Ancient and recent riverine gene flow contributed to the adaptive radiation of sailfin silversides in Wallace's Dreampond

De Keyzer E L R[1], Herder F[2], Böhne A[2], Campuzano Jiménez F[1], Burskaia V[1], Kukowka S[2], Tracey A[3], Denton A[3], Oatley G[3], Wellcome Sanger Institute Tree of Life programme[3,4], Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective[3,5], Tree of Life Core Informatics collective[3,6], Mokodongan D F[7], Wowor D[7], Svardal H[1,8]

[1] Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

[2] Leibniz-Institute for the Analysis of Biodiversity Change (LIB), Museum Koenig Bonn, 53113 Bonn, Germany

[3] Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK

[4] members: https://doi.org/10.5281/zenodo.4783585

[5] members: https://doi.org/10.5281/zenodo.4790455

[6] members: https://doi.org/10.5281/zenodo.7116866

[7] Museum Zoologicum Bogoriense, Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, 16911, Indonesia

[8] Naturalis Biodiversity Center, 2333 Leiden, The Netherlands

Corresponding author: Els L. R. De Keyzer

Email: els.dekeyzer@uantwerpen.be

Keywords

Telmatherina, hybridisation, speciation, hybrid zone, phylogenomics

Abstract

While adaptive radiations significantly contribute to the world's biodiversity, much is unknown about the genetic and ecological factors underlying these rapid successions of speciation. It has been suggested that hybridisation can facilitate the speciation process by generating genetic diversity on which diversifying selection can act. Sailfin silverside fishes (Telmatherinidae) in the Malili Lakes system in Sulawesi have diversified within the last 2 million years. We assembled and annotated a chromosome-scale reference genome of the riverine sailfin silverside *Telmatherina bonti* and generated whole genome sequences of all species of *Telmatherina* in Lake Matano, South Sulawesi, Indonesia, one of the world's oldest and deepest lakes. We reconstructed the phylogenetic relationships within the adaptive radiation of sailfin silversides and inferred past and ongoing introgression patterns. Genome-wide tests confirmed two monophyletic clades, sharpfins and roundfins. However, within clades, we found mismatches between morphology-based taxonomic assignments and genome-wide genetic relationships. We found signs of both old and ongoing introgression between river-dwelling *T. bonti* and the lacustrine sharpfin group, as shown in elevated D-statistic, f4-ratio and f-branch statistic. Levels of excess allele sharing between riverine species and the three most common lacustrine species declined with increasing distance from the river-inlet, indicating ongoing introgression at the lake-river interface. This combination of past and ongoing hybridisation in a radiating species flock makes Lake Matano *Telmatherina* a particularly valuable system to study fundamental mechanisms driving rapid speciation under genomic exchange. The phylogenomic framework elaborated in this study provides the foundation for studies of the processes shaping this charismatic radiation.

Introduction

An important aspect of biodiversity is how new species arise and establish. Adaptive radiations, rapid bursts of ecological speciation originating from a small number of ancestral lineages, have significantly contributed to the world's biodiversity (Berner and Salzburger 2015). The molecular mechanisms underlying these rapid speciation events are still largely unclear. Specifically, it is not well known what genetic and ecological factors underlie the difference between radiating lineages and lineages which do not diversify despite apparent ecological opportunity. Understanding phylogenomic relationships and patterns of evolutionary history of an adaptive radiation opens opportunities to study mechanisms of speciation within the radiation.

The genomic diversity required for rapid speciation can be obtained through introgressive hybridization with a related lineage (Seehausen 2004; Arnold 2006; Nolte and Tautz 2010). This was for example the case in Darwin’s finches, for which introgressive hybridization is the major factor contributing to genomic diversity (Lamichhaney et al. 2015). Hybridisation has often been detected within young adaptive radiations (Malinsky et al. 2018; Kozak et al. 2021). Although genetic exchange can break down reproductive isolation, potentially slowing down or reversing speciation (Seehausen et al. 2008), recent studies emphasise its potential role as a source of genetic variation. Introgression can even give rise to novel traits, increasing adaptive potential and thereby facilitating speciation (Seehausen 2004; Seehausen 2013; Marques et al. 2019; Peñalba et al. 2024). Due to the complexity of the evolutionary mechanisms underlying adaptive radiation, there are still many unknown factors about the molecular basis of adaptive radiation and natural systems in which the effects of ongoing genetic exchange can be studied are rare. To get a clear understanding of the genomic makeup and especially the role of introgression in an ongoing radiation, ideally these processes need to be studied in an ecologically well documented system with

relatively smaller complexity compared to the larger and often interconnected adaptive radiations such as East African cichlid fishes or Heliconius butterflies.

The Malili Lakes in Central Sulawesi, Indonesia, were termed Wallace's Dreamponds (Herder, Nolte, et al. 2006) for their status as a hotspot of aquatic biodiversity, with adaptive radiations and high degrees of endemism in snails, shrimp, crabs, and fishes (Vaillant et al. 2011; von Rintelen et al. 2012). Sailfin silverside fishes (Telmatherinidae) have diversified in the lakes within the last 1-2 million years (Stelbrink et al. 2014). Previous studies suggest 19 distinct species (of which 15 are formally described) with distinctive phenotypic and ecological adaptations (Herder, Schwarzer, et al. 2006). In Lake Matano, the most upstream and deepest lake of the Malili Lakes system, an adaptive radiation of sailfin silversides has produced species that conquered a diverse set of ecological niches (Herder et al. 2008; Pfaender et al. 2016). Lake Matano sailfin silversides can be attributed to two ecologically and morphologically well-defined groups distinguished by the shape of the second dorsal and the anal fin (Kottelat 1991), (Herder, Nolte, et al. 2006); the bentho-pelagic roundfins and the predominantly benthic sharpfins (Herder, Schwarzer, et al. 2006), shown to be genetically distinct based on AFLP markers (Herder and Schliewen 2010). Diversification within the roundfin and sharpfin lineages had been dated to approximately 1.0 and 0.9 million years ago (Mya), respectively, using mitochondrial markers (Stelbrink et al. 2014). Both clades have diversified and inhabit a variety of ecological niches, with corresponding morphological adaptations related to resource use, including variations in body shape and size, oral and pharyngeal jaw structure and the length of gill rakers (Pfaender et al. 2010; Wasiljew et al. 2021). Morphological differentiation is complex, covering different structures rather than single key innovations (Pfaender et al. 2010).

Two species of roundfins have formally been described: *T. prognatha* Kottelat, 1991, and *T. antoniae* Kottelat, 1991 (Kottelat 1991), yet large differences in morphology, AFLP markers, trophic ecology, habitat use as well as the observation of assortative mating suggest *T.*

antoniae consists of two separate species (Herder et al. 2008; Pfaender et al. 2011; Wasiljew et al. 2021), referred to as *T. antoniae* 'small' and *T. antoniae* 'large'. Roundfin *Telmatherina* have, to different degrees, a pelagic ecology, contrasted to the predominantly epibenthic sharpfins (Herder et al 2008). The sharpfin species flock is the most diverse lineage of sailfin silversides in terms of adaptive traits (Herder, Schwarzer, et al. 2006; Pfaender et al. 2010; Pfaender et al. 2016). The six reported species of sharpfins (of which four are formally described) (Herder, Schwarzer, et al. 2006), have mainly been distinguished by morphometric and meristic typing (Kottelat 1991). Trophic groups partly coincide with the species groups, and include insect (*T. opudi* Kottelat, 1991), shrimp (*T.* sp. 'thicklip') and mollusc (*T. wahjui* Kottelat, 1991) feeders, fish predators (*T.* sp. 'elongated', *T. abendanoni* Weber, 1913) and a specialised fish egg predator (*T. sarasinorum* Kottelat, 1991) (Pfaender et al. 2010; Pfaender et al. 2016). Sharpfins show species-specific behaviour, often related to feeding strategies (Cerwenka et al. 2012) and mating (Gray and McKinnon 2006). To some extent, a correlation between genotypic and phenotypic variation has been observed (Schwarzer et al. 2008; Pfaender et al. 2016), but previous genetic investigations have been limited to AFLP and mitochondrial data. At the same time, trophic niches of sharpfin species widely overlap and there is substantial overlap across species in morphological variation (Pfaender et al. 2016), indicative of a young, evolving species flock (Herder et al. 2008; Pfaender et al. 2010).

Although the majority of morphological and species diversity within *Telmatherina* is represented by lacustrine species, riverine populations are common in the rivers around the Malili lakes (Herder, Schwarzer, et al. 2006). *T. wahjui* from the outlet river Petea and the area in Lake Matano around Petea river displays physical and behavioural adaptations to a riverine habitat, such as a wider head and adaptations in mating behaviour (Gray and McKinnon 2006). Riverine populations of *T.* cf. *bonti* occur in the streams surrounding and

122 between the Malili Lakes (Herder, Schwarzer, et al. 2006; Herder, Nolte, et al. 2006). Finally,
123 *T. bonti* Weber & de Beaufort, 1922 has been reported from Lawa river, the inlet stream of
124 Lake Matano (Hadiaty and Wirjoatmodjo 2002). *T. bonti* are ram feeders, specialised on
125 obtaining small food items in their fast-flowing stream habitats (Wasiljew et al. 2022).

126

127 Reproductive isolation among different species of sailfin silversides is substantial (Herder et
128 al 2006a, 2008), yet there are signals of ongoing gene flow between some species (Herder,
129 Nolte, et al. 2006; Herder et al. 2008; Schwarzer et al. 2008; Stelbrink et al. 2014). Previous
130 marker-based efforts failed to completely resolve phylogenetic relationships between
131 *Telmatherina* species and mixed signals were interpreted as gene flow both within sharpfins
132 and within roundfins (Herder et al. 2008; Stelbrink et al. 2014). There are indications of high
133 levels of introgression between lacustrine and riverine lineages (Herder, Nolte, et al. 2006;
134 Schwarzer et al. 2008), including between the riverine *T. wahjui* population in river Petea
135 and lacustrine sharpfins in Lake Matano (Schwarzer et al. 2008). Sharpfins were found to be
136 monophyletic based on AFLP data (Herder, Nolte, et al. 2006), yet they cluster into two
137 mitochondrial haplogroups (Stelbrink et al. 2014). While some individuals carry haplotypes
138 with a sister group relationship to the lake's roundfins, others carry mitochondria related to
139 riverine lineages. This signal was interpreted as riverine introgression early in the divergence
140 of sharpfin silversides (Herder, Nolte, et al. 2006; Stelbrink et al. 2014), or even ongoing
141 hybridization (Schwarzer et al. 2008). However, given the limited number of investigated
142 markers this data is not sufficient to provide a conclusive picture. A better understanding of
143 the complete evolutionary history of these species will require expanding information through
144 obtaining high resolution genome wide markers.

145

146 Disentangling the evolutionary history of the adaptive radiation of sailfin silversides, including
147 past and ongoing introgression, is an important step towards developing this system as a

model to study speciation mechanisms. In this study we assembled a reference genome of the riverine *T. bonti* and investigated the phylogenomic relationships and patterns of past and ongoing introgression between sailfin silversides (*Telmatherina*) from Lake Matano, Sulawesi and from stream populations. Using whole genome sequences of 36 specimens attributed to eleven species of sailfin silversides, we tested for signals of past introgression from riverine species into Lake Matano sharpfins, relative to their sister group roundfins, and presumed hybridisation on ecological timescales between riverine and lake species. Our analysis reveals a complex evolutionary history of Lake Matano *Telmatherina* and implicates a role of both ancient and ongoing introgression in shaping and maintaining current genetic diversity.

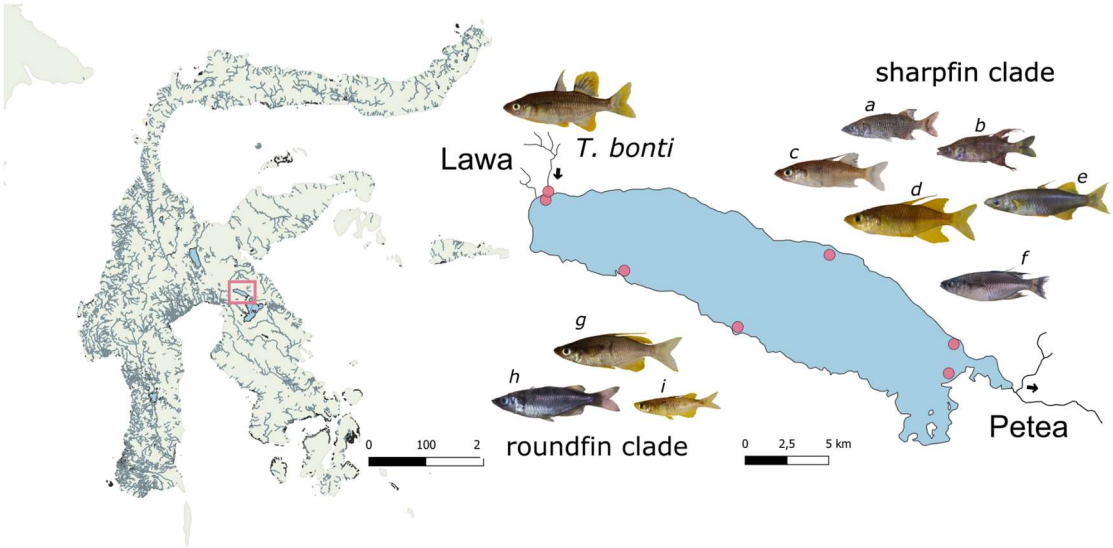
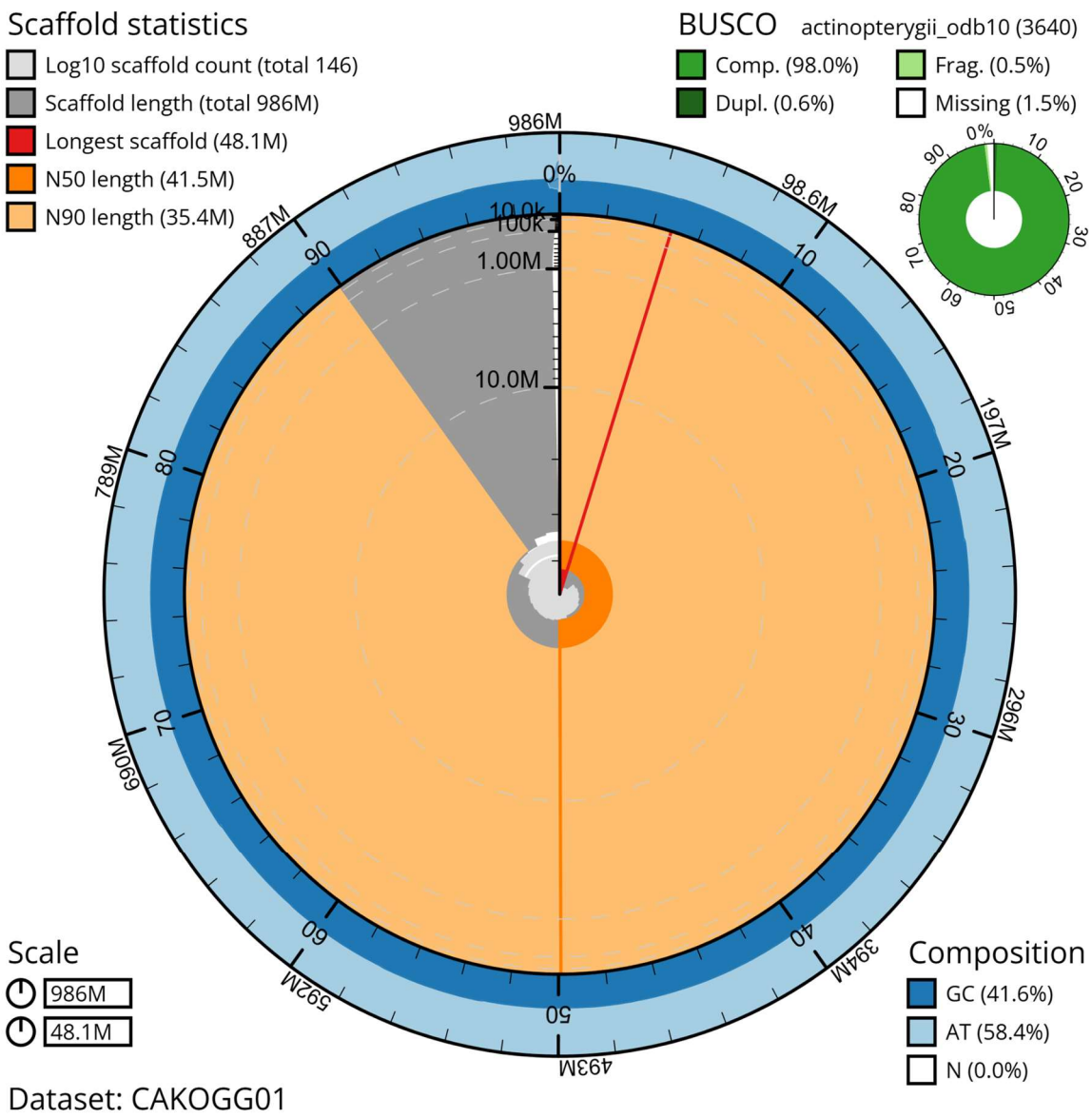


Fig. 1 - **Map of sampling locations.** Location of Lake Matano in the Malili Lakes system, Sulawesi (left). Detail of lake Matano and the inflow river Lawa and outflow river Petea (right). Dots represent the six sampling sites along Lake Matano and one in the inflow Lawa river. Pictures depict the six species of sharpfin Telmatherina (*a T. abendanoni* (n = 3)*, b T.* sp. 'thicklip' (n = 3)*, c T. opudi* (n = 6)*, d T. sarasinorum* (n = 6)*, e T.* sp. 'elongated' (n = 3)*, f T. wahjui* (n = 3)) and three species of roundfin Telmatherina *(g T. antoniae* 'large' (n = 3), *h T. prognatha* (n = 3)*, i T. antoniae* 'small' (n = 3)) and the riverine *T. bonti* (n = 3)*.*

166 Results

167 **High quality reference genome and annotation for *Telmatherina bonti***

168 A high-quality reference genome of *T. bonti* (fTelBon1.1, GCA_933228915.1) was
169 constructed as part of the tree of life project (https://www.sanger.ac.uk/programme/tree-of-
170 life/) (supplementary figure S1, supplementary figure S2). The final assembly had a total
171 length of 986,039,709 bp in 146 sequence scaffolds with scaffold N50 and N90 sequence
172 lengths of 41,523,244 and 35,388,860 bp respectively (fig. 2a). 99.6% of the assembly
173 sequence was assigned to 25 chromosomal-level scaffolds, representing 24 autosomes,
174 numbered by sequence length, and the mitochondrion (Supplementary table S1). The
175 assembly had a BUSCO completeness of 98% using the actinopteryggi_odb10 reference set
176 (fig. 2b). The Ensembl gene annotation system (Aken et al. 2016) was used to generate an
177 annotation for the fTelBon1.1 assembly. The resulting annotation included 47,104 transcripts
178 assigned to 21,413 coding and 4,664 non-coding genes.


Scaffold statistics
Log10 scaffold count (total 146)
Scaffold length (total 986M)
Longest scaffold (48.1M)
N50 length (41.5M)
N90 length (35.4M)
BUSCO actinopterygii_odb10 (3640)
Comp. (98.0%)
Frag. (0.5%)
Dupl. (0.6%)
Missing (1.5%)
986M
0%
98.6M
197M
296M
394M
493M
592M
690M
789M
887M
10.0k
100k
1.00M
10.0M
Scale
986M
48.1M
Composition
GC (41.6%)
AT (58.4%)
N (0.0%)
Dataset: CAKOGG01

179

Fig. 2 - **Assembly statistics for reference genome assembly of *Telmatherina bonti***. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 986,039,709 bp assembly. The distribution of sequence lengths is shown in dark grey (length measured from outside to inside) with the plot radius scaled to the longest sequence present in the assembly shown in red (48,058,054 bp). Orange and pale-orange arcs show the N50 and N90 sequence lengths (41,523,244 and 35,388,860 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. Top right: summary of complete, fragmented, duplicated and missing BUSCO genes in the actinopterygii_odb10 set. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Telmatherina%20bonti/dataset/CAKOGG01/snail.

Genome data supports roundfin and sharpfin clades and most taxa

To investigate whether previous morphology-based taxonomic assignments are consistent with genome-wide genetic relationships, we analysed whole genome sequences of 36 *Telmatherina* specimens from Lake Matano (fig. 1, supplementary figure S3, Supplementary table S2). We constructed a neighbour joining (NJ) tree based on genome-wide pairwise differences (supplementary figure S4). In addition, we built genome-wide and mitochondrial maximum-likelihood (ML) phylogenies and performed PCA (fig. 3-4). All methods showed a clear separation between the bentho-pelagic roundfins and the predominantly epibenthic sharpfins. Within roundfins we found clear support for the previously identified ecomorphological groups *T. prognatha*, *T. antoniae* ‘small’, and *T. antoniae* ‘large’, which each formed monophyletic groups in the ML consensus tree and the genome-wide NJ tree (fig. 3a, supplementary figure S4) and well-defined clusters in a PC analysis only including roundfin samples (supplementary figure S5). We found similarly strong support for three sharpfin species *T.* sp. ‘elongated’, *T.* sp. ‘thicklip’, and *T. wahjui* (fig. 3c, fig. 4a-c). Conversely, samples of the most common sharpfin species in Lake Matano, *T. sarasinorum*, *T. opudi*, and *T. abendanoni*, did not show any consistent pattern of grouping by taxonomic assignment in any of the phylogenetic trees (fig. 2,

212 supplementary figure S4) or PC analyses (fig. 4b-c). Instead, those individuals showed a
213 relatively large spread intermediate between *T.* sp. ‘thicklip’ and *T.* sp. ‘elongated’ along PC2
214 of an analysis only including sharpfin samples (fig. 4c).

215

216 **Complex relationships between lake and river dwelling species**

217 Our sample set contained members of two stream-dwelling species: *T. wahjui* from Petea
218 river, the outflow of Lake Matano, and *T. bonti* from Lawa river, the major tributary of Lake
219 Matano. As mentioned above, *T. wahjui* clearly belongs to the sharpfin genetic clade, in
220 which they fall basally to all other sharpfins (fig. 3a). That said, there was a relatively large
221 spread of these individuals in the PC analysis of sharpfins (fig. 4b) consistent with variable
222 ancestry proportions. *T. bonti*, the second stream dwelling species, showed an even more
223 extreme spread among samples in the PC analysis, albeit along different axes of variation
224 (PC2 in the analysis of all samples, fig. 4a; PC1 in an analysis of only sharpfins and *T. bonti*,
225 fig. 4b). Consistent with this, *T. bonti* samples also showed inconsistent phylogenetic
226 clustering for different inference methods (fig. 3, supplementary figure S4). Specifically, while
227 in the genome-wide neighbour joining tree (supplementary figure S4) the three samples
228 formed a sister group to the sharpfins, they formed an outgroup to both sharpfins and
229 roundfins in the ML consensus tree (fig. 3a). An astral consensus of NJ based local distance
230 trees, on the other hand, gave an intermediate pattern with two *T. bonti* samples clustering
231 as a global outgroup and one as a sister group to sharpfins (supplementary figure S4).
232 Together, these analyses suggest that river-dwelling *Telmatherina* show within-population
233 variation in their genetic relationships with the lacustrine species groups.

234

235 **Mitochondrial phylogeny splits into three haplotype groups**

236 Our sample set was divided into three groups based on mitochondrial haplotypes (fig. 3b).
237 One group, consisting entirely of roundfin samples, was in a sister relationship with a group
238 that included approximately half of the sharpfin samples. The third group contained the
239 remaining sharpfin samples along with all *T. bonti* samples. A similar pattern was reported
240 by Stelbrink et al. (2016), who referred to the roundfin and sharpfin haplotypes in the sister
241 group as "the original Matano haplotype". All samples of the stream-dwelling sharpfin
242 species *T. wahjui* and the lacustrine predator *T. sp.* 'elongated' shared a haplotype similar to
243 *T. bonti*. In contrast, all samples of the shrimp-eating sharpfin species *T. sp.* 'thicklip' carried
244 the original sharpfin haplotype. The remaining three sharpfin species—*T. opudi*, *T.*
245 *sarasinorum*, and *T. abendanoni*—did not consistently group with either haplotype cluster:
246 roughly half of the samples had a *T. bonti*-like haplotype, while the others carried the original
247 sharpfin haplotype. None of the roundfin nor sharpfin species were monophyletic in their
248 haplotype groups.

249

250 **Incomplete lineage sorting and low variation within clades**

251 Given the relatively young age of the radiation, the incomplete sorting of standing genetic
252 variation between speciation events (incomplete lineage sorting, ILS), could lead to variation
253 in genetic relationships among species along the genome. And indeed, we found that among
254 9150 maximum likelihood trees, based on non-overlapping genomic windows of 100kb, there
255 was substantial topological variation. Consistent with this we found allele sharing across all
256 groups (Supplementary figure S6). Pairwise distance was lowest among roundfin samples
257 and highest between roundfin and *T. bonti* and sharpfin samples (fig. 4d).

258 **Signals of introgression between lake and river dwelling species**

259 To identify genetic introgression among lake species and between lake and river dwelling
260 species, we tested for signals of excess allele sharing. Elevated D-statistic, f4-ratio and f-
261 branch statistics point towards introgression between *T. bonti* and all the sharpfin species
262 (fig. 5a, supplementary figure S7, supplementary figure S8) and to a lesser extent the
263 roundfin species, and between *T. wahjui* and other sharpfin species. Comparison of
264 admixture proportions of *T. bonti* into the sharpfins *T. abendanoni*, *T. sarasinorum* and *T.*
265 *opudi* from different sampling sites revealed a spatial gradient of declining admixture with
266 increasing distance from the river-lake interface (fig. 5c-d). This is suggestive of relatively
267 recent gene flow having happened at the river-lake interface that is diffusing into lake
268 sharpfin populations through migration. One of the samples in our dataset, "*T. bonti* 1", while
269 morphologically identified as *T. bonti*, appears to be a recent *T. bonti* x sharpfin hybrid based
270 on its position in the phylogenetic trees (fig. 3a), in the PCA (fig. 4c), and its high level of
271 heterozygosity (fig. 4e).

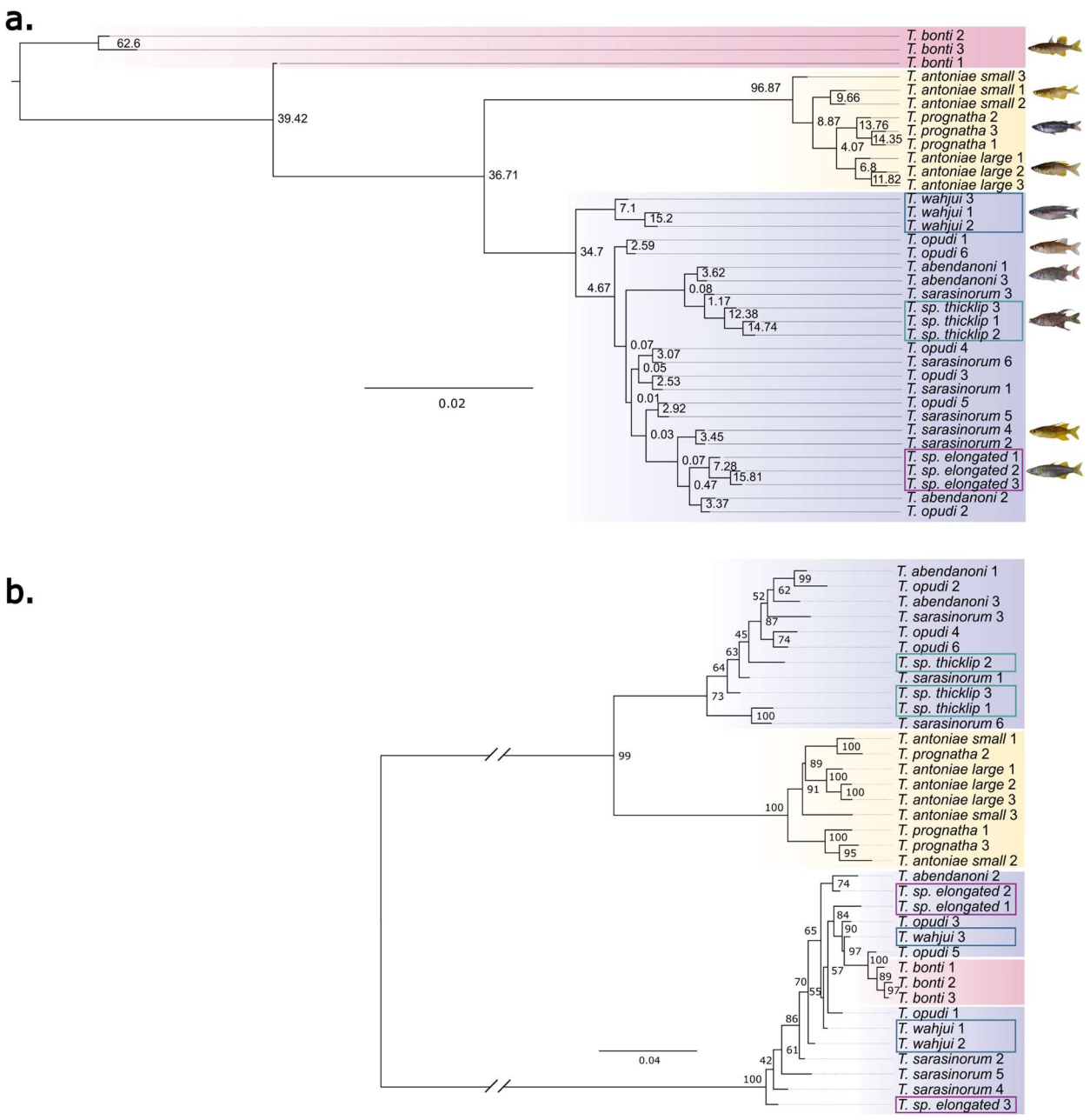


Fig. 3 - **Phylogeny based on maximum likelihood (ML) trees.** Sharpfin species are highlighted in purple, roundfins in yellow and the *T. bonti* samples are indicated in red background shading. **a.** Consensus phylogeny based on 9150 ML trees in 100 kb windows across the genome. Numbers indicate the proportion of windows supporting this split. **b.** Phylogeny based on mitochondrial DNA. Numbers indicate bootstrap support for 1000 bootstrap replicates. Coloured boxes highlight the positions of the three sharpfin species monophyletic in the genome-wide tree shown in panel a.

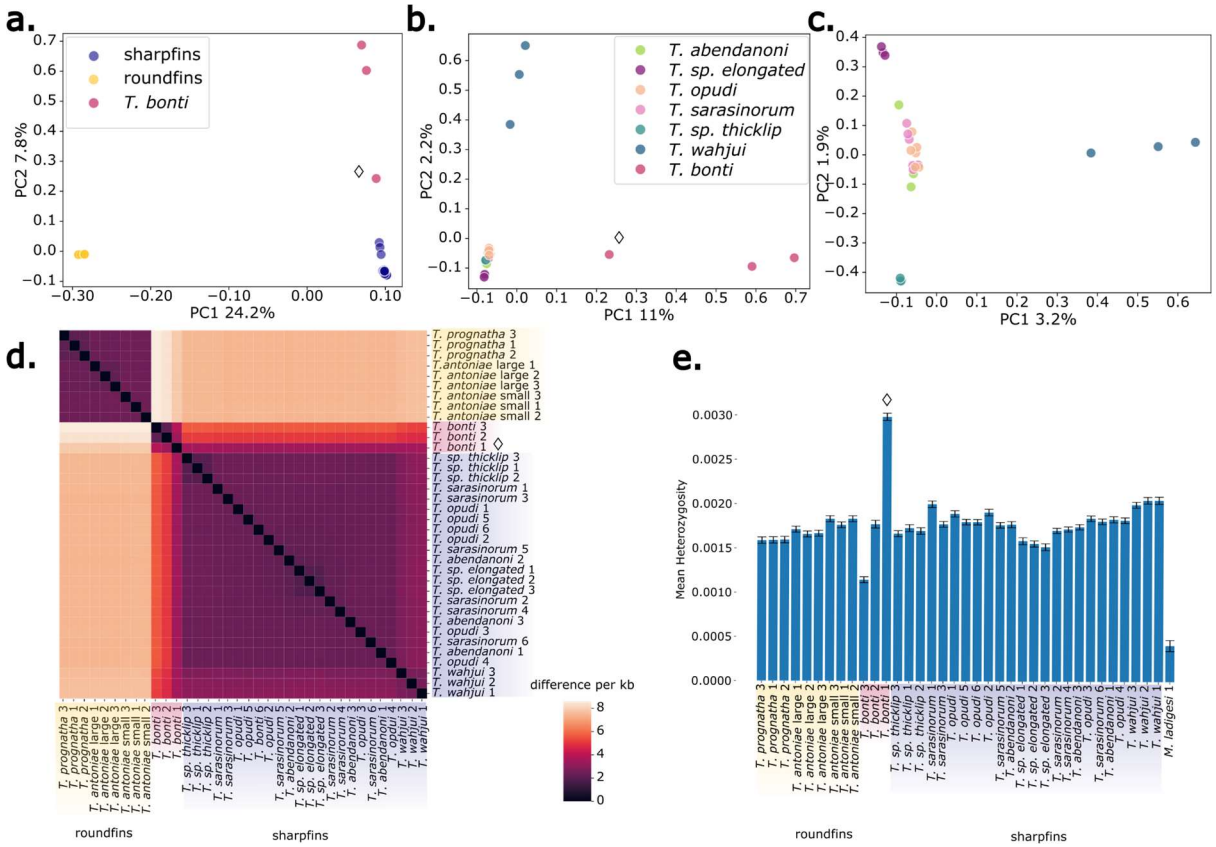


Fig. 4 - **Principal component analysis and pairwise sequence divergence. a-c**. Visualisation of the first and second genomic principal components. **a.** shows the Lake Matano sharpfins and roundfins, and the riverine *T. bonti*. **b**. Lake Matano sharpfins separated by species and *T. bonti* **c.** only Lake Matano sharpfin species. **d.** Pairwise sequence divergence between individual samples visualised in a heatmap. **e.** Mean heterozygosity in individual samples calculated over 200kb windows, error bars indicate 95% confidence interval between windows. Diamond shape (◊) indicates *T. bonti* 1, the presumed recent hybrid sample in panels a, b, d and e.

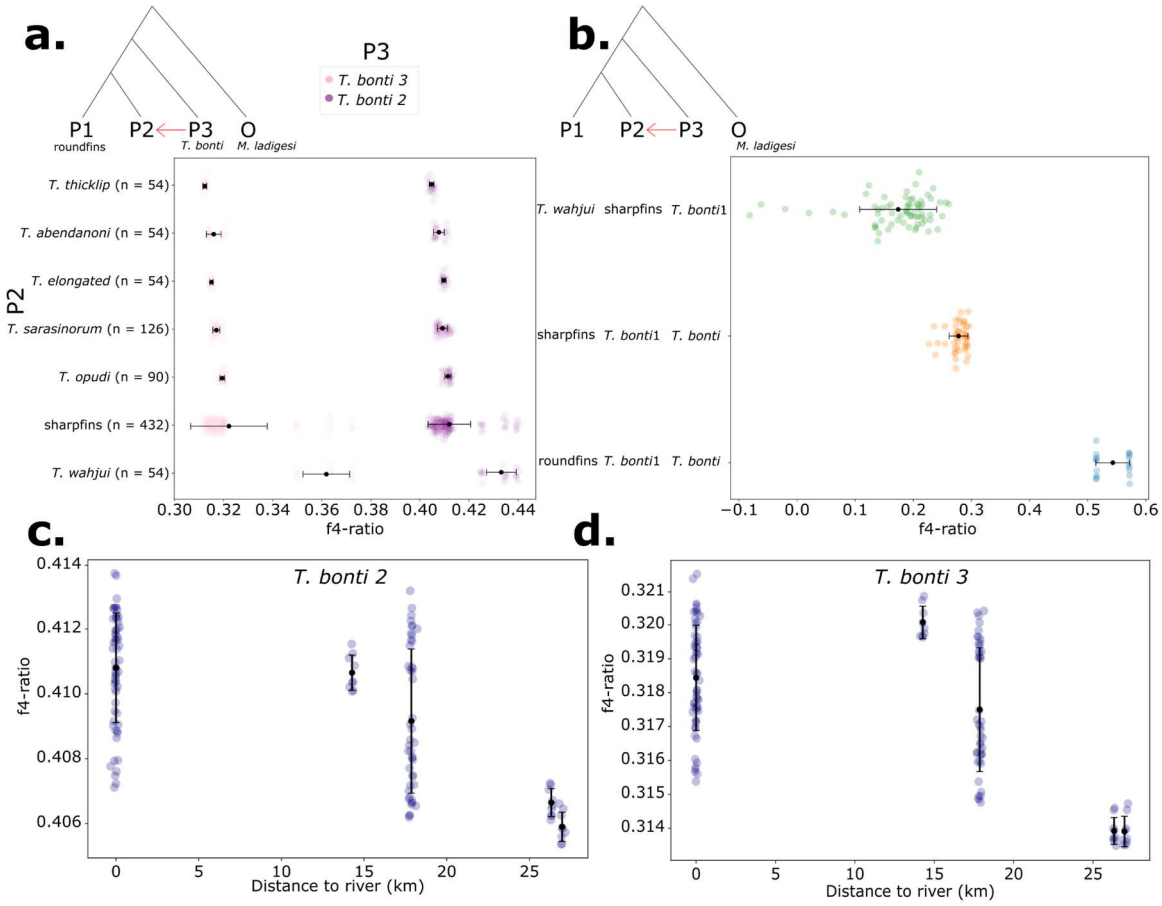


Fig. 5 - **Introgression statistics. a.** Excess allele sharing (f4-ratio) between *T. bonti* 3 (pink; P3) and *T. bonti* 2 (purple; P3) and the sharpfin species (P2), relative to the roundfins (P1). **b**. Excess allele sharing (f4-ratio) between different groups, P1, P2 and P3 are indicated on the Y-axis. *T. bonti* 1 is the presumed recent hybrid individual, *T. bonti* are all other *T. bonti* samples. Higher values indicate more introgression from P3 into P2, relative to P1 **c-d.** excess allele sharing (f4-ratio) of *T. bonti* 2 (**c**) and *T. bonti* 3 (**d**) (P3) into three sharpfin species (*T. sarasinorum*, *T. opudi*, *T. abendanoni*; P2) compared to roundfins (P1) at different distances from the river-lake interface. Higher values indicate more introgression from *T. bonti* into the three sharpfin species.

Discussion

Using whole genome sequences, we reconstructed the phylogenomic relationships and ancient and recent introgression history within the adaptive radiation of *Telmatherina*. We find that morphology-based taxonomic assignments are not consistent with genome-wide genetic relationships for three of the lacustrine species, *T. opudi*, *T. sarasinorum* and *T. abendanoni*. In addition, river-dwelling *Telmatherina* show within-population variation in their

308 genetic relationships with the lacustrine species groups. We confirm a mito-nuclear
309 discordance in the sharpfin clade, indicative of introgression between sharpfins and *T. bonti*.
310 Introgression between riverine species and the lacustrine *Telmatherina* is apparent in
311 elevated levels of excess allele sharing, which decline with increasing distance from the
312 lake-river interface.

313

314 **New species status for sharpfins and roundfins**

315 Our results confirm a phylogenomic separation between roundfins and sharpfins, both are
316 monophyletic clades within the Lake Matano sailfin silversides. Our results support the
317 sharpfins *T. wahjui*, *T.* sp. 'thicklip', and *T.* sp. 'elongated' as separate taxonomic entities, as
318 suggested by previous work based on morphology, behaviour and diet (Herder, Schwarzer,
319 et al. 2006), supporting formal taxonomic descriptions of *T.* sp. 'thicklip' and *T.* sp.
320 'elongated'. A fourth group of sharpfins consists of all samples of *T. opudi*, *T. sarasinorum*
321 and *T. abendanoni*, which together form one cluster in PCA, but are inconsistently placed in
322 phylogenetic trees. While previous research shows morphological and ecological variation
323 between these species (Herder, Schwarzer, et al. 2006; Pfaender et al. 2016), we could not
324 differentiate between them in our genome wide analysis. Possible interpretations include
325 ongoing hybridisation between these three species (Pfaender et al. 2016), hybridisation with
326 a species outside of the sharpfin clade, distorting the phylogenomic signal within the clade,
327 or an absence of separate gene pools.

328 The phylogenetic signal for the roundfin clade is clear throughout our analysis and consistent
329 with the described morphospecies, *T. prognatha*, *T. antoniae* 'small' and *T. antoniae* 'large'
330 (Herder et al. 2008; Pfaender et al. 2011). This result extends previous studies using fewer
331 genetic markers, which found no (Herder, Nolte, et al. 2006) or only low (Herder et al. 2008)
332 genetic differentiation between the three roundfin species. Based on whole genome

333 sequences, we find *T. antoniae* ‘large’ to be more closely related to *T. prognatha* than to *T.*
334 *antoniae* ‘small’. We consider *T. antoniae* ‘large’ and *T. antoniae* ‘small’ as separate species,
335 and urge for an official taxonomic revision of the roundfin clade, based on our findings
336 regarding the genomic phylogeny and previous findings regarding morphology, adaptive
337 niche differentiation, assortative mating, and diet (Herder et al. 2008; Pfaender et al. 2011).
338 The species status of *T. prognatha* has not been under debate in the literature and does also
339 not raise any concerns in our analysis. Even though the overall phylogenetic signal in the
340 roundfins is clear, we found alternative topologies where roundfin species are not
341 monophyletic in some parts of the genome, and the maximum likelihood tree based on
342 mitochondrial DNA does not clearly distinguish between the three roundfin species. This
343 inconsistency could indicate that reproductive isolation between the species is recent or be a
344 signal of incomplete isolation.

345 Similar discordance between morphological and ecological species definitions and genomic
346 signals had been reported for example for *Coenonympha* butterflies (Greenwood et al. 2025)
347 and *Eudesmia* trees (McLay et al. 2023). This discordance had been attributed to amongst
348 others incomplete lineage sorting and interspecific gene flow, complicating genomic species
349 delineation. The fact that our genomic investigation revealed discordance between described
350 species and genetic groups observed in Lake Matano *Telmatherina*, calls for an investigation
351 of species hypotheses throughout the broader silverside radiation, including the
352 *Telmatherina* and closely related *Paratherina* and *Tominanga* of the other lakes in the Malili
353 system and nearby lakes.

354

355 **Speciation and signatures of past introgression in *Telmatherina***

356 In this study, we found strong signals of past introgression of the riverine *T. bonti* into
357 lacustrine sharpfins, shown as elevated values of the f4-ratio and f-branch statistic. We

speculate that this introgression might have played a role in increasing diversity of the sharpfin clade, by providing access to a broader range of alleles for selection to act upon. Indeed, the lacustrine *Telmatherina* show an increased number of trophic traits and colour variation compared to other families of endemic Lake Matano fishes (Pfaender et al., 2010, 2016). In addition, of the two clades of *Telmatherina* in Lake Matano, the sharpfin clade, which is more species rich and shows higher variation in morphology, behaviour and trophic adaptations (Roy et al. 2004; Gray and McKinnon 2006; Pfaender et al. 2010; Cerwenka et al. 2012), is also the clade which received most introgression from *T. bonti.* Thus, our results offer a preliminary indication based on whole genome data that, as has been suggested before (Herder, Nolte, et al. 2006; Schwarzer et al. 2008), the evolutionary history of sharpfin *Telmatherina* has in part been shaped by introgression from a lineage related to the riverine *T. bonti* into an ancestor of the sharpfin clade. Similar processes have been suggested in cichlids, where gene flow between lake species and river species played an important role in exchanging genetic diversity, potentially catalysing speciation in the lake populations (Loh et al. 2013; Berner and Salzburger 2015). Ancient hybridization at the base of the Malawi cichlid radiation led to elevated divergence between splits in the radiation due to the persisting genetic variation (Svardal et al. 2020). Hybridization between two divergent lineages and multiple subsequent hybridisation events facilitated diversification in the Lake Victoria Region Superflock of Haplochromine cichlid fishes into over 700 diverse species which arose in a timespan of 150.000 years (Meier et al. 2017; Meier et al. 2023). Furthermore, it was shown recently that differentiation between two Heliconius species was driven by introgression of ecological traits (Rosser et al. 2024). Detailed assessment of divergence patterns and introgression patterns in *Telmatherina* will provide insight into the effect of gene flow on diversity in the sharpfin clade.

383 Overall, most species of sharpfins have similar levels of excess allele sharing with *T. bonti*,
384 suggesting that introgression of *T. bonti* into the sharpfin clade happened long ago, either
385 early into the sharpfin lineage or into one of the sharpfin species after which hybridisation
386 between sharpfins homogenised introgressed alleles. In contrast, the sharpfin species *T.*
387 *wahjui*, which is confined to river Petea and the lake area close to its outflow, and has
388 morphological and behavioural adaptations to the riverine environment (Gray and McKinnon
389 2006; Schwarzer et al. 2008), shows more introgression from *T. bonti* compared to other
390 sharpfin species. We speculate that the current population of *T. wahjui* is the result of past
391 intermixing between a *T. bonti*-like lineage in the outflow river Petea and the lacustrine
392 sharpfins. This way *T. wahjui* genetically clusters with sharpfins, while retaining genomic and
393 phenotypic building blocks from its *T. bonti*-like ancestral genetic backbone. This
394 interpretation is supported by our finding that the recent sharpfin-bonti hybrid (*T. bonti* 1) in
395 our dataset shares fewer alleles with *T. wahjui* compared to the other sharpfins, in contrast
396 to the other *T. bonti* samples. Alternatively, *T. wahjui* may have received *T. bonti*
397 introgression in the same ancient introgression event as the rest of the sharpfin species and
398 may have retained more of the *T. bonti* alleles, as these would be more adaptive in their
399 riverine habitat. A thorough timing of hybridisation events and insight into the selection
400 processes in *T. wahjui* will shed light on these questions.

401

402 **Recent and ongoing gene flow between Matano sharpfins and the riverine group**

403 Besides signals of past gene flow, we also detected recent gene flow between the stream-
404 dwelling *T. bonti* and Matano species. In the maximum likelihood (ML) consensus trees,
405 assuming a GTR model, *T. bonti* is positioned as an outgroup to both the sharpfins and
406 roundfins. On the other hand, the NJ tree suggests closer genetic similarity between *T. bonti*
407 and sharpfins relative to roundfins, possibly due to reduced genetic distance resulting from
408 recent gene flow. While the NJ method reflects genetic distance, not accounting for

speciation models, ML is more robust for complex evolutionary histories, including variable substitution rates and heterogeneity across the genome. Given these strengths, we interpret the ML tree as more likely reflecting true phylogeny. In addition to inconsistency between NJ and ML phylogenies, we also find that mitochondrial and nuclear patterns are discordant. While the nuclear trees show monophyletic groups for sharpfins and roundfins, in the mitochondrial tree, sharpfins are split between a group that is a sister group to roundfins, and a more basal group which also contains the riverine *T. bonti*. This latter group contains besides *T. bonti,* all samples of *T. wahjui* and *T.* sp. 'elongated', and about half of the samples of *T. sarasinorum, T. opudi* and *T. abendanoni*. These results suggest gene flow from the riverine *T. bonti* into most species of the sharpfin clade. Using whole genome data, our study confirms the patterns identified in previous studies based on AFLP markers (Herder, Nolte, et al. 2006; Stelbrink et al. 2014).

One of the three samples in our dataset which were initially identified as *T. bonti* is a hybrid between *T. bonti* and a Lake Matano sharpfin species. This is evident from its position in the PCA intermediate between the other *T. bonti* samples and Lake Matano sharpfins. Heterozygosity in this individual is twice as high compared to the other samples, suggesting this to be a recent hybrid. The specimen in question was captured at the lake-river interface, where river-dwelling and lake-dwelling species are in direct contact. Since in our limited sample-set we already find a recent hybrid, this suggests that hybridisation might be a locally common process. The other two samples of *T. bonti* in our dataset have less sharpfin-like DNA, which suggests that these are backcrosses between *T. bonti* x sharpfin hybrids and *T. bonti*. This finding suggests the interesting possibility of ongoing introgression of sharpfin alleles into the riverine population, the opposite direction compared to ancient introgression, which introduced alleles from *T. bonti* into Lake Matano sharpfins. Sampling upstream Lawa river will reveal if there is a spatial gradient of proportion of introgressed sharpfin alleles in *T. bonti*.

435 We detected clear signals of decreased excess allele sharing (lower f4-ratios) with
436 increasing distance to the river-lake interface between *T. bonti* and three sharpfin species
437 which occur over the entire lake, *T. sarasinorum*, *T. opudi* and *T. abendanoni*. This pattern is
438 consistent with a hybrid zone at the interface between the inlet river Lawa and the lake,
439 possibly facilitated by natural selection favouring hybrid individuals in this ecologically
440 intermediate zone. This spatial decline in excess allele sharing is a strong indication of
441 ongoing or recent gene flow, as signals of older gene flow would over time have
442 homogenised over the extent of the lake, erasing spatial signals.

443

444 **Signals of incomplete speciation**

445 Our results show weak genomic clustering and weak lineage sorting in three sharpfin
446 species; *T. sarasinorum*, *T. opudi* and *T. abendanoni*. Incomplete reproductive isolation
447 between these species (Herder et al, 2006b, 2008, Schwarzer et al. 2008, Stelbrink et al.
448 2014), and the occurrence of sharpfin *Telmatherina* with morphological traits intermediate
449 between species, suggest that the sharpfin system is in a dynamic stage of the species flock
450 formation process, evolving under disruptive ecological selection. Previous studies found
451 that variation in morphology between the sharpfin species was linked to variation in a small
452 proportion (13%) of nuclear markers, and intermediate, less fit individuals were common,
453 indicative of a young species flock (Pfaender et al. 2016). This signature of recent species
454 flock formation is surprising since diversification within sharpfins has been estimated to have
455 started relatively long ago, around 1 mya (Stelbrink et al. 2014). Why does sharpfin
456 speciation remain incomplete? An interesting contributing factor could be that the influx of
457 similar *T. bonti* alleles into the sharpfin species through the ongoing hybridisation with
458 riverine *T. bonti* continues to homogenise the sharpfin genomes. This continuous influx of *T.*
459 *bonti* alleles, combined with species-specific adaptation pressures, could create a balance
460 between homogenisation of the genome and ecological selection, so that speciation remains

incomplete. To understand the processes behind the incomplete separation of these sharpfin species, future studies will need to gain insight into the interplay between genetic influx of *T. bonti* into the sharpfin species, the ongoing mixing between the sharpfin species, and the strength and type of selection that introgressed alleles undergo.

Conclusion

This study provides data supporting both ancient and ongoing hybridisation in the *Telmatherina* of Lake Matano, which may have accelerated adaptation to a wide variety of ecological niches. This combination of ancient and ongoing hybridisation in a comprehensive, still radiating flock makes the *Telmatherina* of Lake Matano a particularly valuable study system to unravel the fundamental mechanisms driving rapid speciation under genomic exchange. The reference genome and phylogenomic framework elaborated in this study provide the foundation for more in depth studies of this charismatic radiation.

Methods

Sampling and sequencing

We selected 36 samples of the genus *Telmatherina* from the ichthyology collections of the Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig in Bonn, Germany (Supplementary table S2). Samples were collected in Indonesia between 2002 and 2010, sampling was done by gillnet. Sex was determined visually upon capture, the fish were photographed and fish were euthanized by administering an overdose of MS222. A finclip was immediately stored in 97% ethanol; fish specimens were stored in formalin for several days before being transferred to ethanol. Two specimens of *Marosatherina ladigesi* were sampled from an aquarium population to serve as an outgroup in this study. These fish were

485 euthanized by administering an overdose of MS222, photographed and a part of the right
486 pectoral fin was stored separately in 97% ethanol. DNA for short read sequencing was
487 extracted from fin clips using the Zymo tissue kit according to manufacturer's guidelines.
488 Extracted DNA was sent to BGI, Hong Kong and submitted for whole genome resequencing
489 on DNBSEQ for 150bp paired end short read sequencing.

490 For the reference genome construction and annotation, one specimen of *T. bonti* from an
491 aquarium stock at ZFMK was immediately dissected after euthanization with an overdose of
492 MS222. Fin and muscle tissue samples were placed in pre-chilled tubes, immediately
493 submerged in liquid nitrogen for snap freezing and kept at -80° C until high-molecular weight
494 DNA extraction for long read sequencing. The workflow for high-molecular weight (HMW)
495 DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory
496 includes a sequence of core procedures: sample preparation; sample homogenisation, DNA
497 extraction, fragmentation, and clean-up. In sample preparation, the fTelBon1 sample was
498 weighed and dissected on dry ice (Jay et al. 2023). Tissue from muscle was homogenised
499 using a PowerMasher II tissue disruptor (Denton et al. 2023). HMW DNA was extracted
500 using the Automated MagAttract v1 protocol (Sheerin et al. 2023). DNA was sheared into an
501 average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic et al. 2023).
502 Sheared DNA was purified by solid-phase reversible immobilisation (Strickland et al. 2023),
503 using AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA.
504 The concentration of the sheared and purified DNA was assessed using a Nanodrop
505 spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit.
506 Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

507 RNA was extracted from muscle tissue of fTelBon1 in the Tree of Life Laboratory at the WSI
508 using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral et al. 2023).
509 The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit

510 Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the
511 RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

512 Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed
513 according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed
514 using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by
515 the Scientific Operations core at the WSI on Pacific Biosciences Sequel IIe (HiFi) and
516 Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from muscle
517 tissue of fTelBon1 using the Arima-HiC v2 kit. The Hi-C sequencing was performed using
518 paired-end sequencing with a read length of 150 bp on the Illumina NovaSeq 6000
519 instrument.

520

521 **Reference genome construction**

522 The reference genome for *T. bonti* (GCA_933228915.1;
523 https://www.ncbi.nlm.nih.gov/assembly/GCA_933228915.1) was sequenced from a single
524 male *T. bonti* specimen bred at the LIB Bonn, Germany. A total of 36-fold coverage in
525 PacBio long reads. The HiFi reads were first assembled using Hifiasm version 0.16.1-r375
526 (Cheng et al. 2021) with the --primary option. Haplotypic duplications were identified and
527 removed using purge_dups version 1.2.3 (Guan et al. 2020). The Hi-C reads were mapped
528 to the primary contigs using bwa-mem2 (Vasimuddin et al. 2019). The contigs were further
529 scaffolded using the provided Hi-C data (Rao et al. 2014) in YaHS version yahs-1.1.91eebc2
530 (Zhou et al. 2023) using the --break option. The scaffolded assemblies were evaluated using
531 Gfastats (Formenti et al. 2022), BUSCO (Manni et al. 2021) and MERQURY.FK (Rhie et al.
532 2020).

533 The mitochondrial genome was assembled using MitoHiFi version 2 (Uliano-Silva et al.
534 2023), which runs MitoFinder (Allio et al. 2020) and uses these annotations to select the final
535 mitochondrial contig and to ensure the general quality of the sequence.

536 Manual curation was primarily conducted using PretextView 0.2.5 (Harry 2022), with
537 additional insights provided by JBrowse2 (Diesh et al. 2023) and HiGlass (Kerpedjiev et al.
538 2018). Scaffolds were visually inspected and corrected as described by Howe et al. (Howe et
539 al. 2021). Manual assembly curation corrected 207 missing joins or mis-joins and one
540 haplotypic duplication, which reduced the scaffold number by 54.26 and increased the
541 scaffold N50 by 11.67%. The process is documented at https://gitlab.com/wtsi-grit/rapid-
542 curation.

543 In total 24 chromosome-scale scaffolds were confirmed by the Hi-C data, which have been
544 named in order of size (large to small). Besides the 24 chromosomes and the mitochondrial
545 genome, the genome assembly includes 115 unplaced scaffolds.

546

547 **Short read quality control and mapping**

548 Whole genome sequencing reads were filtered by BGI using the SOAPnuke software (Chen
549 et al. 2018). Sequencing reads which matched > 25% adapter sequence, had a read length
550 of < 150 bp, > 0.1% N content or a Phred score < 33 were filtered out. Sequencing yielded
551 ~80.1 million clean reads (after filtering) per sample. Sequences were quality checked using
552 FastQC v0.11.9 (Andrews 2010) and MultiQC v1.12 (Ewels et al. 2016). Reads were aligned
553 against the *T. bonti* (NCBI:txid446457) reference genome fTelBon1.1 with BWA-MEM (BWA
554 version 0.7.17, (Li 2013)). The resulting SAM files were compressed to cram format and
555 quality checked using samtools flagstat. After filtering and mapping each sample to the *T.*
556 *bonti* reference genome, the resulting effective sequencing coverage varied between 13.8-

557 22.9 fold, except for the outgroup samples (*M. ladigesi*) which yielded on average 10.5 fold
558 coverage, related to their larger distance to the reference species.

559 Only single nucleotide polymorphisms (SNPs) were retained for genetic investigation.
560 Variant sites were called jointly for all samples with the mpileup function in BCFtools v1.14
561 (Danecek et al. 2021). Total depth and mean coverage were calculated per site. Any sites
562 with an overall mapping quality Phred-scaled score of less than 50 (4.9% of SNPs) or with
563 more than 10% of mapped reads with a mapping quality zero (4.0% of SNPs) were masked.
564 Additionally, filters were applied to sites with excess heterozygosity (InbreedingCoeff < 0.2;
565 0.03% of SNPs), before filtering and sites with a Phred-scaled p-value of binomial test for
566 allele balance violation of > 40 (0.06% of SNPs). Sites with > 40% missing genotypes (2.0%
567 of SNPs) as well as sites where the sum of overall depth for all samples was unusually high
568 (> 3 units of standard deviation above individual mean coverage; 3.6% of SNPs) were
569 filtered out. In total, 5.2% of SNPs were removed from the dataset (Supplementary figure
570 S9), note that most sites were removed due to multiple filters. After filtering, 43.862.995
571 biallelic SNPs across 24 chromosomes remained. The mitogenome of the *T. bonti* reference
572 was 16.560 bp long. Filters were applied to mitochondrial sites with a mapping quality Phred-
573 scaled score under 10 (19% of SNPs), sites with more than 50% of mapped reads with a
574 mapping quality of zero (11% of SNPs), sites with a Phred-scaled p-value of binomial test for
575 allele balance violation of > 40 (1.2% of SNPs), and sites with > 40% missing genotypes
576 after application of other filters (23% of SNPs)(Supplementary figure S9). After filtering, 1836
577 SNPs were retained in the aligned mitogenome.

578 **Phylogeny**

579 *Nuclear genome*

580 To infer evolutionary relationships between the different *Telmatherina* species we
581 constructed genomic phylogenies based on whole genome sequences. The SNP dataset
582 was split into genomic windows of 100kb based on SNP coordinates, using a custom python
583 script, resulting in 9831 windows. Of these, 681 windows containing fewer than 10 SNPs
584 were removed since these did not contain sufficient information to construct a reliable tree,
585 resulting in a final dataset of 9150 windows, containing 10 - 7905 SNPs (median = 5793.5).

586 In these remaining genomic windows of 100 kb over the nuclear genome, local neighbour
587 joining (NJ) trees were calculated based on distance matrices and a consensus tree was
588 calculated with Astral-III (version 5.7.1, (Zhang et al. 2018)). For the same dataset,
589 Maximum likelihood (ML) trees were inferred using the General time reversible (GTR) model
590 in IQ-TREE v2.2.0 (Minh et al. 2020), which accounts for unequal substitution rates between
591 nucleotide pairs and unequal base frequencies. Bootstrap support was calculated for 1000
592 bootstrap replicates with ultrafast bootstrap approximation (UFBoot2; (Hoang et al. 2018)).
593 The local trees were summarised as a majority consensus tree with SumTrees v4.5.2,
594 distributed with the DendroPy v4.5.2 Python library (Sukumaran and Holder 2010), with
595 minimum clade frequency set to zero. The consensus trees were visualised using Figtree
596 v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

597 For all trees, a specimen of *Marosatherina ladigesi* was used as an outgroup. The split
598 between the ancestors of *M. ladigesi* and *Telmatherina* has been estimated between 12.9
599 and 42.9 Mya (Stelbrink et al. 2014).

600 *Mitochondrial tree*

601 We constructed a mitochondrial phylogeny from the SNP set (1836 nucleotide sites) of the
602 mitogenomes with IQ-TREE v2.2.0 run with the GTR-model of nucleotide evolution and
603 default parameters, using a *M. ladigesi* sample as an outgroup. The consensus tree was
604 constructed from 1000 bootstrap trees with the UltraFast bootstrap (UFBoot2) setting in IQ-
605 Tree (Hoang et al. 2018).

606

607 **Phylogenomic discordance**

608 To gain insight into phylogenomic discordance in the *Telmatherina* adaptive radiation, we
609 assessed the set of local maximum likelihood trees. We counted the number of unique
610 topologies as well as the relative frequencies of the most frequent topologies with Newick
611 utilities v1.6 (Junier and Zdobnov 2010). In total, there were 9150 unique topologies,
612 meaning that every window resulted in a unique topology.

613

614 **Genomic diversity**

615 To visualise similarity between individuals, principal component analysis (PCA) was
616 performed on the SNP dataset with PLINK v2.0 (Purcell et al. 2007; Chang et al. 2015). For
617 PCA analysis we only retained SNPs with a minor allele frequency (maf) of at least 0.05. PC
618 analysis was done on the entire dataset, on the sharpfins and roundfins respectively alone,
619 and on the sharpfins and riverine species (*T. bonti*) jointly. Figures were plotted with seaborn
620 v 0.12.2 (Waskom 2021).

621 We calculated Hamming distance in the nuclear genome with PLINK v2.0 (Purcell et al.
622 2007; Chang et al. 2015), and divided by size of the filtered genome (accessible genome) to

623 assess pairwise distance between all samples. We calculated heterozygosity of the entire
624 genomes in windows of 200kb with ANGSD v0.921 (Korneliussen et al. 2014), based on the
625 folded site frequency spectrum, using the *T. bonti* reference as the ancestral and filtering for
626 a minimum read quality of 20, minimum mapping quality of 30 and adjusting mapping quality
627 for excessive mismatches.

628 To assess whether the sharpfin, roundfin and riverine groups were genetically well
629 separated or still shared substantial variation, we measured allele sharing across all filtered
630 SNPs using a custom Python script with *cyvcf2* v0.31.1 (Pedersen and Quinlan 2017). For
631 each SNP, we recorded whether both reference and alternative alleles were present within
632 the roundfin, sharpfin, and *T. bonti* groups, as well as in every pairwise and three-way
633 combination of these groups. We identified SNPs that were unique to each group and those
634 shared between groups. For this analysis we did not include the presumed hybrid specimen
635 *T. bonti 1*. Missing alleles were not considered when classifying genotypes as heterozygous.
636 The results were visualised in a Venn diagram using *matplotlib-venn* v1.1.2. As the
637 reference genome used for mapping and SNP calling belongs to *T. bonti*, this analysis is
638 biased toward detecting fewer non-reference alleles in the *T. bonti* group, potentially
639 underestimating its shared variation with the other groups. In addition, the sample sets were
640 unbalanced across groups, which may have led to an underestimation of allele sharing
641 between *T. bonti* (represented by fewer samples) and the other groups.

642

643 **Introgression**

644 We calculated D-statistics (ABBA-BABA test) to assess gene flow between populations, as it
645 allows us to distinguish between incomplete lineage sorting and introgressive hybridisation.
646 This test compares allele sharing patterns between four groups: two sister groups (P1 and
647 P2), a closely related third group (P3) and an outgroup (O). In the absence of gene flow, the

648 two sister groups P1 and P2 would be expected to share the same amount of derived alleles
649 (different from the allele in the outgroup O) with the third, related species P3. If however
650 there is gene flow from P3 into P2, and not in P1, there will be an excess of shared derived
651 alleles between P3 and P2, compared to P1, counted as a higher ratio of ABBA/BABA
652 patterns. This pattern of excess allele sharing was tested for all possible (unordered) trios of
653 individual samples in our dataset, by computing genome-wide D statistics and f4-ratio with
654 Dsuite v0.4 (Malinsky et al. 2021). For each test, an *M. ladigesi* sample was used as an
655 outgroup. F4-ratios of excess allele sharing between *T. bonti* and the sharpfin species
656 relative to the roundfins is visualised, as well as excess allele sharing between *T. bonti* and
657 the supposed hybrid specimen *T. bonti* 1.

658 If hybridisation events are more prevalent closer to the river-lake interface, this could lead to
659 a spatial signal in the proportion of allele sharing of lake sharpfins with riverine *T. bonti.* We
660 compared f4-ratios of allele sharing between *T. bonti* and sharpfins from several sampling
661 sites at different distances from the lake-river interface, compared to a non-introgressed
662 outgroup, the roundfins. Here we orient the roundfins as P1, the sharpfins as P2, and the
663 riverine *T. bonti* as P3. This way, a higher f4-ratio indicates more allele sharing between *T.*
664 *bonti* (P3) and the sharpfins (P2). We conducted this test for the three most common
665 sharpfin species *T. sarasinorum*, *T. opudi* and *T. abendanoni*.

666 To correct for false signals caused by correlations of admixture ratios between taxa which
667 are in the same clade as taxa involved in introgression, we calculated the f-branch statistic.
668 This statistic takes into account the timing of introgression by investigating the internal
669 branches of the phylogeny (Malinsky et al. 2018).

670 Data availability

671 The genome assembly fTelBon1.1 is available on NCBI (https://www.ncbi.nlm.nih.gov/)
672 under bioproject PRJEB51590
673 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_933228915.1/)

674 Raw reads of the resequencing data are available at the European nucleotide archive (ENA;
675 https://www.ebi.ac.uk/) under study PRJEB45386

676

677 Abbreviations

678 GTR - General time reversible

679 Maf - minor allele frequency

680 ML - Maximum-likelihood

681 Mya - million years ago

682 NJ - Neighbour-joining

683

684 Acknowledgements

685 No samples have been collected for this study. The specimens used in this study were
686 available from museum collections of the Museum Zoologicum Bogoriense (MZB) in Bogor,
687 Indonesia and the Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn,
688 Germany. Original sampling was conducted between 2002 and 2010 with permits from the
689 Indonesian government (Surat Izin Penelitian) 5602/SU/KS/2002 (LIPI) and
690 0093/SIP/FRP/SM/V/2010 (RISTEK). We thank the Heads of Research Center for Biology,

Indonesian Institute of Sciences (LIPI; presently Research Center for Biosystematics and Evolution, National Research and Innovation Agency or Badan Riset dan Inovasi Nasional [BRIN] and the Kementerian Negara Riset dan Teknologi [RISTEK] presently also BRIN) for permissions to conduct research in Indonesia. The late R. K. Hadiaty enabled joint fieldwork in Sulawesi, funded by the Deutsche Forschungsgemeinschaft (DFG SCHL 567/2, HE 5707/2-1, HE 5707/3-1), providing voucher samples re-analyzed in the present project. Computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government. Access to VSC infrastructure was facilitated by the CalcUA team at Universiteit Antwerpen. ELRDK is supported as postdoctoral fellow fundamental research by the Flemish Research Foundation (FWO postdoctoral scholarship 1275922N). Research and collaboration between Indonesia and Europe was supported by Research Foundation - Flanders (FWO) (project G0A9B24N to HS and travel grant V408023N to ELRDK), VLIR-UOS (SI project ID2022SIN349A102 to HS and GMSR project 54084 to ELRDK) and Universiteit Antwerpen (BOF Seal of approval and KPBOF project 48620 to ELRDK).

Author contributions

ELRDK, FH, AB and HS conceived the study. ELRDK, CCJ, VB and HS conducted analysis on resequencing data. SK, AD and GO conducted molecular laboratory work. The Wellcome Sanger Institute Tree of life programme organised and conducted fTelBon1.1 reference genome assembly and annotation. AT led manual curation of the fTelBon1.1 reference genome. DW, DFM and FH organised and conducted data collection. ELRDK and HS wrote the manuscript with input from all co-authors.

715 References

716 Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K,
717 García Girón C, Hourlier T, et al. 2016. The Ensembl gene annotation system.
718 *Database (Oxford)* 2016:baw093.

719 Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020.
720 MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target
721 enrichment phylogenomics. *Mol. Ecol. Resour.* 20:892–905.

722 do Amaral RJV, Denton A, Yatsenko H, Jay J, Howard C, Bates A. 2023. Sanger Tree of Life
723 RNA Extraction: Automated MagMax™ mirVana. Available from:
724 https://www.protocols.io/view/sanger-tree-of-life-rna-extraction-automated-magma-
725 cxxnxpme.html

726 Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

727 Arnold M. 2006. Evolution through genetic exchange. Available from:
728 https://academic.oup.com/book/3287

729 Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by
730 adaptive radiations. *Trends Genet.* 31:491–499.

731 Cerwenka AF, Wedekind JD, Hadiaty RK, Schliewen UK, Herder F. 2012. Alternative egg-
732 feeding tactics in Telmatherina sarasinorum, a trophic specialist of Lake Matano's
733 evolving sailfin silversides fish radiation. *Hydrobiologia* 693:131–139.

734 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation
735 PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.

736 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo
737 assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18:170–175.

738 Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, et al. 2018.
739 SOAPnuke: a MapReduce acceleration-supported software for integrated quality control
740 and preprocessing of high-throughput sequencing data. *Gigascience* 7:1–6.

741 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
742 McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools.
743 *Gigascience* 10. Available from: http://dx.doi.org/10.1093/gigascience/giab008

744 Denton A, Oatley G, Cornwell C, Quail M, Howard C. 2023. Sanger Tree of Life Sample
745 Homogenisation: PowerMash. Available from: https://www.protocols.io/view/sanger-
746 tree-of-life-sample-homogenisation-powermas-cyf8xtrw.html

747 Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S,
748 Zhang J, Bridge C, et al. 2023. JBrowse 2: a modular genome browser with views of
749 synteny and structural variation. *Genome Biol.* 24:74.

750 Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for
751 multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.

752 Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, Fedrigo O, Jarvis ED.
753 2022. Gfastats: conversion, evaluation and manipulation of genome sequences using

754 assembly graphs. *Bioinformatics* 38:4214–4216.

755 Gray SM, McKinnon JS. 2006. A comparative description of mating behaviour in the
756 endemic telmatherinid fishes of Sulawesi's malili lakes. *Environ. Biol. Fishes* 75:471–
757 482.

758 Greenwood MP, Capblancq T, Wahlberg N, Després L. 2025. Whole genome data confirm
759 pervasive gene discordance in the evolutionary history of *Coenonympha* (Nymphalidae)
760 butterflies. *Mol. Phylogenet. Evol.* 202:108222.

761 Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing
762 haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898.

763 Hadiaty RK, Wirjoatmodjo S. 2002. Studi Pendahuluan biodiversitas Dan distribusiikan Di
764 Danau Matano, Sulawesi Selatan (preliminary study: Biodiversity and distribution of
765 fishes in lake Matano, south Sulawesi). *J. Iktiologi Indones.* 2:23–29.

766 Harry E. 2022. PretextView (Paired REad TEXTure Viewer): A desktop application for
767 viewing pretext contact maps.

768 Herder F, Nolte AW, Pfaender J, Schwarzer J, Hadiaty RK, Schliewen UK. 2006. Adaptive
769 radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in
770 the Malili Lakes of Sulawesi. *Proc. Biol. Sci.* 273:2209–2217.

771 Herder F, Pfaender J, Schliewen UK. 2008. Adaptive sympatric speciation of polychromatic
772 "roundfin" sailfin silverside fish in Lake Matano (Sulawesi). *Evolution* 62:2178–2195.

773 Herder F, Schliewen UK. 2010. Beyond Sympatric Speciation: Radiation of Sailfin Silverside
774 Fishes in the Malili Lakes (Sulawesi). *Evolution in Action*:465–483.

775 Herder F, Schwarzer J, Pfaender J, Hadiaty RK, Schliewen UK. 2006. Preliminary checklist
776 of sailfin silversides (Teleostei: Telmatherinidae) in the Malili Lakes of Sulawesi
777 (Indonesia), with a synopsis of systematics and threats. *Verhandlungen der*
778 *Gesellschaft für Ichthyologie* 5:139–163.

779 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
780 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522.

781 Howe K, Chow W, Collins J, Pelan S, Pointon D-L, Sims Y, Torrance J, Tracey A, Wood J.
782 2021. Significantly improving the quality of genome assemblies through curation.
783 *Gigascience* 10:giaa153.

784 Jay J, Yatsenko H, Narváez-Gómez JP, Mbye H, Morra M, Strickland M, Cornwell C,
785 Howard C. 2023. Sanger Tree of Life Sample Preparation: Triage and Dissection.
786 Available from: https://www.protocols.io/view/sanger-tree-of-life-sample-preparation-
787 triage-and-cztex6je.html

788 Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree
789 processing in the UNIX shell. *Bioinformatics* 26:1669–1670.

790 Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM,
791 Ouellette SB, Azhir A, Kumar N, et al. 2018. HiGlass: web-based visual exploration and
792 analysis of genome interaction maps. *Genome Biol.* 19:125.

793 Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation

794 Sequencing Data. *BMC Bioinformatics* 15:356.

795 Kottelat M. 1991. Sailfin silversides (Pisces: Telmatherinidae) of Lake Matano, Sulawesi,
796 Indonesia, with descriptions of six new species. *Ichthyol Explor Freshw* 1:321–344.

797 Kozak KM, Joron M, McMillan WO, Jiggins CD. 2021. Rampant genome-wide admixture
798 across the Heliconius radiation. *Genome Biol. Evol.* 13:evab099.

799 Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A,
800 Promerová M, Rubin C-J, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches
801 and their beaks revealed by genome sequencing. *Nature* 518:371–375.

802 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
803 MEM. *arXiv [q-bio.GN]*. Available from: http://arxiv.org/abs/1303.3997

804 Loh Y-HE, Bezault E, Muenzel FM, Roberts RB, Swofford R, Barluenga M, Kidd CE, Howe
805 AE, Di Palma F, Lindblad-Toh K, et al. 2013. Origins of shared genetic variation in
806 African cichlids. *Mol. Biol. Evol.* 30:906–917.

807 Malinsky M, Matschiner M, Svardal H. 2021. Dsuite - Fast D-statistics and related admixture
808 evidence from VCF files. *Mol. Ecol. Resour.* 21:584–595.

809 Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018.
810 Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected
811 by gene flow. *Nat Ecol Evol* 2:1940–1955.

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

815 Marques DA, Meier JI, Seehausen O. 2019. A Combinatorial View on Speciation and
816 Adaptive Radiation. *Trends Ecol. Evol.* 34:531–544.

817 McLay TGB, Fowler RM, Fahey PS, Murphy DJ, Udovicic F, Cantrill DJ, Bayly MJ. 2023.
818 Phylogenomics reveals extreme gene tree discordance in a lineage of dominant trees:
819 hybridization, introgression, and incomplete lineage sorting blur deep evolutionary
820 relationships despite clear species groupings in Eucalyptus subgenus Eudesmia. *Mol.*
821 *Phylogenet. Evol.* 187:107869.

822 Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient
823 hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363.

824 Meier JI, McGee MD, Marques DA, Mwaiko S, Kishe M, Wandera S, Neumann D, Mrosso H,
825 Chapman LJ, Chapman CA, et al. 2023. Cycles of fusion and fission enabled rapid
826 parallel adaptive radiations in African cichlids. *Science* 381:eade2833.

827 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear
828 R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in
829 the Genomic Era. *Mol. Biol. Evol.* 37:1530–1534.

830 Nolte AW, Tautz D. 2010. Understanding the onset of hybrid speciation. *Trends Genet.*
831 26:54–58.

832 Peñalba JV, Runemark A, Meier JI, Singh P, Wogan GOU, Sánchez-Guillén R, Mallet J,
833 Rometsch SJ, Menon M, Seehausen O, et al. 2024. The Role of Hybridization in

Species Formation and Persistence. *Cold Spring Harb. Perspect. Biol.*

Pfaender J, Hadiaty RK, Schliewen UK, Herder F. 2016. Rugged adaptive landscapes shape a complex, sympatric radiation. *Proc. Biol. Sci.* 283.

Pfaender J, Miesen FW, Hadiaty RK, Herder F. 2011. Adaptive speciation and sexual dimorphism contribute to diversity in form and function in the adaptive radiation of Lake Matano's sympatric roundfin sailfin silversides. *J. Evol. Biol.* 24:2329–2345.

Pfaender J, Schliewen UK, Herder F. 2010. Phenotypic traits meet patterns of resource use in the radiation of "sharpfin" sailfin silverside fish in Lake Matano. *Evol. Ecol.* 24:957–974.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–1680.

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21:245.

von Rintelen T, von Rintelen K, Glaubrecht M, Schubart CD, Herder F. 2012. Aquatic biodiversity hotspots in Wallacea:: the species flocks in the ancient lakes of Sulawesi, Indonesia. In: Gower D, Johnson K, Richardson J, Rosen B, Ruber L, Williams S, editors. Biotic Evolution and Environmental Change in Southeast Asia. Cambridge: Cambridge University Press. p. 290–315.

Rosser N, Seixas F, Queste LM, Cama B, Mori-Pezo R, Kryvokhyzha D, Nelson M, Waite-Hudson R, Goringe M, Costa M, et al. 2024. Hybrid speciation driven by multilocus introgression of ecological traits. *Nature* 628:811–817.

Roy D, Docker MF, Hehanussa P, Heath DD, Haffner GD. 2004. Genetic and morphological data supporting the hypothesis of adaptive radiation in the endemic fish of Lake Matano. *J. Evol. Biol.* 17:1268–1276.

Schwarzer J, Herder F, Misof B, Hadiaty RK, Schliewen UK. 2008. Gene flow at the margin of Lake Matano's adaptive sailfin silverside radiation: Telmatherinidae of River Petea in Sulawesi. *Hydrobiologia* 615:201–213.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19:198–207.

Seehausen O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* 26:279–281.

Seehausen O, Takimoto G, Roy D, Jokela J. 2008. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* 17:30–44.

Sheerin E, Sampaio F, Oatley G, Todorovic M, Strickland M, do Amaral RJV, Howard C. 2023. Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.1. Available from: https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-extraction-automated-m-czh2x38e.html

874 Stelbrink B, Stöger I, Hadiaty RK, Schliewen UK, Herder F. 2014. Age estimates for an
875 adaptive lake fish radiation, its mitochondrial introgression, and an unexpected sister
876 group: Sailfin silversides of the Malili Lakes system in Sulawesi. *BMC Evol. Biol.* 14:94.

877 Strickland M, Cornwell C, Howard C. 2023. Sanger Tree of Life Fragmented DNA clean up:
878 Manual SPRI. Available from: https://www.protocols.io/view/sanger-tree-of-life-
879 fragmented-dna-clean-up-manual-czhkx34w.html

880 Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing.
881 *Bioinformatics* 26:1569–1571.

882 Svardal H, Quah FX, Malinsky M, Ngatunga BP, Miska EA, Salzburger W, Genner MJ,
883 Turner GF, Durbin R. 2020. Ancestral Hybridization Facilitated Species Diversification in
884 the Lake Malawi Cichlid Fish Adaptive Radiation. *Mol. Biol. Evol.* 37:1100–1113.

885 Todorovic M, Sampaio F, Howard C. 2023. Sanger Tree of Life HMW DNA Fragmentation:
886 Diagenode Megaruptor®3 for PacBio HiFi. Available from:
887 https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-fragmentation-diagenod-
888 cztmx6k6.html

889 Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium,
890 Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, et al. 2023. MitoHiFi:
891 a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.
892 *BMC Bioinformatics* 24:288.

893 Vaillant JJ, Haffner GD, Cristescu ME. 2011. The ancient lakes of Indonesia: towards
894 integrated research on speciation. *Integr. Comp. Biol.* 51:634–643.

895 Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of
896 BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed
897 Processing Symposium (IPDPS). IEEE. p. 314–324.

898 Wasiljew BD, Pfaender J, Wipfler B, Gabelaia M, Utama IV, Wantania LL, Herder F. 2021.
899 Sexual dimorphism in an adaptive radiation: Does intersexual niche differentiation result
900 in ecological character displacement? *Ecol. Evol.* 11:14615–14629.

901 Wasiljew BD, Pfaender J, Wipfler B, Wantania LL, Herder F. 2022. Functional morphology of
902 prey capture in stream-dwelling sailfin silversides (Telmatherinidae) based on high-
903 speed video recordings. *Zoomorphology* 141:347–359.

904 Waskom M. 2021. seaborn: statistical data visualization. *J. Open Source Softw.* 6:3021.

905 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree
906 reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153.

907 Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool.
908 *Bioinformatics* 39:btac808.