Potential Contribution of Ancient Introgression to the Evolution of a Derived Reproductive Strategy in Ricefishes

Jana M. Flury [1,2,*], Karen Meusemann[1], Sebastian Martin[1], Leon Hilgers[1], Tobias Spanke[1], Astrid Böhne [1], Fabian Herder[1], Daniel F. Mokodongan[3], Janine Altmüller[4], Daisy Wowor[3], Bernhard Misof[1], Arne W. Nolte[5], and Julia Schwarzer [1,*]

[1]Leibniz-Institute for the Analysis of Biodiversity Change (LIB), Museum Koenig Bonn, Bonn, Germany

[2]Department of Environmental Sciences, University of Basel, Basel, Switzerland

[3]Museum Zoologicum Bogoriense, Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, West Java, Indonesia

[4]Cologne Center for Genomics (CCG), Cologne University, Cologne, Germany

[5]Department of Ecological Genomics, Carl von Ossietzky Universität, Oldenburg, Germany

*Corresponding authors: E-mail: j.flury@leibniz-lib.de, j.schwarzer@leibniz-lib.de.

Accepted: 13 July 2023

Abstract

Transitions from no parental care to extensive care are costly and involve major changes in life history, behavior, and morphology. Nevertheless, in Sulawesi ricefishes, pelvic brooding evolved from transfer brooding in two distantly related lineages within the genera *Adrianichthys* and *Oryzias*, respectively. Females of pelvic brooding species carry their eggs attached to their belly until the fry hatches. Despite their phylogenetic distance, both pelvic brooding lineages share a set of external morphological traits. A recent study found no direct gene flow between pelvic brooding lineages, suggesting independent evolution of the derived reproductive strategy. Convergent evolution can, however, also rely on repeated sorting of preexisting variation of an admixed ancestral population, especially when subjected to similar external selection pressures. We thus used a multispecies coalescent model and D-statistics to identify gene-tree–species-tree incongruencies, to evaluate the evolution of pelvic brooding with respect to interspecific gene flow not only between pelvic brooding lineages but also between pelvic brooding lineages and other Sulawesi ricefish lineages. We found a general network-like evolution in Sulawesi ricefishes, and as previously reported, we detected no gene flow between the pelvic brooding lineages. Instead, we found hybridization between the ancestor of pelvic brooding *Oryzias* and the common ancestor of the *Oryzias* species from the Lake Poso area. We further detected signs of introgression within the confidence interval of a quantitative trait locus associated with pelvic brooding in *O. eversi*. Our results hint toward a contribution of ancient standing genetic variation to the evolution of pelvic brooding in *Oryzias*.

Key words: ancient gene flow, convergent evolution, pelvic brooding, reproductive strategy, Sulawesi ricefishes.

Significance

The evolution of pelvic brooding in *Oryziaseversi* (Beloniformes: Adrianichthyidae) was recently described to be independent from another pelvic brooding ricefish lineage from Sulawesi (*Adrianichthys*). We confirmed these results and detected no gene flow between the two distantly related pelvic brooding lineages. Instead, we found ancient gene flow from another *Orzyias* lineage into the pelvic brooding *Oryzias* lineage. One of the previously described genetic regions associated with pelvic brooding overlaps with a region of high introgression signal. Therefore, we assume that not only de novo mutations contributed to the evolution of pelvic brooding in *Orzyias*, but that introgressed ancient genetic variation was potentially also recruited for the evolution of this derived brooding strategy.

Introduction

The repeatability of evolution exemplified by convergent traits intrigued biologists since the early days (Darwin 1859). Similar selective regimes were predicted to be a main driver of convergent evolution resulting in similar adaptive phenotypes (Schluter and Nagel 1995; Elmer and Meyer 2011). With the advance of next-generation sequencing technologies, the genetic basis of repeatedly evolved traits has been discovered in a variety of organisms (Stern 2013). Genetic mechanisms for the evolution of similar phenotypic traits in different species can be independent de novo mutations (e.g., Zhang et al. 2021), introgression (e.g., Grant et al. 2004; The Heliconius Genome Consortium et al. 2012; Jones et al. 2020), or repeated sorting of shared standing genetic variation (e.g., Veale and Russello 2017; Van Belleghem et al. 2018; Greenway et al. 2020). However, de novo mutations fix rather slowly (Hermisson and Pennings 2005; Barrett and Schluter 2008) and are rarely beneficial (Ohta 1992). Introgressed variation from one into another population or even species may provide fast access to additional potentially adaptive genetic variation (Waters and McCulloch 2021), if species boundaries are weak (Seehausen 2004; Mallet 2005; Whitney et al. 2006; Arnold 2007; Baack and Rieseberg 2007; Castric et al. 2008; Rieseberg 2009; The Heliconius Genome Consortium et al. 2012). Accordingly, high levels of hybridization promoted rapid radiations of *Heliconius* butterflies (Edelman et al. 2019), similar to Darwin's finches (Lamichhaney et al. 2015) and African cichlids (Meier et al. 2017). Convergent evolution by repeated sorting (Waters and McCulloch 2021) involves alleles identical by descent. A typical example is the stickleback, in which standing genetic variation was repeatedly sorted, resulting in convergent phenotypes in different populations on a global scale (Foster and Baker 2004; Colosimo et al. 2005; Terekhanova et al. 2014). Although the underlying polymorphism originated millions of years in the past, the convergent ecomorphs only evolved after the last ice age (Nelson and Cresko 2018). In some examples, the polymorphisms used in repeated sorting originated from ancient hybridization events, therefore combining both, introgression and repeated sorting as mechanisms leading to convergent evolution (Meier et al. 2017; Veale and Russello 2017; Van Belleghem et al. 2018). This reassembling of old variants partly deriving from ancient introgression events leads to a large potential of new combinations of genetic variation used in speciation and adaptive radiation (reviewed in (Marques et al. 2019).

In Sulawesi ricefishes (Beloniformes; Adrianichthyidae), a group of freshwater fishes endemic to the island of Sulawesi, Indonesia, an exceptional reproductive strategy ("pelvic brooding") evolved in two distantly related lineages (~16 Myr divergence time, fig. 1) (Mokodongan and Yamahira 2015; Hilgers and Schwarzer 2019): in the genus *Adrianichthys* and in three closely related *Oryzias* species: *O. eversi, O. sarasinorum*, and *O. kalimpaaensis* (Kottelat 1990; Parenti 2008; Mokodongan and Yamahira 2015; Gani et al. 2022). It was suggested that this brooding strategy evolved twice convergently (Montenegro et al. 2022). Females of pelvic brooding ricefish species carry a cluster of eggs attached to their gonoduct for up to 3 weeks, until the fry hatches (Kottelat 1990). They have elongated pelvic fins and shorter ribs, forming a concavity where the egg cluster is situated (Kottelat 1990; Spanke et al. 2021). In contrast, in the more common and ancestral brooding strategy transfer brooding (Parenti 2008), females deposit the eggs after several hours (Yamamoto 1975; Wootton and Smith 2014). Changing reproductive strategies entail not only severe changes in life history but also morphological adaptations which in this case are related to prolonged carrying of eggs (Parenti 2008; Herder et al. 2012; Spanke et al. 2021). Why and how such major transitions of reproductive strategies evolve in some groups but not in others remains largely unclear (e.g., Bainbridge 2014).

Gene flow is abundant in Central Sulawesi ricefishes and has been detected between pelvic brooding *Oryzias* (PBO) species (*O. sarasinorum* and *O. eversi*) (Horoiwa et al. 2021) and within two Lake systems (Poso [Sutra et al. 2019] and Malili [Mandagi et al. 2021]). The Malili Lakes species, the Lake Poso species (including *O. soerotoi* from Lake Tiu, which is geographically very close to Lake Poso, Sutra et al. 2019), and the PBO together with *O. dopingdopingensis* are closely related (figs. 1 and 2*A*). Despite the prevalent gene flow within the Central Sulawesi species (CSS, fig. 1), no direct gene flow was found between the two pelvic brooding lineages: PBO and pelvic brooding *Adrianichthys* (PBA) (Montenegro et al. 2022). In the present study, we contrast a presumably independent origin of pelvic brooding in the two ricefish lineages with the introgression of genetic variants via gene flow in the PBO, by using a phylogenomic approach. We reconstructed 1,907 gene-trees based on single-copy protein-coding orthologous genes of ten Sulawesi ricefishes, five mainland ricefishes, and four outgroup species (two poeciliid and two killifish species each). We used these gene-trees as basis for a multispecies coalescence analysis to investigate gene-tree–species-tree incongruencies and to reconstruct a phylogenetic network. We further evaluated introgression on a genome-wide level on published genomes of 17 Sulawesi ricefishes (including four pelvic brooding species) and two non-Sulawesi ricefish species.

Results

Gene-tree–Species-tree Incongruencies and Convergent Evolution of Pelvic Brooding

The ancestral state reconstruction based on our maximum likelihood (ML) tree supported that pelvic brooding

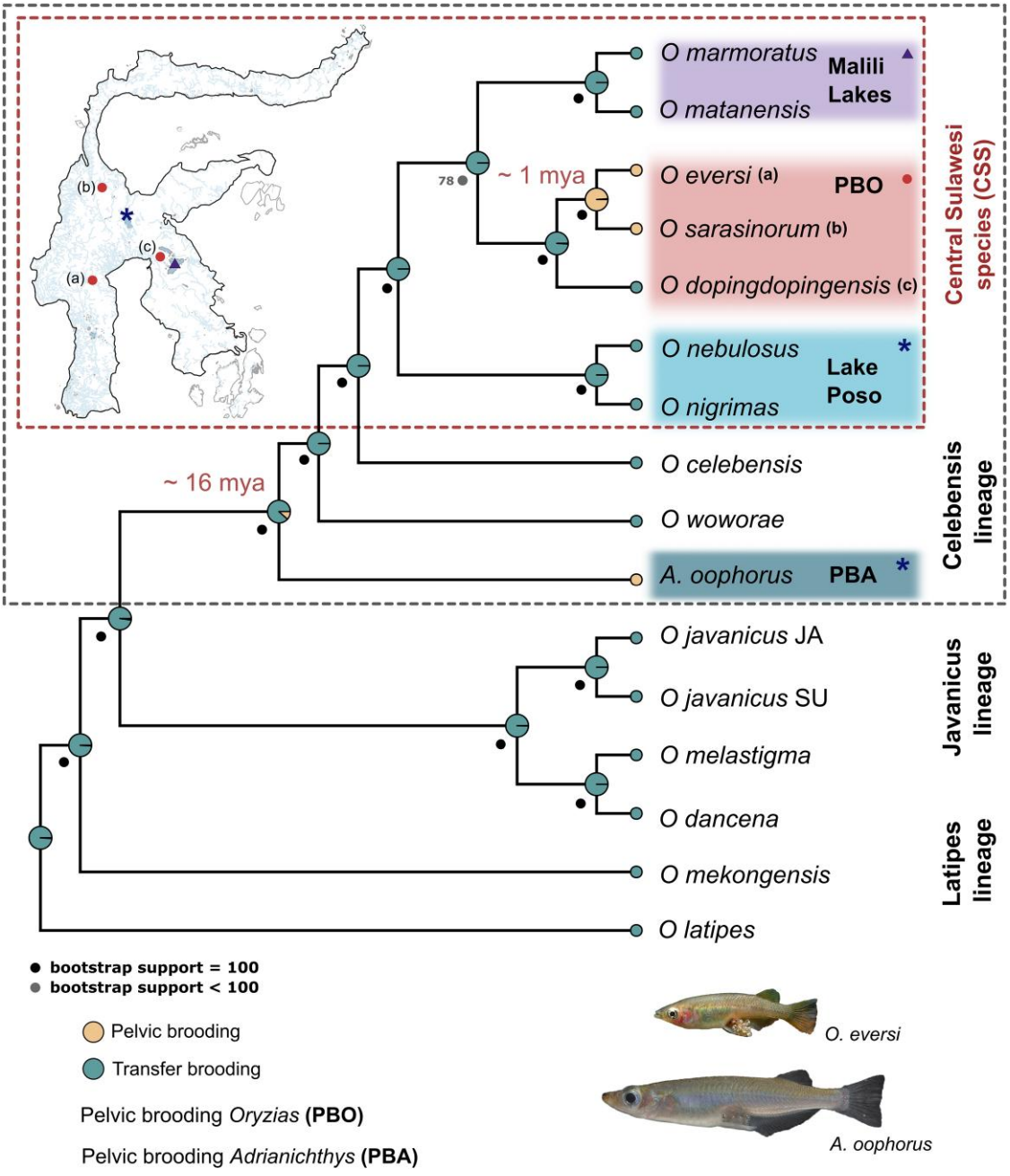

Fig. 1—Ancestral state reconstruction supporting convergent evolution of pelvic brooding in Sulawesi *Oryzias* (PBO) and *Adrianichthys* (PBA). Shown is the maximum likelihood tree generated in this study based on 1,907 orthologous single-copy genes. Suggested node ages for the pelvic brooding lineages are taken from Mokodongan and Yamahira (2015). Black dots below nodes indicate maximal bootstrap support (BS = 100). One node (gray dot), supporting the sister group relationship of the PBO and *O. dopingdopingensis* lineage with the Malili Lakes species, has a bootstrap support of BS = 78. The distribution centers of the Central Sulawesi ricefishes included in the tree are indicated on the map in the upper left corner; symbols on the map correspond to those indicated in the phylogeny. In the lower right corner, brooding females of a representative of both pelvic brooding ricefish lineages PBO (*O. eversi*) and PBA (*A. oophorus*) are shown. Map of Sulawesi including water shed information was obtained from open-source.

evolved convergently in *Oryzias* and *Adrianichthys* (fig. 1). The ML tree (fig. 1), based on the concatenated supermatrix of 1,907 orthologous genes (DS1 in the following, supplementary table S1*a*, Supplementary Material online), revealed a highly congruent topology with the one published by Mokodongan and Yamahira (2015). The main difference was that in our ML tree, the split within the CSS showed only moderate bootstrap support (bootstrap

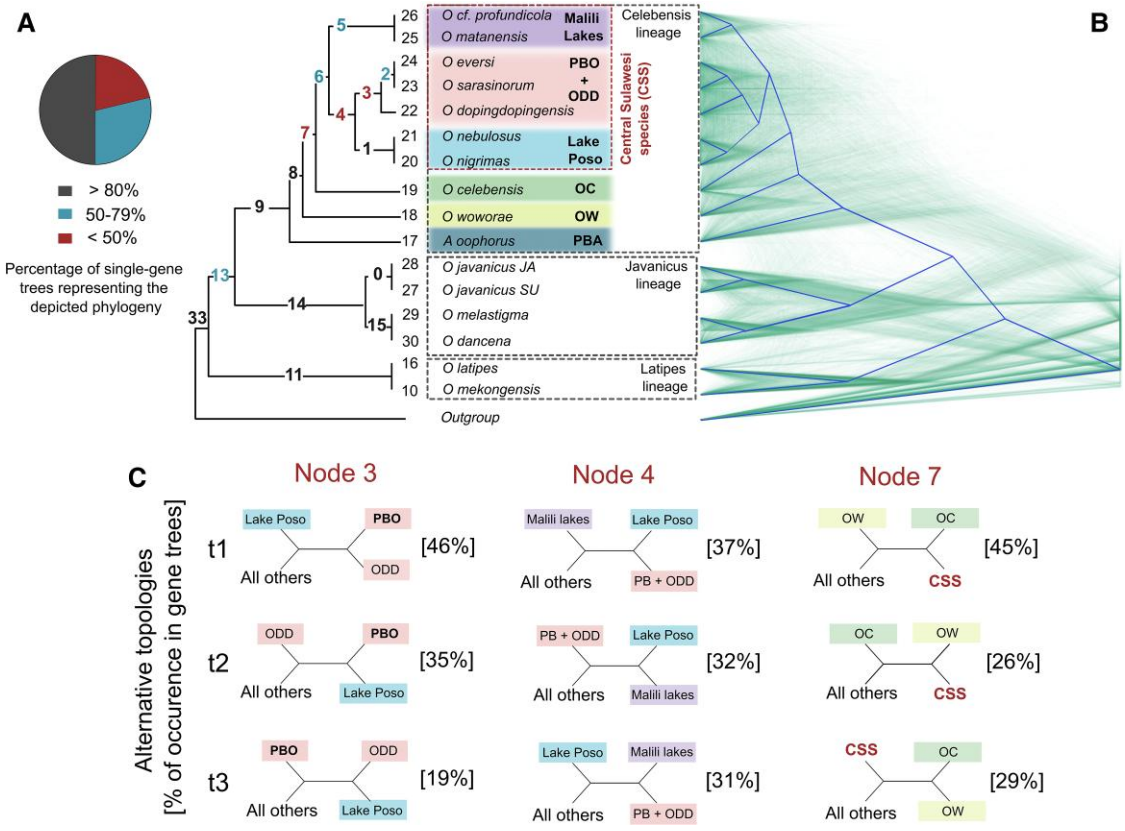

Fig. 2—Results of multispecies-coalescence approach to investigate gene-tree–species-tree incongruencies. (*A*) Astral species tree with visualizations of gene tree discordances based on Discovista. Branch quartet frequencies are coded at each node with different size and color of the respective node number: numbers in black depict quartets in which >80% of all gene trees follow the shown topology. Numbers in turquoise (branch 2,5,6 and 13) indicate that quartets following the shown topology are present in 50–79% of all gene trees, and red numbers (branch 3, 4 and 7) indicate splits where the shown topology of the quartet was found in <50% of all gene trees. The pie chart in the upper left corner shows the frequency of each well (black, >80%), moderately (turquoise, 50–79%), and poorly (red, <50%) supported quartet in the Astral species tree. (*B*) Densitree based on 1,907 single gene trees generated by IQtree. The thin blue line shows the consensus phylogeny (species tree). Blurry areas indicate gene-tree–species-tree incongruencies. (*C*) Quartets depicted in the Astral species tree (t1) and alternative quartets (t2 and t3) of splits with poorly supported quartet scores (<50%). All quartet scores and alternative quartets for other splits are shown in detail in supplementary figure S2, Supplementary Material online.

support (BS) = 78, fig. 1). Further, in the species tree inferred by the multispecies coalescent model in this study, the topology within the CSS was inconsistent, as indicated by low quartet scores for splits 3 and 4 (fig. 2*A* and *C*). Matching this inconsistency, we found that these clades had many gene-tree–species-tree incongruencies (fig. 2*B*). Even though the posterior support values were "1" in each branching event in the species tree determined by Astral (supplementary fig. S1, Supplementary Material online), quartet scores imply a high degree of uncertainty in splits 3, 4, and 7 (fig. 2*A*, marked in red) with almost equal quartet frequencies for alternative topologies (fig. 2*C* and supplementary fig. S2, Supplementary Material online). The split between the PBA and the Sulawesi *Oryzias* was strongly supported, both in the ML tree (fig. 1) as well as in the species tree inferred by the multispecies coalescent model (fig. 2*A*).

Gene Flow within the Central Sulawesi *Oryzias*

Using the Dtrios method implemented in Dsuite (Malinsky et al. 2021) on a dataset of 27 *Oryzias/Adrianichthys* whole-genome sequences comprising ~38 million genomic single nucleotide polymorphisms (SNPs) (DS2 in the following, supplementary table S1*b*, Supplementary Material online), we neither found hints for gene flow between *A. oophorus* and *O. eversi* nor between *A. oophorus* and *O. sarasinorum*. However, we found a signal of hybridization of species from Lakes Poso and Tiu (*O. nebulosus*, *O. nigrimas*, *O. orthognathus*, and *O. soerotoi*) with *O. sarasinorum* and *O. eversi*. Further, we found several signals of gene flow within the Malili Lakes *Oryzias*, with an especially high signal between *O. marmoratus* from Mahalona and *O. matanensis* (all D-values and z-scores of the Drios analysis can be found in supplementary table S12, Supplementary Material online).

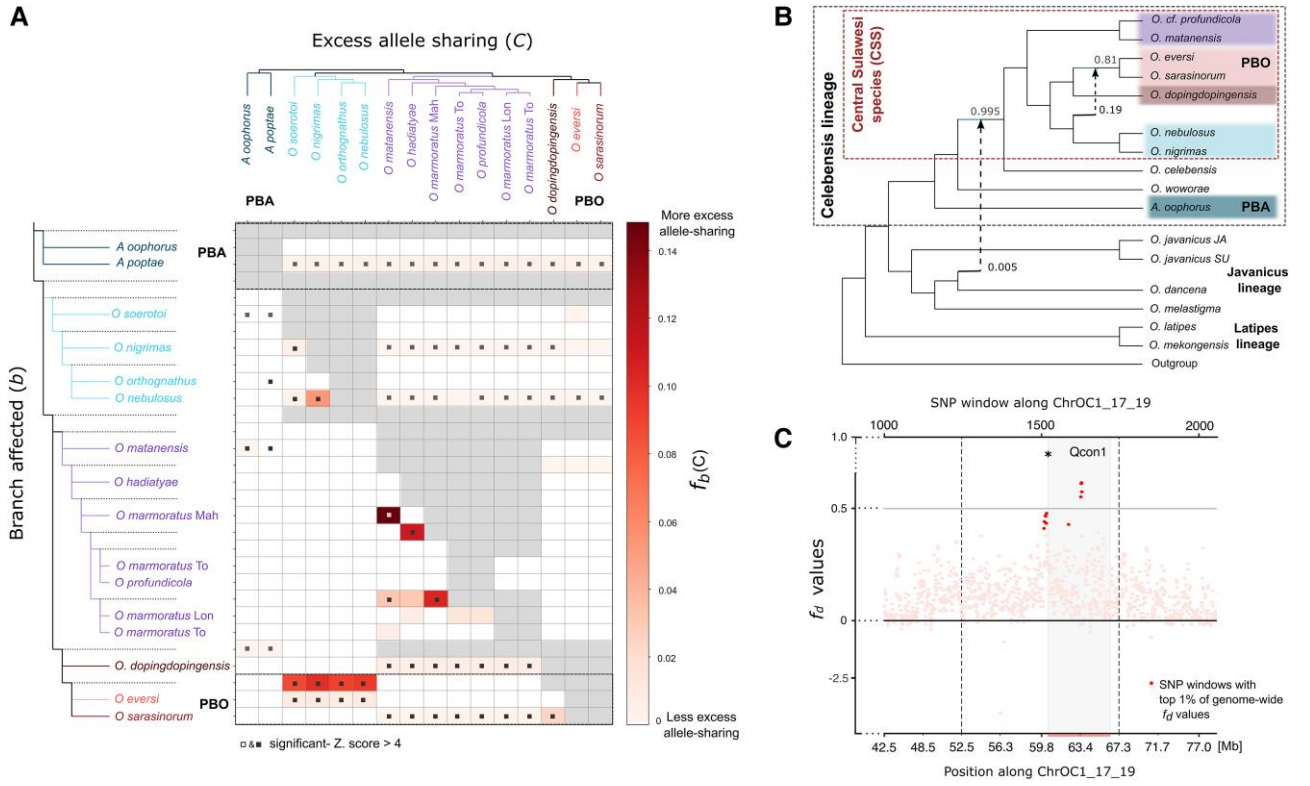

Fig. 3—Signatures of introgression between Central Sulawesi ricefishes. (*A*) F-branch [*fb*(*C*)] statistics across the Central Sulawesi ricefishes. The excess of allele-sharing is shown between tips of the tree (horizontally arranged along the *x* axis) and each other tips (solid lines) and nodes (dashed lines) in the phylogenetic tree (vertically arranged on the *y* axis). The tree that was used as a basis for the branch statistic was a Neighbor-Joining tree generated based on the SNP dataset (DS2). The redness of each cell in the matrix indicates the degree of excess allele sharing between each tree tip (*C*) and each branch (*b*). Excess of allele-sharing was significant when the z-score was >4 (equivalent to Bonferroni multiple-testing corrected *P*-value of 0.001), which is indicated by small black squares. Gray cells in the matrix correspond to tests that are not consistent with the phylogeny. Colors in the tree correspond to phylogenetic lineages as in figure 1. (*B*) Phylogenetic network revealed by SNaQ including representative of all ricefish lineages, indicates gene flow between ancestor of Lake Poso *Oryzias* and ancestor of PBO. Another hybrid edge is indicated between *O. dancena* and *O. celebensis* and Sulawesi lineages, though the proportion consisting of introgressed *O. dancena* genome was low, with 0.5%. The network with two hybrid edges was chosen due to its low network score, which did not improve with more hybrid edges. Networks with three and four hybrid edges are shown in supplementary figure S4, Supplementary Material online. (*C*) Region on the *O. celebensis* reference genome OCchr1_17_19 showing high introgression signal (top 1% of f_d in color) and overlapping with one (QCon1) of the three QTL windows correlating with the extent of the ventral concavity in *O. eversi* (Montenegro et al. 2022). The asterisk marks the location of the QTL within the QTL confidence interval (Montenegro et al. 2022).

Excess of allele-sharing was computed using the f-branch statistic fb(C) (Malinsky et al. 2018, 2021), which was calculated from the f4 admixture ratios of all Sulawesi ricefish species according to our phylogeny (fig. 1, Celebensis lineage). The highest amount of excess of allele-sharing (indicated by a high, and significant, fb(C) value) was present within the Malili lakes *Oryzias* species (10.4–14.8%), between the ancestor of the two PBO species and the Lake Poso *Oryzias* species (8.7–9.8%, fig. 3*A* and supplementary table S13, Supplementary Material online) and between *O. nigrimas* and *O. nebulosus* from Lake Poso (5.3%). In all other species or clades showing a significant excess of allele-sharing with other species or clades across our dataset, the proportion of shared alleles was very low (all <0.6%, fig. 3*A* and supplementary table S13, Supplementary Material online). Such a low (but significant) excess of allele-sharing was, for example, evident between the two *Adrianichthys* species and *O. soerotoi*, *O. orthognathus*, *O. matanensis*, and between the *Adrianichtys* species and the stem lineage of *O. dopingdopingensis*, *O. eversi*, and *O. sarasinorum* (between 0.03% and 0.09%) and in *A. poptae*, which showed significant, but low excess of allele-sharing (0.36–0.42%) with all tested *Oryzias* species (fig. 3*A* and supplementary table S13, Supplementary Material online).

The results of the D- and f-branch statistics were congruent with phylogenetic networks generated with SNaQ: all indicated introgression from the Lake Poso *Oryzias* into the pelvic brooders *O. eversi* and *O. sarasinorum* (19% of shared alleles, fig. 3*B*). We also observed a hybrid edge

between *O. dancena* and the ancestor of *O. celebensis* and the younger *Oryzias* lineages. However, the proportion of introgression here was rather low at 0.46% (fig. 3*B*). Further, with three hypothesized hybrid edges, we observed a hybrid edge between the ancestor of *O. matanensis* and *O.* cf. *profundicola* (supplementary fig. S4, Supplementary Material online). This was in line with the ambiguous node 4 in the quartet tree of Astral (fig. 2*A*) and the moderately supported node in the ML tree (fig. 1).

We used a sliding-window approach performed with DInvestigate from Dsuite to search for genomic regions with high introgression (Malinsky et al. 2021). We found the strongest signals of introgression (top 1% of f_d values) on chromosomes 9, 12_20_13, and 21 (supplementary fig. S3, Supplementary Material online) of the *O. celebensis* reference genome (OCchr) (supplementary fig. S5, Supplementary Material online). In one region on chromosome OCchr 1_17_19, we found five of the top 1% genome wide f_d values for the tested trio (P1: *O. dopingdopingensis*, P2: PBO, P3: Lake Poso *Oryzias*) within the confidence interval of a quantitative trait locus (QTL) for the ventral concavity (labelled QCon1,fig. 3C), described by Montenegro et al., (2022). This genomic region is one of three regions associated with the ventral concavity, where the egg cluster is situated in pelvic brooding females. To test if this number of high f_d values within an interval of the size of the QCon1 QTL can be observed by chance, we sampled intervals with the same number of windows of 50 informative SNPs as the QCon1 QTL (which are $N = 169$ windows, based on our empirical dataset) for 10,000 times on each chromosome. Here, a higher or equally high number of f_d values was on average only recovered in 4.72% of all intervals of the same size than QCon1. For other QTLs identified in Montenegro et al., 2022, the same or higher number of top 1% f_d values were found in between >17% to 100% of the 10,000 permutations in intervals the size of the respective QTL (supplementary table S15, Supplementary Material online).

Discussion

Analyzing 1,907 single-copy protein-coding genes and ~ 38 mio genomic SNPs did not reveal any sign of direct gene flow between the two pelvic brooding ricefish lineages (fig. 3*A* and *B*), which is in line with previously published results (Montenegro et al. 2022). Different phenotypic expressions of pelvic brooding traits and a differing extent of sexual dimorphism between species of the two pelvic brooding lineages, *Adrianichthys* and *Oryzias* (Spanke et al. 2021), provide further support for a scenario of an independent origin of pelvic brooding in each lineage. We found strong signals of gene flow from Lake Poso *Oryzias* into the PBO lineage (fig. 3*A* and supplementary table S12, Supplementary Material online) and a very small, but significant, amount of excess allele-sharing (0.09% and 0.08%) between both *Adrianichthys* species and the ancestor of *O. dopingdopingensis* (ODD) and the PBO (ODD + PBO) (fig. 3*A*). Besides, the most striking significant introgressions (indicated by a high, and significant, fb(C) and significant D, in combination with high z-scores, supplementary tables S12 and S13, Supplementary Material online) were evident between Malili Lakes species, from the Lake Poso species into the stem lineage of the PBO (indicated also by SNaQ, fig. 3*B*) and between *O. nigrimas* and *O. nebulosus* from Lake Poso. Further introgression, though subtle, was evident between several Central Sulawesi ricefish species (fig. 3*A*). Many of the above-mentioned species inhabit waterbodies or streams that are nowadays not connected. This, in combination with the—in some cases—rather low excess of shared alleles (all <1%), indicates that most of these patterns might be genomic footprints of ancient gene flow events.

Apart from our results, several gene flow events were already described for Sulawesi *Oryzias*, for example, locally within lake systems as in the Malili lakes including *O. marmoratus*, *O. matanensis*, and *O. profundicola* (Mandagi et al. 2021; fig. 3*A*) or between species that are now separated by large spatial distances as described for *O. sarasinorum* and *O. eversi* (Horoiwa et al. 2021) or for *O. soerotoi* in Lake Tiu and the ancestor of *O. nebulosus* and *O. orthognathus* in Lake Poso (Horoiwa et al. 2021). All this evidence, together with our data (figs. 2, 3*A* and *B*), indicates that Sulawesi ricefish evolution was likely complex, heavily influenced by gene flow in ancestral water bodies shaped by the complex geological history of the island of Sulawesi, and that it is better described in a network-like evolutionary structure (figs. 2 and 3*A*), rather than by a simple bifurcating tree.

The Quasi-independent Origin of a Complex Reproductive Strategy

The QTL of one trait associated with pelvic brooding (QCon1, Montenegro et al. 2022) lies within regions with high (top 1% f_d) signal for introgression (fig. 3*C*). This QTL is one of three QTL regions found to correlate with the expression of the ventral concavity in female ricefishes (Montenegro et al. 2022) and is the QTL with highest LOD (logarithm of the odds, >6; compared with <4 for Qcon2; and between 5 and 6 for Qcon3) and the narrowest confidence interval for this trait (4.8 cM, compared with 39.1 cM for Qcon2 and 23.6 cM for Qcon3; Montenegro et al. 2022). This overlap (supplementary fig. S5, Supplementary Material online) indicates that introduced genetic variants from the Lake Poso *Oryzias* might have played a role in the evolution of the ventral concavity. Hence, the genetic architecture of pelvic brooding traits is rather complex and potentially based on a mixture of de novo mutations and ancient variation.

Hybridization between Lake Poso species and ancestors of the PBO occurred after the splitting of *O. dopingdopingensis* and the PBO (~1.4–1.8 Ma based on Ansai et al. 2021; Mokodongan and Yamahira 2015). This coincides with the estimated age for the origin of Lake Poso, which is assumed to be less than 1–2 Myr old (von Rintelen et al. 2004; von Rintelen and Glaubrecht 2006), rendering it possible that the ancestors of present-day Lake Poso *Oryzias* and PBO species met within the Lake or in that area in rivers. Intriguingly, all PBA species are endemic to Lake Poso, which would allow an indirect transmission of pelvic brooding alleles from Lake Poso *Adrianichthys* over Lake Poso *Oryzias* into the stem lineage of the PBO. This assumption, however, has to remain speculative; although there is a significant excess of shared alleles between both *Adrianichthys* species with Lake Poso *Oryzias* species as well as with the clade comprising *O. dopingdopingensis* and the PBO, their amount is rather small (between 0.03% and 0.09%; fig. 3*A* and supplementary table S13, Supplementary Material online), and the paleo-geological history of this area is still unclear.

Based on our data, we found that a small percentage (0.46%; fig. 3*B*) of the *O. dancena* genome, a non-Sulawesi ricefish, has introgressed into the ancestor of *O. celebensis* and the younger *Oryzias* lineages hinting towards that even older hybridization events occurred between different ricefish lineages. Ancient hybridization events have the potential to serve as resource for adaptive genetic variation (Marques et al. 2019). Available ecological niche space allows selection to act upon genetic diversity originating from differently combined old alleles, generating adapted and divergent phenotypes (Seehausen 2004, 2013). For example, in the apple maggot *Rhagoletis pomonella*, introgressed genetic regions were found, deriving from an ancient hybridization with Mexican Altiplano highland fruit flies about 1.6 Ma. However, the emergence of new species using this old variation to adapt to new hosts happened more recently (Feder et al. 2003). Under this combinatorial view, reassembling old variants into new combinations promotes rapid speciation and adaptive radiation (reviewed in Marques et al. 2019). To further disentangle the genomic background of pelvic brooding in the two distantly related ricefish lineages, in-depth comparative genomic studies are needed.

Habitat Dependence—Why Did Pelvic Brooding Evolve Twice?

Irrespective of the genetic background, strong selective pressures must have shaped the evolution from ancestral transfer brooding to the extended care (of pelvic brooding), which is presumably costly in terms of increased predation risk and increased energetic costs of the care-giver (Cooke et al. 2006; Wootton and Smith 2014). A similar environment yielding similar selection pressures was proven to be important for the repeated evolution of adaptive traits (Losos et al. 1998; Arendt and Reznick 2008; Gompel and Prud'homme 2009). Present day macrohabitats of pelvic brooding ricefishes range from lakes to small karst ponds (reviewed in Parenti 2008; e.g., Herder et al. 2012; Mandagi et al. 2018); in Lake Poso, PBA mainly occupy open-water habitats, leading to the hypothesis that it evolved in adaptation to the absence of suitable spawning substrates in pelagic habitats (Herder et al. 2012). *Oryzias eversi*, however, was described from a small karst pond where potential spawning substrates are abundant (Herder et al. 2012). Thus, present day habitats make it rather complicated to define a general selective regime that might have favored the evolution of pelvic brooding. Furthermore, the paleogeographical history of Sulawesis water bodies is barely known and complicated (Wilson and Moss 1999; Hall 2001), leaving plenty of room for speculations, like presumable ancient connections between waterbodies or the existence of a paleolake in central Sulawesi (Utama et al. 2022). Thus, even an ancestral syntopic distribution of *A. oophorus*, the ancestor of *O. eversi* and *O. sarasinorum* and ancestors of the lake Poso *Oryzias*, seems possible. Given the costs and direct fitness effects of switching reproductive strategies and the rather fast convergent evolution of pelvic brooding in two distinct lineages, it appears likely that similar environmental selection pressures in combination with gene flow fueled its evolution.

Conclusions

Several gene flow events within Sulawesi ricefishes indicate that the Sulawesi ricefish radiation did not follow a tree-like evolution. In this study, we found no direct gene flow between the two distantly related pelvic brooding lineages, but detected gene flow into the stem lineage of PBO from Lake Poso *Oryzias* and a small amount of excess allele-sharing between the *Adrianichthys* species and the clade comprising the PBO and their transfer brooding sister group. The presence of elevated f_d values within the confidence interval of one QTL for the ventral concavity raises the possibility that genetic variants were recruited for the evolution of pelvic brooding in *Oryzias* via ancient hybridziation with the ancestor of the *Oryzias* species from Lake Poso.

Materials and Methods

Taxon Sampling

Fish were collected in Sulawesi between 2011 and 2013, and some species were bred in the aquarium until they were sacrificed (for details, see supplementary table S1*a*, Supplementary Material online). For dataset 1 (DS1), we

sequenced the transcriptomes of 12 ricefish species (*A. oophorus*, *O. mekongensis*, *O. dancena*, *O.* cf. *profundicola*, *O. nebulosus*, *O. javanicus*, *O. sarasinorum*, *O. eversi*, *O. nigrimas*, *O. woworae*, *O. celebensis*, and *O. matanensis*) and the genome of *O. dopingdopingensis*. Furthermore, a published genome of *O. javanicus* (GenBank Bioproject: PRJNA505405, Biosample: SAMN10417210) and the official gene sets (supplementary table S5, Supplementary Material online, used reference genomes in supplementary table S8, Supplementary Material online) of two ricefish species (*O. latipes* and *O. melastigma*) and four outgroup species (*Poecilia formosa*, *Xiphophorus maculatus*, *Nothobranchius furzeri*, and *Austrofundulus limaneus*) derived from the orthologous gene sets of Actinopterygii (TaxID 7898) from OrthoDB (supplementary table S5, Supplementary Material online) were included in our analyses.

We created a second dataset (DS2) to assess introgression on a genomic level based on published whole genome data of 26 individuals of 19 ricefish species downloaded from GenBank (supplementary table S1*b*, Supplementary Material online). Besides the pelvic-brooding species *O. eversi*, *O. sarasinorum*, *A. oophorus*, and *A. poptae*, the dataset included the transfer brooding species: *O. celebensis*, *O. dopingdopingensis*, *O. nigrimas*, *O. nebulosus*, *O. asinua*, *O. hadiatyae*, *O. matanensis*, *O. orthognathus*, *O. profundicola*, *O. soerotoi* and *O. marmoratus*, *O. wolasi*, and *O. woworae* (from different localities) (Bioproject PRJDB10385) as well as *O. javanicus* (Bioproject PRJNA505405, accession number SRR8467745) and *O. melastigma* (Bioproject PRJNA556761, accession number SRR12442554).

Sequencing and Prediction of Orthologous Genes

We extracted RNA from complete fish and prepared non-stranded Truseq mRNA libraries which were sequenced on an Illumina Hiseq2000 at the CCG in Cologne (we provide a shortened version of Materials and Methods here, details about all used methods are in the Supplementary Material). We used the Qiagen DNeasy Blood & Tissue Kit to extract DNA of *O. dopingdopingensis*. A genomic short read TruSeq DNA PCR free library was prepared by Macrogen sequencing company; finally, 15,698,917,340 bases and 103,966,340 reads were sequenced. The raw transcriptomic reads were quality-filtered using trim-fast.pl from the PoPoolation pipeline (Kofler et al. 2011) and assembled using Trinity v2.8.4. (Grabherr et al. 2011; Haas et al. 2013). A de novo whole-genome assembly of *O. dopingdopingensis* short reads was generated according to (Böhne et al. 2019) and (Malmstrøm et al. 2017) and annotated using funannotate v. 1.5.3 (Palmer and Stajich 2019) based on gene predictions of *O. latipes* and *O. melastigma*, a protein set from *O. javanicus* and the set of orthologous genes generated in this study (see supplementary methods, Supplementary Material online for more detailed description of assembly and annotation methods). Busco completeness (based on dataset Actinopterygii Odb9) ranged from 60.7% to 80.9% in the transcriptomes and was 97% in the annotated genome assembly (supplementary table S3, Supplementary Material online).

We generated a reference set consisting of 8,390 single-copy protein-coding genes representing Beloniformes and closest relatives derived from OrthoDB v.9.1 (Waterhouse et al. 2013) (*A. limnaeus*, *Centrocoris variegatus*, *Fundulus heteroclitus*, *Kryptolebias marmoratus*, *N. furzeri*, *O. latipes*, *O. melastigma*, *P. formosa*, *P. latipinna*, *P. mexicana*, *P. reticulata*, and *X. maculatus*, NCBI Accession numbers in supplementary table S5, Supplementary Material online). The hierarchical split was set to Actinopterygii (ID 7898). We then used Orthograph v0.7.1 (Petersen et al. 2017) to identify orthologous genes in our transcriptome assemblies and the annotated genome assembly. The orthologous genes were then aligned using MAFFT v7.221 with the L-INS-I algorithm (Katoh and Standley 2013) on the amino-acid level; outlier sequences (supplementary table S9, Supplementary Material online) within gene alignments were identified according to Misof et al. (2014) and subsequently removed from further analysis, retaining 7,475 genes. Keeping only genes present in each sample resulted in 2,437 genes. We then created nucleotide alignments with the amino acid alignments as a control with a modified version of Pal2Nal v14 (Suyama et al. 2006; Misof et al. 2014) and masked ambiguously aligned regions (maximum number of pairwise sequence comparisons for each multisequence alignment [option –r], and the -e option for gap-rich data sets, leaving remaining settings to defaults) in 1,806 genes with ALISCORE v2.0 (supplementary table S10*a* and *b*, Supplementary Material online); 631 genes did not need masking. Subsequently we removed all sections, which were indicated as randomly similar aligned (16.5% of all base pairs) with ALICUT v2.3 (Misof and Misof 2009; Kück et al. 2010; Misof et al. 2014).

Phylogenetic and Coalescence Analyses

We used IQ-TREE v1.6.12 to create a tree for each of the 2,437 remaining genes separately (Nguyen et al. 2014) (for details on the methods described in this paragraph, see Supplementary Material). For each gene, a substitution model was estimated (supplementary table S11, Supplementary Material online). For 23 genes, the information content was too low after masking for IQ-TREE to estimate a substitution model and hence was removed from the analysis. The resulting 2,414 gene trees were tested for paralogous sequences with Phylotreepruner (Kocot et al. 2013), which returned no suspicious genes. Additionally, we ran TreeShrink (Mai and Mirarab 2018) on each gene tree. TreeShrink identifies outlier species in a tree by evaluating its impact on tree diameter, which gives

the maximum distance between any of two leaves. Gene trees with sequences that had an unexpectedly high impact on the tree diameter compared with the other species were labeled as outliers. We removed all marked gene trees ($N = 507$) after additional manual inspection resulting in 1,907 orthologous genes (2.415.561 bp) present in all 16 ricefish and four outgroup species (will be deposited on Dryad after acceptance).

We used Astral v5.7.3 with default parameters to run a multispecies coalescent model on all gene trees generated by IQ-tree as input (Zhang et al. 2018) and displayed the results using DiscoVista (Sayyari et al. 2018). To display the potential of gene-tree–species-tree incongruencies in Densitree v2.01 (Bouckaert and Heled 2014), gene trees were made ultrametric using the function "chronos" in R v4.1.2 (package "ape" v5.6-2 (Paradis and Schliep 2019). For phylogenetic reconstruction, genes were ordered according to their position on the reference genome of *O. latipes* (RefSeq GCF_002234675.1, Bioprject PRJNA325079) and in this order concatenated using FASconCAT-G (Kück and Longo 2014). The resulting supermatrix (2.415.561 bp) was partitioned, and a suitable model was searched for each partition taking codon position into account using IQ-TREE. We ran 20 single tree searches and 50 nonparametric, slow bootstrap replicates.

Finally, we did an ancestral state reconstruction in R using the package "phytools" (Revell 2012) with the implemented function "rerootingMethod" based on the rerooting method of (Yang et al. 1995). We assigned a state of pelvic brooding to *O. eversi*, *O. sarasinorum*, and *A. oophorus* and transfer brooding to the other ricefishes. *Oryzias latipes* was used as outgroup. We used the model "ER" as recommended by the author (Revell 2012).

Gene Flow Analyses

Orthologous genes (DS1) were used as input to estimate a phylogenetic network with SNaQ. SNaQ is part of the PhyloNetworks package (Solís-Lemus et al. 2017) and performs maximum pseudolikelihood estimation of phylogenetic networks using the multispecies coalescent model on networks (Meng and Kubatko 2009; Yu et al. 2014). The approach incorporates uncertainty in estimated gene trees and gene tree discordance due to ILS and considers quartet topologies only, which makes the approach more robust to rate variation across genes and across lineages (Solís-Lemus and Ané 2016). Whole genome sequences (DS 2) were mapped on the *O. celebensis* reference genome from Genbank (GCA_014656515.1, Bioproject PRJDB10371 (Ansai et al. 2021)) using bowtie2 (Langmead and Salzberg 2012), and SNPs were called using mpileup from bcftools (Li 2011). After filtering using bcftools (QUAL > 25, DP > 30, MQ > 25, only SNPs), 38.183142 SNPs remained. Dtrios from Dsuite v0.4r41 was used to calculate the D (ABBA-BABA) and f4-ratio statistics for all possible trios of ricefish species with *O. javanicus* (Bioproject PRJNA505405, accession number SRR8467745) as outgroup (Malinsky et al. 2021) (for detailed methods, see supplementary methods, Supplementary Material online).

The "f-branch" metric or $f_b(C)$ implemented in the Dsuite package (Malinsky et al. 2018) was applied on a neighbor joining tree generated with the R package Ape v. 5.6-2 (Paradis and Schliep 2019) based on genome wide SNPs (DS2), including only the Sulawesi ricefish species. The resulting topology matches the ML tree (fig. 1). The summary of f-scores on the tree captures the excess allele-sharing between a species C and a branch b compared with the sister branch of b (fig. 3*A*, *x* axis). An $f_b(C)$ score is thus specific to the branch b (on the *y* axis in fig. 3*A*), but a single introgression event can still lead to significant $f_b(C)$ values across multiple related C values (Malinsky et al. 2018). DInvestigate was used to identify significantly elevated f_d values in sliding windows of 50 informative SNPs, followed by steps of 25 informative SNPs (which are the default options); here, we defined *O. dopingdopingensis* as one population (P1), *O. eversi* and *O. sarasinorum* as one (pelvic brooders, P2), the Poso/Tiu *Oryzias* (*O. nebulosus*, *O. nigrimas*, *O. orthognathus*, *O. soerotoi*) as one population (P3), and *A. poptae* as outgroup (it was chosen, because it is the most distant relative but still a ricefish from Sulawesi, and there was no introgression detected in this species using Dtrios).

To determine chromosomes with the strongest signal of introgression, we counted the number of windows with top 1% f_d values per chromosome and divided them by the number of total windows per chromosome (supplementary fig. S3, Supplementary Material online). The regions of elevated f_d values were compared with the quantitative trait loci (QTLs) associated with pelvic fin length, the extent of the concavity, and duration of egg carrying found in (Montenegro et al. 2022).

To test whether the observed accumulation of high f_d values (top 1%) in previously published QTL intervals associated with pelvic-brooding (Montenegro et al. 2022; supplementary fig. S5, Supplementary Material online) differs from what is expected by chance, we randomly sampled intervals in the size of the QTLs (e.g., QCon1 had 169 windows of 50 informative SNPs) within each chromosome 10,000 times and calculated how often we found the same or a higher amount of windows with a high f_d value within the intervals.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We thank Rob Waterhouse for the advice on using OrthoDB. Further, we are thankful for the help of Alexandros

Vasilikopoulos with compiling the ortholog set. We thank Michael Matschiner and Thore Koppetsch for their insights in the pitfalls of different evolutionary rates in hybridization analysis and their recommendations. We thank Jan Möhring and two anonymous reviewers for the very helpful and constructive feedback that clearly improved our manuscript. We are very grateful for the permission to use the beautiful photo of *O. eversi* taken by Hans Evers for this publication. Many thanks to the Zoology Division of Research Center for Biology, Indonesian Institute of Sciences (LIPI) (presently Research Center for Biology, National Research and Innovation Agency or Badan Riset dan Inovasi Nasional [BRIN]), and the Kementerian Negara Riset dan Teknology (RISTEK) for permission to conduct research in Indonesia. Calculations for the genome assembly were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at University of Basel. Further, we used the annotation scripts from the Sigenae platform for the genome annotation. This work was supported by the Leibniz Association, grant P91/2016.

Benefit-Sharing Statement

Benefits generated: A research collaboration was developed with scientists from the countries providing genetic samples, all collaborators are included as coauthors, and the results of research have been shared with the provider communities and the broader scientific community (see above). More broadly, our group is committed to international scientific partnerships, as well as shares knowledge about the establishing and maintaining of scientific collections.

Author Contributions

J.M.F., J.S., and K.M. designed research. A.W.N. and F.H. were leading the field expeditions in Sulawesi where the samples were collected. A.W.N. provided the transcriptome sequences. L.H. did the transcriptome assemblies. J.M.F., K.M., and J.S. performed research and analyzed data. A.B. did the genome assembly. S.M. annotated the genome and supported J.M.F. with the NCBI submission process. J.M.F. and J.S. wrote initial draft of the manuscript. All authors contributed to writing the final manuscript.

Data Availability

Genetic data: Raw sequence reads are deposited in the SRA (see supplementary table S2, Supplementary Material online). Genome annotation data are available on zenodo (10.5281/zenodo.8064517). Sample metadata: Related metadata can be found in Supplementary Material (supplementary table S1a, Supplementary Material online).

Literature Cited

Ansai S, et al. 2021. Genome editing reveals fitness effects of a gene for sexual dichromatism in Sulawesian fishes. Nat Commun. 12(1):1350.
Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol Evol. 23(1):26–32.
Arnold ML. 2007. Evolution through genetic exchange. USA: Oxford University Press.
Baack EJ, Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev. 17:513–518.
Bainbridge DRJ. 2014. The evolution of pregnancy. Early Hum Dev. 90(11):741–745.
Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. Trends Ecol Evol. 23:38–44.
Böhne A, et al. 2019. Repeated evolution versus common ancestry: sex chromosome evolution in the haplochromine cichlid *Pseudocrenilabrus philander*. Genome Biol Evol. 11(2):439–458.
Bouckaert RR, Heled J. 2014. DensiTree 2: seeing trees through the forest. bioRxiv 012401.
Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genet. 4:e1000168.
Colosimo PF, et al. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307(5717): 1928–1933.
Cooke SJ, Philipp DP, Wahl DH, Weatherhead PJ. 2006. Energetics of parental care in six syntopic centrarchid fishes. Oecologia 148(2): 235–249.
Darwin C. 1859. The origin of species by means of natural selection. London: John Murray.
Edelman NB, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. Science 366(6465):594–599.
Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends Ecol Evol. 26(6):298–306.
Feder JL, et al. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. Proc Natl Acad Sci U S A. 100(18):10314–10319.
Foster SA, Baker JA. 2004. Evolution in parallel: new insights from a classic system. Trends Ecol Evol. 19(9):456–459.
Gani A, et al. 2022. A new endemic species of pelvic-brooding ricefish (Beloniformes: Adrianichthyidae: Oryzias) from Lake Kalimpa'a, Sulawesi, Indonesia. Bonn Zool. Bull. 71:77–85.
Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. Dev Biol. 332(1):36–47.
Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.
Grant PR, Grant BR, Markert JA, Keller LF, Petren K. 2004. Convergent evolution of Darwin's Finches caused by introgressive hybridization and selection. Evolution 58(7):1588–1599.
Greenway R, et al. 2020. Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. Proc Natl Acad Sci U S A. 117(28):16424–16430.
Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8(8):1494–1512.
Hall R. 2001. Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and sea. In: Ian Metcalfe, Jeremy Smith, Mike Morwood, Iain Davidson (Eds.), *Faunal and floral migrations and evolution in SE Asia–Australasia*. Darwin, Australia: CRC Press. p. 35–56.

The Heliconius Genome Consortium, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487(7405):94.
Herder F, Hadiaty RK, Nolte AW. 2012. Pelvic-fin brooding in a new species of riverine ricefish (Atherinomorpha: Beloniformes: Adrianichthyidae) from Tana Toraja, Central Sulawesi, Indonesia. Raffles Bull Zool. 60(2):467–476.
Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352.
Hilgers L, Schwarzer J. 2019. The untapped potential of medaka and its wild relatives. Elife 8:1–14.
Horoiwa M, et al. 2021. Mitochondrial introgression by ancient admixture between two distant lacustrine fishes in Sulawesi Island. PLoS One. 16(6):1–14.
Jones MR, Mills LS, Jensen JD, Good JM. 2020. Convergent evolution of seasonal camouflage in response to reduced snow cover across the snowshoe hare range*. Evolution 74(9): 2033–2045.
Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kocot KM, Citarella MR, Moroz LL, Halanych KM. 2013. Phylotreepruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. Evol Bioinform Online. 9:429–435.
Kofler R, et al. 2011. Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One. 6(1):e15925.
Kottelat M. 1990. Synopsis of the endangered Buntingi (Osteichthyes: Adranichtyidae and Oryziidae) of Lake Poso, Central Sulawesi, Indonesia, with a new reproductive guild and descriptions of three new species. Ichthyol Explor Freshw. 1:46–67.
Kück P, et al. 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front Zool. 7(1):10.
Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool. 11(1):81.
Lamichhaney S, et al. 2015. Evolution of Darwin's Finches and their beaks revealed by genome sequencing. Nature 518(7539): 371–375.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359. http://dx.doi.org/10.1038/nmeth.1923
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987–2993. http://dx.doi.org/10.1093/bioinformatics/btr509
Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279(5359):2115–2118.
Mai U, Mirarab S. 2018. Treeshrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics. 19(S5):272.
Malinsky M, Matschiner M, Svardal H. 2021. Dsuite—fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 21(2):584–595.
Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nature Ecology & Evolution 2(12):1940–1955. http://dx.doi.org/10.1038/s41559-018-0717-x
Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol Evol. 20(5):229–237.
Malmstrøm M, Matschiner M, Tørresen OK, Jakobsen KS, Jentoft S. 2017. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. Sci Data. 4(1):160132.
Mandagi IF, et al. 2021. Species divergence and repeated ancient hybridization in a Sulawesian lake system. J Evol Biol. 34(11): 1767–1780.
Mandagi IF, Mokodongan DF, Tanaka R, Yamahira K. 2018. A new riverine ricefish of the genus *Oryzias* (Beloniformes, Adrianichthyidae) from Malili, Central Sulawesi, Indonesia. Copeia 106(2):297–304.
Marques DA, Meier JI, Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. Trends Ecol Evol. 34(6): 531–544.
Meier JI, et al. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat Commun. 8(1):14363.
Meng C, Kubatko LS. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. Theor Popul Biol. 75(1):35–45. http://dx.doi.org/10.1016/j.tpb.2008.10.004
Misof B, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346(6210):763–767.
Misof B, Misof K. 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Syst Biol. 58(1):21–34.
Mokodongan DF, Yamahira K. 2015. Origin and intra-island diversification of Sulawesi endemic Adrianichthyidae. Mol Phylogenet Evol. 93:150–160.
Montenegro J, et al. 2022. Genetic basis for the evolution of pelvic-fin brooding, a new mode of reproduction, in a Sulawesian fish. Mol Ecol. 31(14):3798–3811.
Nelson TC, Cresko WA. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evol Lett. 2(1):9–21.
Nguyen L-T, Schmidt HA, von Haeseler A, Quang Minh B. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.
Ohta T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 23:263–286..
Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation. http://dx.doi.org/10.5281/zenodo.1134477
Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3): 526–528.
Parenti LR. 2008. A phylogenetic analysis and taxonomic revision of rice fishes, *Oryzias* and relatives (Beloniformes, Adrianichthydae). Zool J Linn Soc. 154:494–610.
Petersen M, et al. 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics. 18(1):111.
Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things): *phytools: R package*. Methods Ecol Evol. 3(2):217–223.
Rieseberg LH. 2009. Evolution: replacing genes and traits through hybridization. Curr Biol. 19(3):R119–R122.
Sayyari E, Whitfield JB, Mirarab S. 2018. Discovista: interpretable visualizations of gene tree discordance. Mol Phylogenet Evol. 122: 110–115.
Schluter D, Nagel LM. 1995. Parallel speciation by natural selection. Am Nat. 146(2):292–301.
Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol Evol. 19(4):198–207.

Seehausen O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. J Evol Biol. 26(2):279–281.
Solís-Lemus C, Ané C. 2016. Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. PLoS Genet. 12(3):e1005896. http://dx.doi.org/10.1371/journal.pgen.1005896
Solís-Lemus C, Bastide P, Ané C. 2017. Phylonetworks: a package for phylogenetic networks. Mol Biol Evol. 34(12):3292–3298.
Spanke T, et al. 2021. Complex sexually dimorphic traits shape the parallel evolution of a novel reproductive strategy in Sulawesi ricefishes (Adrianichthyidae). BMC Ecol Evol. 21(1):57.
Stern DL. 2013. The genetic causes of convergent evolution. Nat Rev Genet. 14(11):751–764.
Sutra N, et al. 2019. Evidence for sympatric speciation in a Wallacean ancient lake. Evolution 73(9):1898–1915.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34(Web Server issue): W609–W612.
Terekhanova NV, et al. 2014. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. PLoS Genet. 10(10):e1004696.
Utama IV, et al. 2022. Deeply divergent freshwater fish species within a single river system in central Sulawesi. Mol Phylogenet Evol. 173: 107519.
Van Belleghem SM, et al. 2018. Evolution at two time frames: polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. PLoS Genet. 14(11):e1007796.
Veale AJ, Russello MA. 2017. Genomic changes associated with reproductive and migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). Genome Biol Evol. 9(10):2921–2939.
von Rintelen T, Glaubrecht M. 2006. Rapid evolution of sessility in an endemic species flock of the freshwater bivalve *Corbicula* from ancient lakes on Sulawesi, Indonesia. Biol Lett. 2(1):73–77.
Von Rintelen T, Wilson AB, Meyer A, Glaubrecht M. 2004. Escalation and trophic specialization drive adaptive radiation of freshwater gastropods in ancient lakes on Sulawesi, Indonesia. Proc Biol Sci. 271(1557):2541–2549.
Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41(D1):D358–D365. http://dx.doi.org/10.1093/nar/gks1116
Waters JM, McCulloch GA. 2021. Reinventing the wheel? Reassessing the roles of gene flow, sorting and convergence in repeated evolution. Mol Ecol. 30(17):4162–4172.
Whitney KD, Randell RA, Rieseberg LH. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. Am Nat. 167:794.
Wilson MEJ, Moss SJ. 1999. Cenozoic palaeogeographic evolution of Sulawesi and Borneo. Palaeogeogr Palaeoclimatol Palaeoecol. 145(4):303–337.
Wootton R, Smith C. 2014. Reproductive biology of teleost fishes. West Sussex, UK: John Wiley & Sons.
Yamamoto T. 1975. Medaka: (Killifish). Biology and strains. Series of stock culture in biological field. Tokyo, Japan: Keigaku Publishing Company.
Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141(4):1641–1650.
Yu Y, Dong J, Liu KJ, Nakhleh L. 2014. Maximum likelihood inference of reticulate evolutionary histories. Proc Natl Acad Sci. 111(46): 16448–16453. http://dx.doi.org/10.1073/pnas.1407950111
Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 19(S6):153.
Zhang X, Rayner JG, Blaxter M, Bailey NW. 2021. Rapid parallel adaptation despite gene flow in silent crickets. Nat Commun. 12(1):50.

Associate editor: Bonnie Fraser