841–6856.

Brauer CJ, Sandoval-Castillo J, Gates K, Hammer MP, Unmack PJ, Bernatchez L, Beheregaray LB. Natural hybridization reduces vulnerability to climate change. *Nat Clim Chang*. 2023;13:282–289.

Buckley SJ, Domingos FM, Attard CR, Brauer CJ, Sandoval-Castillo J, Lodge R, Unmack PJ, Beheregaray LB. Phylogenomic history of enigmatic pygmy perches: implications for biogeography, taxonomy and conservation. *R Soc Open Sci*. 2018;5:172125.

Buckley SJ, Brauer C, Lamin C, Rose P, Vornicu DE, Beheregaray LB. A community-driven captive-breeding and reintroduction program maintains genetic diversity in a threatened freshwater fish. *Conserv Sci Pract*. 2024a;6:e13054.

Buckley SJ, Brauer CJ, Unmack PJ, Hammer MP, Adams M, Beatty SJ, Morgan DL, Beheregaray LB. Long-term climatic stability drives accumulation and maintenance of divergent freshwater fish lineages in a temperate biodiversity hotspot. *Heredity*. 2024b;133:149–159.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:1–9.

Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics*. 2014;48:4.11. 11–14.11. 39.

Challis RJ, Blaxter ML. Snail plots are badges of genome assembly quality. *BiorXiv*. 2025;2025:2020.689594.

Cole TL, Hammer MP, Unmack PJ, Teske PR, Brauer CJ, Adams M, Beheregaray LB. Range-wide fragmentation in a threatened fish associated with post-European settlement modification in the Murray–Darling Basin, Australia. *Conserv Genet*. 2016;17:1377–1391.

Coleman R, Gauffre B, Pavlova A, Beheregaray L, Kearns J, Lyon J, Sasaki M, Leblois R, Sgro C, Sunnucks P. Artificial barriers prevent genetic recovery of small isolated populations of a low-mobility freshwater fish. *Heredity*. 2018;120: 515–532.

Cooke GM, Chao NL, Beheregaray LB. Natural selection in the water: freshwater invasion and adaptation by water colour in the Amazonian pufferfish. *J Evol Biol*. 2012;25:1305–1320.

Coombe L, Kazemi P, Wong J, Birol I, Warren RL. Multi-genome synteny detection using minimizer graph mappings. *BioRxiv: 20242002*. 2024;2007.579356

Coombe L, Warren RL, Birol I. ntSynt-viz: visualizing synteny patterns across multiple genomes. *BioBxiv: 20252001*. 2025; 2015.633221.

Darwall WR, Holland RA, Smith KG, Allen D, Brooks EG, Katarya V, Pollock CM, Shi Y, Clausnitzer V, Cumberlidge N. Implications of bias in conservation research and investment for freshwater species. *Conserv Lett*. 2011;4:474–482.

DiBattista JD, Saenz-Agudelo P, Piatek MJ, Wang X, Aranda M, Berumen ML. Using a butterflyfish genome as a general tool for RAD-Seq studies in specialized reef fish. *Mol Ecol Resour*. 2017;17:1330–1341.

Donaldson MR, Burnett NJ, Braun DC, Suski CD, Hinch SG, Cooke SJ, Kerr JT. *Taxonomic bias and international biodiversity conservation research*. Ottawa, ON: Canadian Science Publishing; 2016.

Dussex N, Van Der Valk T, Morales HE, Wheat CW, Díez-del-Molino D, Von Seth J, Foster Y, Kutschera VE, Guschanski K, Rhie A. Population genomics of the critically endangered kākāpō. *Cell Genom*. 2021;1:100002.

Esin EV, Markevich GN, Shkil FN. Rapid miniaturization of *Salvelinus* fish as an adaptation to the volcanic impact. *Hydrobiologia*. 2020;847:2947–2962.

Flaxman SM, Wacholder AC, Feder JL, Nosil P. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol Ecol*. 2014;23:4074–4088.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci*. 2020;117:9451–9457.

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J. The era of reference genomes in conservation genomics. *Trends Ecol Evol*. 2022;37:197–202.

Galetti P Jr, Aguilar C, Molina W. An overview of marine fish cytogenetics. *Hydrobiologia*. 2000;420:55–62.

Gates K, Sandoval-Castillo J, Beaman JE, Da Silva KB, Saltré F, Belov K, Hogg CJ, Bradshaw CJ, Beheregaray LB. Conservation Arks: genomic erosion and inbreeding in an abundant island population of koalas. *Mol Ecol*. 2025;34:e70097.

Goergen EA, Gilliam DS. Outplanting technique, host genotype, and site affect the initial success of outplanted *Acropora cervicornis*. *PeerJ*. 2018;6:e4433.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–1075.

Hammer MP, Bice CM, Hall A, Frears A, Watt A, Whiterod NS, Beheregaray LB, Harris JO, Zampatti BP. Freshwater fish conservation in the face of critical water shortages in the southern Murray–Darling Basin, Australia. *Mar Freshw Res*. 2013;64:807–821.

Hanken J, Wake DB. Miniaturization of body size: organismal consequences and evolutionary significance. *Annu Rev Ecol Syst*. 1993;24:501–519.

He S, Li L, Lv L-Y, Cai W-J, Dou Y-Q, Li J, Tang S-L, Chen X, Zhang Z, Xu J. Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. *Commun Biol*. 2020;3:1–13.

Hoffmann AA, Miller AD, Weeks AR. Genetic mixing for population management: from genetic rescue to provenancing. *Evol Appl*. 2021;14:634–652.

Johnson G. Percoidei: development and relationships. In: Moser H et al., eds. *Ontogeny and systematics of fishes. Special publication No. 1, American Society of Ichthyologists and Herpetologists Kansas*. USA: Allen Press; 1984:464–498.

Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC. The crucial role of genome-wide genetic variation in conservation. *Proc Natl Acad Sci*. 2021;118:e2104642118.

Kess T, Bentzen P, Lehnert SJ, Sylvester EV, Lien S, Kent MP, Sinclair-Waters M, Morris C, Wringe B, Fairweather R. Modular chromosome rearrangements reveal parallel and nonparallel adaptation in a marine fish. *Ecol Evol*. 2020;10:638–653.

Kolmogorov M, Armstrong J, Raney BJ, Streeter I, Dunn M, Yang F, Odom D, Flicek P, Keane TM, Thybert D. Chromosome assembly of large and complex genomes using multiple references. *Genome Res*. 2018;28:1720–1732.

Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004;5:1–9.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359.

Lautenschlager S, Gill PG, Luo Z-X, Fagan MJ, Rayfield EJ. The role of miniaturization in the evolution of the mammalian jaw and middle ear. *Nature*. 2018;561:533–537

Laverdière J, Lenz P, Nadeau S, Depardieu C, Isabel N, Perron M. Breeding for adaptation to climate change: genomic selection for drought response in a white spruce multi-site polycross test. *Evol Appl*. 2022;15:383–402.

Leitwein M, Duranton M, Rougemont Q, Gagnaire P-A, Bernatchez L. Using haplotype information for conservation genomics. *Trends Ecol Evol*. 2020;35:245–258.

Lennox RJ, Crook DA, Moyle PB, Struthers DP, Cooke SJ. Toward a better understanding of freshwater fish responses to

an increasingly drought-stricken world. *Rev Fish Biol Fish*. 2019;29:71–92.

Lintermans M, Geyle HM, Beatty S, Brown C, Ebner BC, Freeman R, Hammer MP, Humphreys WF, Kennard MJ, Kern P. Big trouble for little fish: identifying Australian freshwater fishes in imminent risk of extinction. *Pac Conserv Biol*. 2020;26:365–377.

Liu S, Hui TH, Tan SL, Hong Y. Chromosome evolution and genome miniaturization in minifish. *PLoS ONE*. 2012;7:e37305.

Löytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci*. 2005;102:10557–10562.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 2012;1:2047-2217X-2041-2018.

MacColl AD, Nagar AE, de Roij J. The evolutionary ecology of dwarfism in three-spined sticklebacks. *J Anim Ecol*. 2013;82:642–652.

Malmstrøm M, Britz R, Matschiner M, Tørresen OK, Hadiaty RK, Yaakob N, Tan HH, Jakobsen KS, Salzburger W, Rüber L. The most developmentally truncated fishes show extensive Hox gene loss and miniaturized genomes. *Genome Biol Evol*. 2018;10:1088–1103.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38:4647–4654.

Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27:764–770.

Marshall IR, Brauer CJ, Wedderburn SD, Whiterod NS, Hammer MP, Barnes TC, Attard CR, Möller LM, Beheregaray LB. Longitudinal monitoring of neutral and adaptive genomic diversity in a reintroduction. *Conserv Biol*. 2022;36:e13889.

Mérot C, Oomen RA, Tigano A, Wellenreuther M. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol*. 2020;35:561–572.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–274.

Pavlova A, Harrisson K, Turakulov R, Lee YP, Ingram B, Gilligan D, Sunnucks P, Gan HM. Labile sex chromosomes in the Australian freshwater fish family Percichthyidae. *Mol Ecol Resour*. 2022;22:1639–1655.

Ralls K, Sunnucks P, Lacy RC, Frankham R. Genetic rescue: a critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biol Conserv*. 2020;251:108784.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432.

Ruzicka L, Howe DG, Ramachandran S, Toro S, Van Slyke CE, Bradford YM, Eagle A, Fashena D, Frazer K, Kalita P. The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Res*. 2019;47:D867–D873

Saddlier S, Koehn JD, Hammer MP. Let's not forget the small fishes—conservation of two threatened species of pygmy perch in south-eastern Australia. *Mar Freshw Res*. 2013;64:874–886.

Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics*. 2014;30:3506–3514.

Sandoval-Castillo J, Beheregaray L, Wellenreuther M. Genomic prediction of growth in a commercially, recreationally, and culturally important marine resource, the Australasian snapper (*Chrysophrys auratus*). *G3 Genes Genomes Genet*. 2022;12:jkac015.

Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005;6:1–11.

Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med*. 2009;1:390–399.

Sovic I, Kronenberg K, Dunn C, Barnett D, Kingan S, Drake J. *IPA HiFi genome assembler*. PacificBiosciences; 2022: https://github.com/PacificBiosciences/pbipa.

Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 2008;24:637–644.

Sunnucks P, Hales DF. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Biol Evol*. 1996;13:510–524.

Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009;4:10.

Thiem JD, Wooden IJ, Baumgartner LJ, Butler GL, Taylor MD, Watts RJ. Hypoxic conditions interrupt flood-response movements of three lowland river fish species: implications for flow restoration in modified landscapes. *Ecohydrology*. 2020;13:e2197.

Todd CR, Koehn JD, Pearce L, Dodd L, Humphries P, Morrongiello JR. Forgotten fishes: what is the future for small threatened freshwater fish? Population risk assessment for southern pygmy perch, *Nannoperca australis*. *Aquat Conserv Mar Freshwat Ecosyst*. 2017;27:1290–1300.

Tonkin Z, Kitchingman A, Ayres R, Lyon J, Rutherfurd I, Stout J, Wilson P. Assessing the distribution and changes of instream woody habitat in south-eastern Australian rivers. *River Res Appl*. 2016;32:1576–1586.

Troyer EM, White WT, Betancur-R R, Arcila D. Parallel shifts in differential gene expression reveal convergent miniaturization in fishes. *Proc Natl Acad Sci*. 2025;122:e2512299122.

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, Glidden S, Bunn SE, Sullivan CA, Liermann CR. Global threats to human water security and river biodiversity. *Nature*. 2010;467:555–561.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9:e112963.

Wold J, Koepfli KP, Galla SJ, Eccles D, Hogg CJ, Le Lec MF, Guhlin J, Santure AW, Steeves TE. Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern. *Mol Ecol*. 2021;30:5949–5965.

Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, Fan G, Liu X, Xu X, Deng L. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience*. 2020;9:giaa094.

Xu H, Ye X, Yang Y, Yang Y, Sun YH, Mei Y, Xiong S, He K, Xu L, Fang Q. Comparative genomics sheds light on the convergent evolution of miniaturized wasps. *Mol Biol Evol*. 2021;38:5539–5554.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–1591.

Ye C, Hill CM, Wu S, Ruan J, Ma ZS. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci Rep*. 2016;6:1–9.

Zhu B-H, Xiao J, Xue W, Xu G-C, Sun M-Y, Li J-T. P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. *BMC Genomics*. 2018;19:1–13.

Journal of Heredity, 2026, 00, 1–10
https://doi.org/10.1093/jhered/esag054
Genome Resources