

## ORIGINAL ARTICLE OPEN ACCESS

# Ancient and Recent Riverine Gene Flow Contributed to the Adaptive Radiation of Sailfin Silversides in Wallace's Dreampond

Els Lea R. De Keyzer<sup>1,2</sup>  | Fabian Herder<sup>3</sup>  | Astrid Böhne<sup>3</sup>  | Francisco Campuzano Jiménez<sup>1</sup>  | Valentina Burskaia<sup>1</sup>  | Sandra Kukowka<sup>3</sup> | Alan Tracey<sup>4</sup>  | Amy Louise Denton<sup>4</sup>  | Graeme Oatley<sup>4</sup>  | Wellcome Sanger Institute Tree of Life programme<sup>5</sup> | Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective<sup>6</sup> | Tree of Life Core Informatics Collective<sup>7</sup> | Daniel F. Mokodongan<sup>8</sup>  | Daisy Wowor<sup>8</sup>  | Hannes Svoldal<sup>1,9</sup> 

<sup>1</sup>Evolutionary Ecology Group, Department of Biology, University of Antwerp, Antwerp, Belgium | <sup>2</sup>Aquatic Macrofauna Biodiversity and Conservation Research Group, Hasanuddin University, Makassar, Indonesia | <sup>3</sup>Leibniz-Institute for the Analysis of Biodiversity Change (LIB), Museum Koenig Bonn, Bonn, Germany | <sup>4</sup>Wellcome Sanger Institute, Hinxton, UK | <sup>5</sup>Members: <https://doi.org/10.5281/zenodo.4783585>, Hinxton, UK | <sup>6</sup>Members: <https://doi.org/10.5281/zenodo.4790455>, Hinxton, UK | <sup>7</sup>Members: <https://doi.org/10.5281/zenodo.7116866>, Hinxton, UK | <sup>8</sup>Museum Zoologicum Bogoriense, Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, Indonesia | <sup>9</sup>Naturalis Biodiversity Center, Leiden, the Netherlands

**Correspondence:** Els Lea R. De Keyzer ([els.dekeyzer@uantwerpen.be](mailto:els.dekeyzer@uantwerpen.be)) | Hannes Svoldal ([hannes.svoldal@uantwerpen.be](mailto:hannes.svoldal@uantwerpen.be))

**Received:** 8 July 2025 | **Revised:** 16 April 2026 | **Accepted:** 12 May 2026

**Keywords:** hybrid zone | hybridisation | phylogenomics | speciation | *Telmatherina*

## 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. To establish a phylogenetic framework and investigate the presence of hybridisation in this radiation, 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 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, f<sub>4</sub>-ratio and f<sub>branch</sub> 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.

**Abbreviations:** GTR, general time reversible; HMW, high-molecular weight; ILS, incomplete lineage sorting; LD, linkage disequilibrium; Maf, minor allele frequency; ML, maximum-likelihood; Mya, million years ago; NJ, neighbour-joining.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

## 1 | 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 hybridisation 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 hybridisation 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). The lakes range in size from 4 to 561 km<sup>2</sup>, the system has been estimated to be 1- to 4-million years old, yet a clear age estimation of the lakes system is still lacking (Herder, Nolte et al. 2006; Vaillant et al. 2011; von Rintelen et al. 2004). 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 (164 km<sup>2</sup>), the most upstream and deepest (590 m) 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 benthopelagic

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: *Telmatherina prognatha* (Kottelat 1991), and *T. antoniae* (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 between the Malili Lakes (Herder, Nolte, et al. 2006; Herder, Schwarzer, et al. 2006). Finally, *T. bonti* Weber & de Beaufort, 1922 has been reported from Lawa river, the inlet stream of Lake Matano (Hadiaty and Wirjoatmodjo 2002). *T. bonti* are ram feeders, specialised on obtaining small food items in their fast-flowing stream habitats (Wasiljew et al. 2022).

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

Disentangling the evolutionary history of the adaptive radiation of sailfin silversides, including 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.

## 2 | Methods

### 2.1 | 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 (Table S1). 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 euthanised 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 ladigesii* were sampled from an aquarium population to serve as an outgroup in this study. These fish were euthanised by administering an overdose

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

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

RNA was extracted from muscle tissue of fTelBon1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax *mir*Vana protocol (do Amaral et al. 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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

### 2.2 | Reference Genome Construction

The reference genome for *T. bonti* (GCA\_933228915.1; [https://www.ncbi.nlm.nih.gov/assembly/GCA\\_933228915.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_933228915.1)) was sequenced from a single male *T. bonti* specimen bred at the LIB Bonn, Germany. A total of 36-fold coverage in PacBio long

reads. The HiFi reads were first assembled using Hifiasm version 0.16.1-r375 (Cheng et al. 2021) with the primary option. Haplotypic duplications were identified and removed using purge\_dups version 1.2.3 (Guan et al. 2020). The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin et al. 2019). The contigs were further scaffolded using the provided Hi-C data (Rao et al. 2014) in YaHS version yahs-1.1.91eebc2 (Zhou et al. 2023) using the break option. The scaffolded assemblies were evaluated using Gfastats (Formenti et al. 2022), BUSCO (Manni et al. 2021) and MERQURY.FK (Rhie et al. 2020).

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

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

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

## 2.3 | Short Read Quality Control and Mapping

Whole genome sequencing reads were filtered by BGI using the SOAPnuke software (Chen et al. 2018). Sequencing reads which matched > 25% adapter sequence, had a read length of < 150bp, > 0.1% N content or a Phred score < 33 were filtered out. Sequencing yielded ~80.1 million clean reads (after filtering) per sample. Sequences were quality checked using FastQC v0.11.9 (Andrews 2010) and MultiQC v1.12 (Ewels et al. 2016). Reads were aligned against the *T. bonti* (NCBI: txid446457) reference genome fTelBon1.1 with BWA-MEM (BWA v 0.7.17; Li 2013). The resulting SAM files were compressed to cram format and quality checked using samtools flagstat. After filtering and mapping each sample to the *T. bonti* reference genome, the resulting effective sequencing coverage varied between 13.8 and 22.9-fold, except for the outgroup samples (*M. ladigesi*) which yielded on average 10.5-fold coverage, related to their larger distance to the reference species.

Only single nucleotide polymorphisms (SNPs) were retained for genetic investigation. Variant sites were called jointly for all samples with the mpileup function in BCFtools v1.14 (Danecek et al. 2021). Total depth and mean coverage were calculated per site. Any sites with an overall mapping quality Phred-scaled score of less than 50 (4.9% of SNPs) or with more than 10% of mapped reads with a mapping quality zero (4.0% of SNPs) were excluded for analysis. Additionally, filters were applied to sites

with excess heterozygosity (InbreedingCoeff < 0.2; 0.03% of SNPs) before filtering and sites with a Phred-scaled *p*-value of binomial test for allele balance violation of > 40 (0.06% of SNPs). Sites with > 40% missing genotypes (2.0% of SNPs) as well as sites where the sum of overall depth for all samples was unusually high (> 3 units of standard deviation above individual mean coverage; 3.6% of SNPs) were filtered out. In total, 5.2% of SNPs were removed from the dataset (Figure S1), note that most sites were removed due to multiple filters. After filtering, 43,862,995 biallelic SNPs across 24 chromosomes remained. The mitogenome of the *T. bonti* reference was 16,560bp long. Filters were applied to mitochondrial sites with a mapping quality Phred-scaled score under 10 (19% of SNPs), sites with more than 50% of mapped reads with a mapping quality of zero (11% of SNPs), sites with a Phred-scaled *p*-value of binomial test for allele balance violation of > 40 (1.2% of SNPs), and sites with > 40% missing genotypes after application of other filters (23% of SNPs; Figure S1). After filtering, 1836 SNPs were retained in the aligned mitogenome.

## 2.4 | Phylogeny

### 2.4.1 | Nuclear Genome

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

In these remaining genomic windows of 100kb over the nuclear genome, local neighbour joining (NJ) trees were calculated based on distance matrices and a consensus tree was calculated with Astral-III v5.7.1 (Zhang et al. 2018). For the same dataset, Maximum likelihood (ML) trees were inferred using the General time reversible (GTR) model in IQ-TREE v2.2.0 (Minh et al. 2020), which accounts for unequal substitution rates between nucleotide pairs and unequal base frequencies. Bootstrap support was calculated for 1000 bootstrap replicates with ultrafast bootstrap approximation (UFBoot2; Hoang et al. 2018). The local trees were summarised as a majority consensus tree with SumTrees v4.5.2, distributed with the DendroPy v4.5.2 Python library (Sukumaran and Holder 2010), with minimum clade frequency set to zero. The consensus trees were visualised using Figtree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

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

### 2.4.2 | Mitochondrial Tree

We constructed a mitochondrial phylogeny from the SNP set (1836 nucleotide sites) of the mitogenomes with IQ-TREE v2.2.0

run with the GTR-model of nucleotide evolution and default parameters, using a *M. ladiges* sample as an outgroup. The consensus tree was constructed from 1000 bootstrap trees with the UltraFast bootstrap (UFBoot2) setting in IQ-Tree (Hoang et al. 2018).

## 2.5 | Phylogenomic Discordance

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

We quantified variation in genealogical relationships among taxa over the genome by weighting alternative topologies in each 100kb genomic window using Twisst (Martin and Van Belleghem 2017). Since the taxa in our study are closely related, incomplete lineage sorting (ILS) is expected to result in non-reciprocally monophyletic groups. For our three groups, *T. bonti*, roundfins and sharpfins, and the fixed outgroup *M. ladiges*, three alternative topologies are possible. We subsampled each windowed tree and quantified the relative occurrence of each topology. Topology 1 corresponds to the inferred species tree while topologies 2 and 3 would point towards incomplete lineage sorting or gene flow leading to gene-tree-species tree discordance.

## 2.6 | Genomic Diversity

To visualise similarity between individuals, principal component analysis (PCA) was performed on the SNP dataset with PLINK v2.0 (Purcell et al. 2007; Chang et al. 2015). For PCA analysis we only retained SNPs with a minor allele frequency (maf) of at least 0.05. PC analysis was done on the entire dataset, on the sharpfins and roundfins respectively alone, and on the sharpfins and riverine species (*T. bonti*) jointly. Figures were plotted with seaborn v0.12.2 (Waskom 2021).

We calculated Hamming distance in the nuclear genome with PLINK v2.0 (Purcell et al. 2007; Chang et al. 2015), and divided by the size of the filtered genome (accessible genome) to assess pairwise distance between all samples. We calculated heterozygosity of the entire genomes in windows of 200kb with ANGSD v0.921 (Korneliussen et al. 2014), based on the folded site frequency spectrum, using the *T. bonti* reference as the ancestral and filtering for a minimum read quality of 20, minimum mapping quality of 30, and adjusting mapping quality for excessive mismatches.

To assess whether the sharpfin, roundfin and riverine groups were genetically well separated or still shared substantial variation, we measured allele sharing across all filtered SNPs using a custom Python script with *cyvcf2* v0.31.1 (Pedersen and Quinlan 2017). For each SNP, we recorded whether both reference and alternative alleles were present within the roundfin, sharpfin and *T. bonti* groups, as well as in every pairwise and

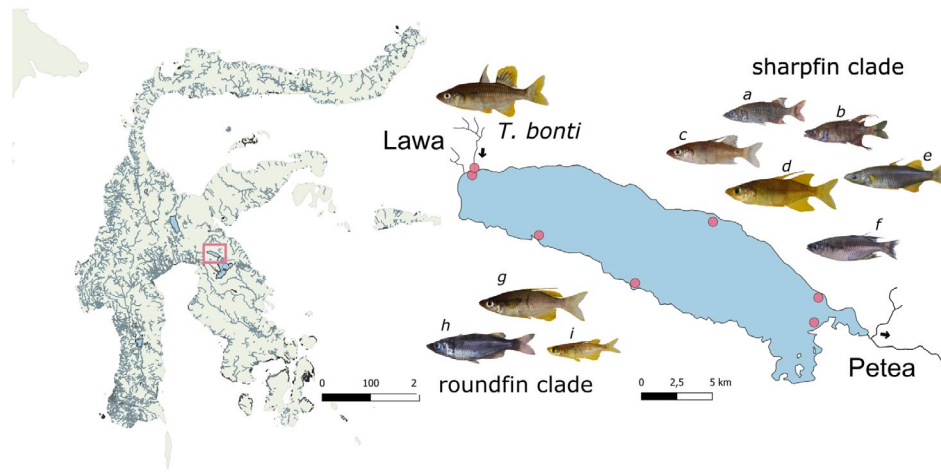
three-way combination of these groups. We identified SNPs that were unique to each group and those shared between groups. For this analysis we did not include the presumed hybrid specimen *T. bonti* 1. Missing alleles were not considered when classifying genotypes as heterozygous. The results were visualised in a Venn diagram using *matplotlib-venn* v1.1.2. As the reference genome used for mapping and SNP calling belongs to *T. bonti*, this analysis is biased towards detecting fewer non-reference alleles in the *T. bonti* group, potentially underestimating its shared variation with the other groups. In addition, the sample sets were unbalanced across groups, which may have led to an underestimation of allele sharing between *T. bonti* (represented by fewer samples) and the other groups.

We performed an admixture analysis to cluster samples into genetic groups without prior information on group assignments. Linkage disequilibrium (LD) pruning was performed on the filtered SNP dataset using plink v1.9 (Purcell et al. 2007; Chang et al. 2015), applying a sliding window of 50 SNPs with a step size of 5 SNPs. For pairs of SNPs exceeding an LD threshold of  $r^2 = 0.2$ , one SNP was removed from the dataset. Individual ancestry proportions were estimated using a maximum likelihood framework implemented in ADMIXTURE v1.3.0 (Alexander et al. 2009). The analysis was repeated for values of K (number of clusters) ranging from 6 to 12, with K = 7 showing the lowest cross validation error (CV error = 0.43848).

## 2.7 | Introgression

We calculated D-statistics (ABBA-BABA test) to assess gene flow between populations, as it allows us to distinguish between incomplete lineage sorting and introgressive hybridisation. This test compares allele sharing patterns between four groups: two sister groups (P1 and P2), a closely related third group (P3) and an outgroup (O). In the absence of gene flow, the two sister groups P1 and P2 would be expected to share the same amount of derived alleles (different from the allele in the outgroup O) with the third, related species P3. If, however, there is gene flow from P3 into P2, and not in P1, there will be an excess of shared derived alleles between P3 and P2, compared to P1, counted as a higher ratio of ABBA/BABA patterns. This pattern of excess allele sharing was tested for all possible (unordered) trios of individual samples in our dataset by computing genome-wide D statistics and f4-ratio with Dsuite v0.4 (Malinsky et al. 2021). For each test, an *M. ladiges* sample was used as an outgroup. F4-ratios of excess allele sharing between *T. bonti* and the sharpfin species relative to the roundfins is visualised, as well as excess allele sharing between *T. bonti* and the supposed hybrid specimen *T. bonti* 1.

If hybridisation events are more prevalent closer to the river-lake interface, this could lead to a spatial signal in the proportion of allele sharing of lake sharpfins with riverine *T. bonti*. We compared f4-ratios of allele sharing between *T. bonti* and sharpfins from several sampling sites at different distances from the lake-river interface, compared to a non-introgressed outgroup, the roundfins. Here we orient the roundfins as P1, the sharpfins as P2, and the riverine *T. bonti* as P3. This way, a higher f4-ratio indicates more allele sharing between *T. bonti* (P3) and the sharpfins (P2). We



**FIGURE 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* ( $n = 3$ ).

conducted this test for the three most common sharpfin species *T. sarasinorum*, *T. opudi*, and *T. abendanoni*.

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

### 3 | Results

#### 3.1 | High Quality Reference Genome and Annotation for *Telmatherina bonti*

A high-quality reference genome of *T. bonti* (fTelBon1.1, GCA\_933228915.1) was constructed as part of the tree of life project (<https://www.sanger.ac.uk/programme/tree-of-life/>) (Figures S2–S4). The final assembly had a total length of 986,039,709 bp in 146 sequence scaffolds with scaffold N50 and N90 sequence lengths of 41,523,244 and 35,388,860 bp respectively (Figure S2). 99.6% of the assembly sequence was assigned to 25 chromosomal-level scaffolds, representing 24 autosomes, numbered by sequence length, and the mitochondrion (Table S2). The assembly had a BUSCO completeness of 98% using the actinopterygii\_odb10 reference set (Figure S2). The Ensembl gene annotation system (Aken et al. 2016) was used to generate an annotation for the fTelBon1.1 assembly. The resulting annotation included 47,104 transcripts assigned to 21,413 coding and 4664 non-coding genes.

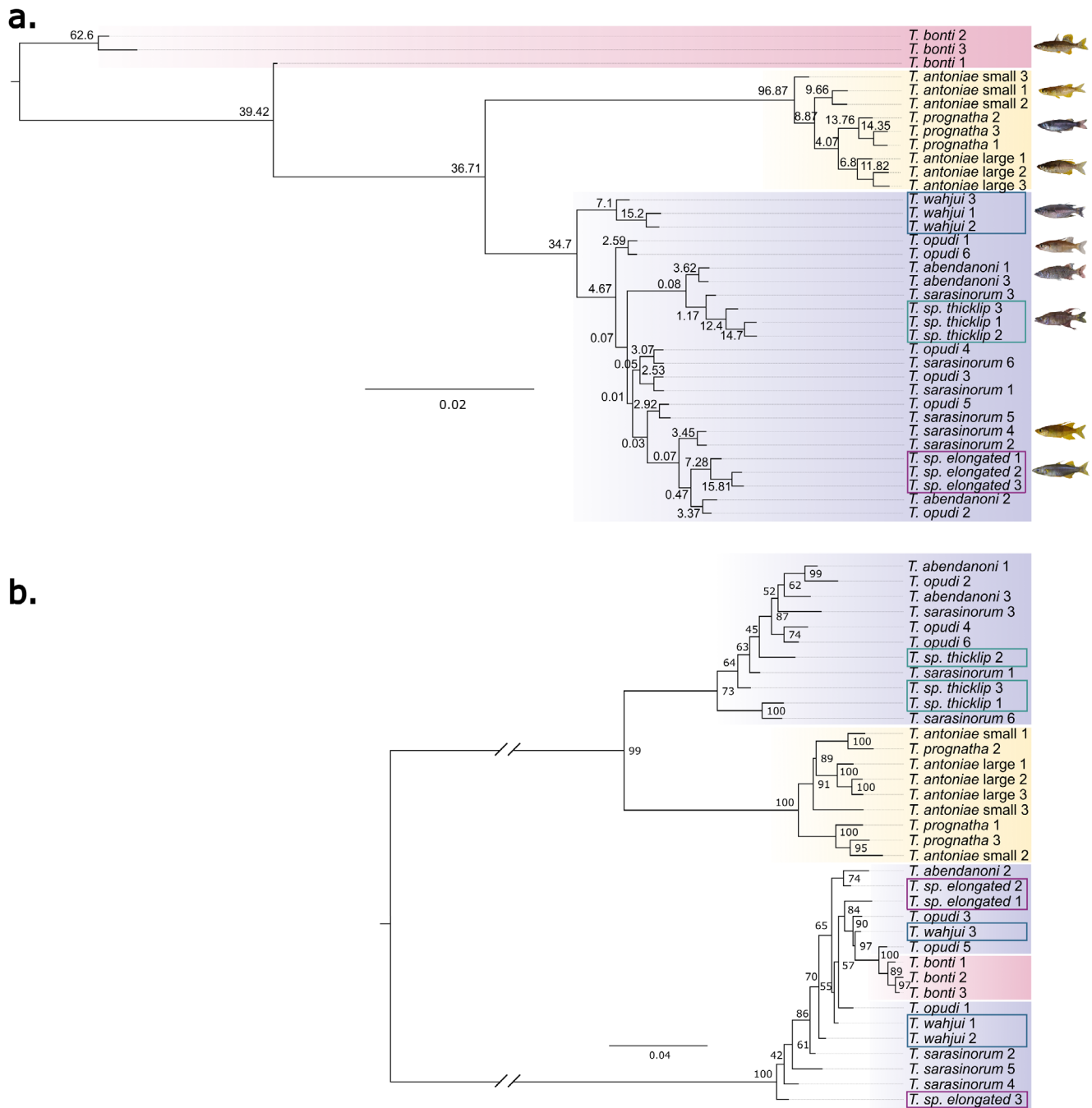
#### 3.2 | Genome Data Supports Roundfin and Sharpfin Clades and Most Taxa

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 (Figures 1 and S5, Table S1). We

constructed a neighbour joining (NJ) tree based on genome-wide pairwise differences (Figure S6). In addition, we built genome-wide and mitochondrial maximum-likelihood (ML) consensus phylogenies based on 100 kb windowed trees, and performed PCA and admixture analysis (Figures 2, 3 and S7–S9). All methods showed a clear separation between the benthic-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 (Figures 2a and S6) and well-defined clusters in a PC analysis only including roundfin samples (Figure S8). We found similarly strong support for three sharpfin species *T. sp. 'elongated'*, *T. sp. 'thicklip'* and *T. wahjui* (Figures 2a and 3a–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 (Figures 2 and S6), admixture analysis (Figure S9) or PC analyses (Figure 3b,c). Instead, those individuals showed a relatively large spread intermediate between *T. sp. 'thicklip'* and *T. sp. 'elongated'* along PC2 of an analysis only including sharpfin samples (Figure 3c).

#### 3.3 | Complex Relationships Between Lake and River Dwelling Species

Our sample set contained members of two stream-dwelling species: *T. wahjui* from Petea river, the outflow of Lake Matano, and *T. bonti* from Lawa river, the major tributary of Lake Matano. As mentioned above, *T. wahjui* clearly belongs to the sharpfin genetic clade, in which they fall basally to all other sharpfins (Figure 2a). That said, there was a relatively large spread of these individuals in the PC analysis of sharpfins (Figure 3b) consistent with variable ancestry proportions. *T. bonti*, the second stream dwelling species, showed an even more extreme spread among samples in the PC analysis, albeit along different axes of variation (PC2 in the analysis of all samples, Figure 3a;



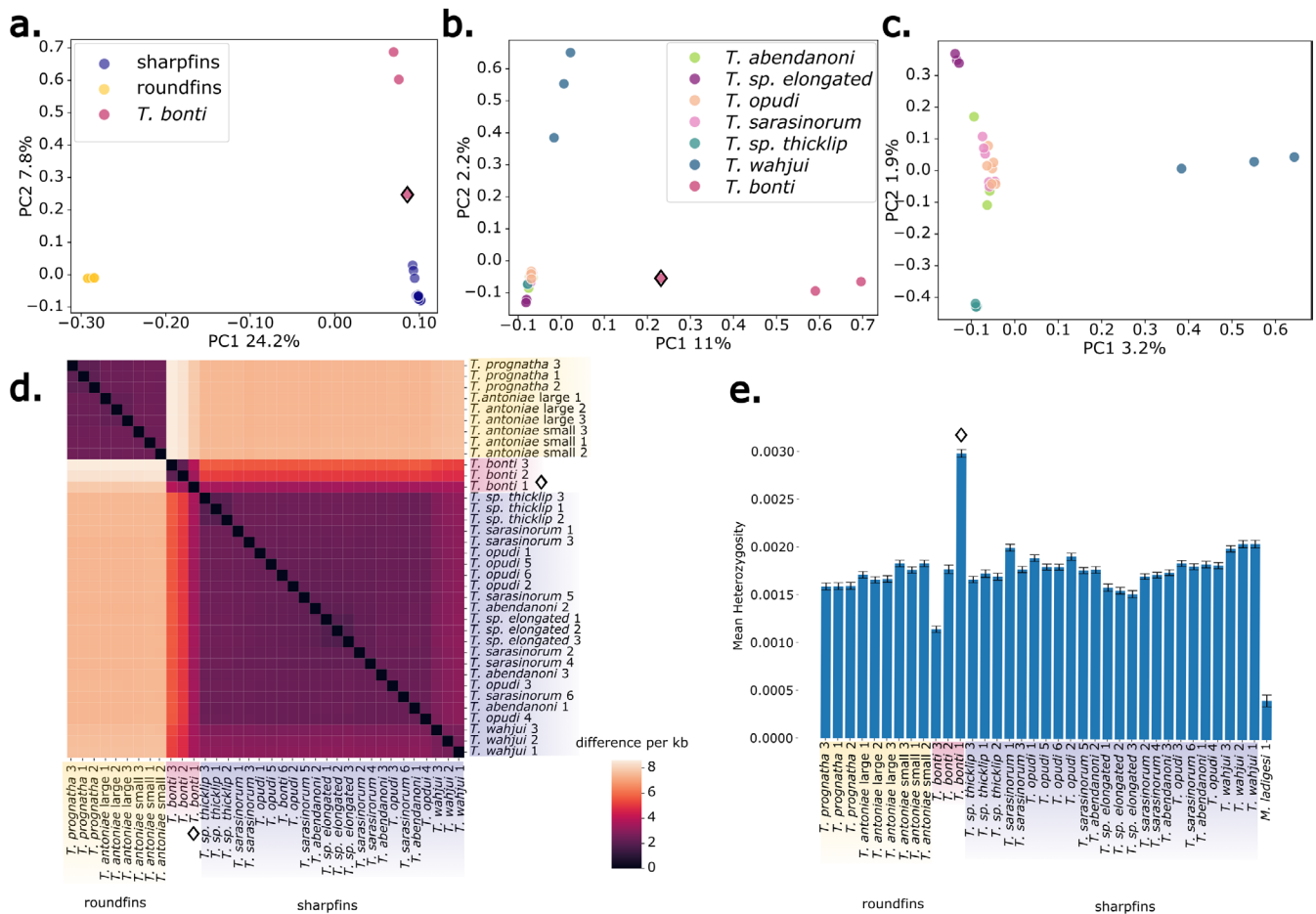
**FIGURE 2** | Phylogeny based on maximum likelihood (ML) trees. Sharpfin species are highlighted in purple, roundfins in yellow and the *T. bonti* samples are indicated in red background shading. Trees are rooted at outgroup species *M. ladigesii* (not shown in the figure, see Figure S7). (a) Consensus phylogeny based on 9150 ML trees in 100kb 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.

PC1 in an analysis of only sharpfins and *T. bonti*, Figure 3b). Consistent with this, *T. bonti* samples also showed inconsistent phylogenetic clustering for different inference methods (Figures 2 and S6). Specifically, while in the genome-wide neighbour joining tree (Figure S6) the three samples formed a sister group to the sharpfins, they formed an outgroup to both sharpfins and roundfins in the ML consensus tree (Figures 2a and S7). An astral consensus of NJ based local distance trees, on the other hand, gave an intermediate pattern with two *T. bonti* samples clustering as a global outgroup and one as a sister group

to sharpfins (Figure S6). Together, these analyses suggest that river-dwelling *Telmatherina* show within-population variation in their genetic relationships with the lacustrine species groups.

### 3.4 | Mitochondrial Phylogeny Splits Into Three Haplotype Groups

Our sample set was divided into three groups based on mitochondrial haplotypes (Figure 2b). One group, consisting entirely of



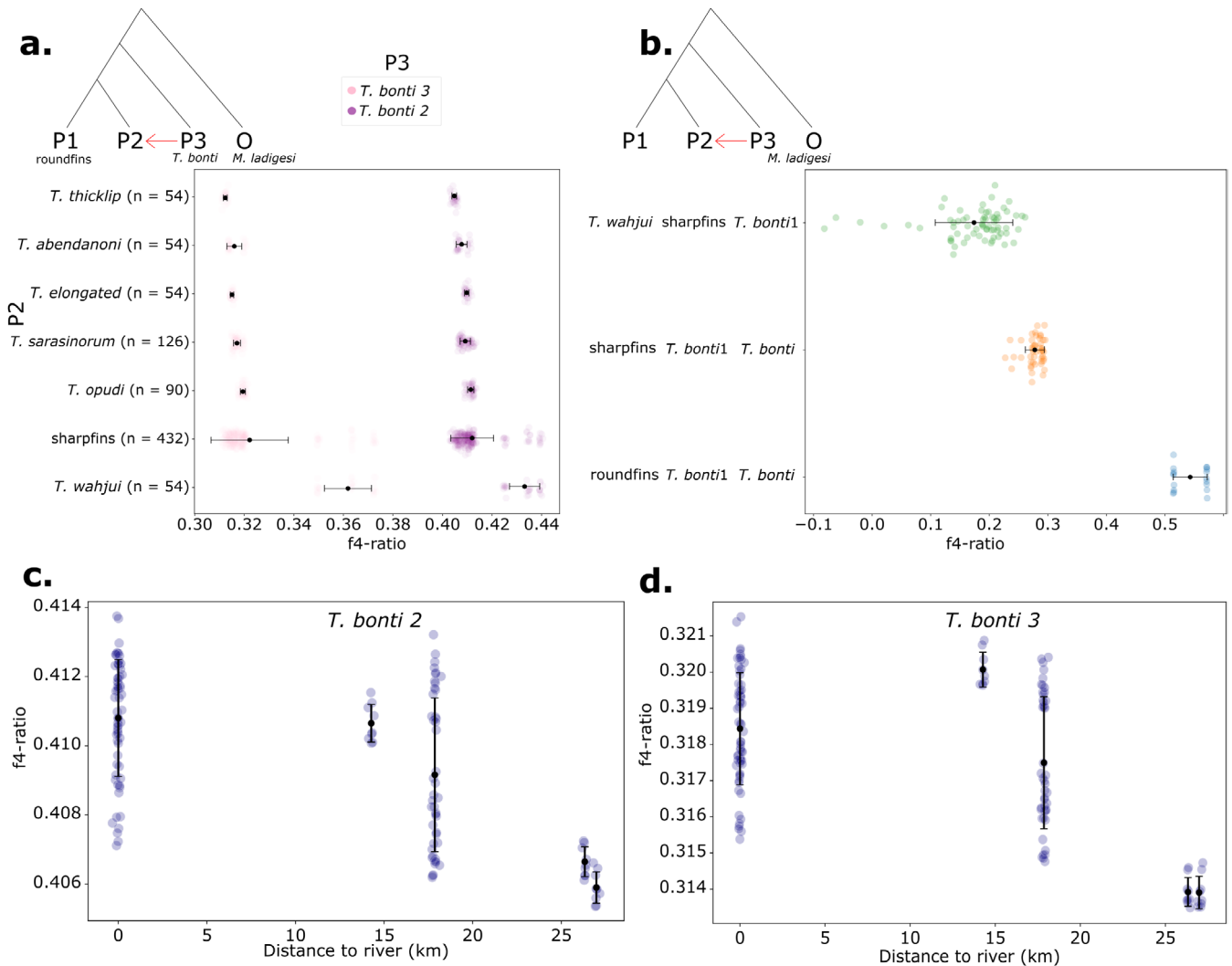
**FIGURE 3** | Principal component analysis and pairwise sequence divergence. (a–c) Visualisation of the first and second genomic principal components. (a) shows the Lake Matano sharpfins and roundfins, and the riverine *T. bonti*. (b) Lake Matano sharpfins separated by species and *T. bonti*. (c) Only Lake Matano sharpfin species. (d) Pairwise sequence divergence between individual samples visualised in a heatmap. (e) Mean heterozygosity in individual samples calculated over 200 kb 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.

roundfin samples, was in a sister relationship with a group that included approximately half of the sharpfin samples. The third group contained the remaining sharpfin samples along with all *T. bonti* samples. A similar pattern was reported by Stelbrink et al. (2014), who referred to the roundfin and sharpfin haplotypes in the sister group as ‘the original Matano haplotype’. All samples of the stream-dwelling sharpfin species *T. wahjui* and the lacustrine predator *T. sp. ‘elongated’* shared a haplotype similar to *T. bonti*. In contrast, all samples of the shrimp-eating sharpfin species *T. sp. ‘thicklip’* carried the original sharpfin haplotype. The remaining three sharpfin species—*T. opudi*, *T. sarasinorum* and *T. abendanoni*—did not consistently group with either haplotype cluster: roughly half of the samples had a *T. bonti*-like haplotype, while the others carried the original sharpfin haplotype. None of the roundfin nor sharpfin species were monophyletic in their haplotype groups.

### 3.5 | Incomplete Lineage Sorting and Low Variation Within Clades

Given the relatively young age of the radiation, the incomplete sorting of standing genetic variation between speciation

events (incomplete lineage sorting, ILS) could lead to variation in genetic relationships among species along the genome. And indeed, we found that among 9150 maximum likelihood trees, based on non-overlapping genomic windows of 100 kb, there was substantial topological variation. Consistent with this, we found shared polymorphism across all groups (Figure S10). Pairwise genetic distance was lowest among roundfin samples and highest between roundfin and both *T. bonti* and sharpfin samples (Figure 3d). When considering relationships among the three groups (sharpfins, roundfins, and *T. bonti*), 45.8% of subsampled trees supported the topology consistent with the genome-wide maximum likelihood tree, in which sharpfins and roundfins are sister groups. A nearly equal proportion of subsampled trees (44.3%) supported an alternative topology placing sharpfins as a sister group to *T. bonti*, and the remaining 9.8% of subsampled trees supported a third topology placing roundfins as a sister to *T. bonti* (Figure S11). The high proportion of alternative topologies is consistent with extensive incomplete lineage sorting. The strong asymmetry in occurrence between the two alternative topologies points towards a prominent contribution of gene flow in shaping local genetic relationships.



**FIGURE 4** | 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).  $n$  = number of trios tested. (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.

### 3.6 | Signals of Introgression Between Lake and River Dwelling Species

To identify genetic introgression among lake species and between lake and river dwelling species, we tested for signals of excess allele sharing. Elevated D-statistic, f4-ratio and f-branch statistics point towards introgression between *T. bonti* and all the sharpfin species (Figures 4a, S12 and S13) and between *T. wahjui* and other sharpfin species (Figure S13). These results are consistent with the above-mentioned topology weighting analysis showing a strong excess of alternative topologies where sharpfins group with *T. bonti* compared to roundfins grouping with *T. bonti*. The strong asymmetry in the frequency of alternative topologies is consistent with introgression, rather than incomplete lineage sorting alone, contributing to the observed phylogenetic discordance

(Figure S11). Interestingly, introgressed regions are most frequent at the center of the chromosomes. Comparison of admixture proportions (f4-ratio) of *T. bonti* into the sharpfins *T. abendanoni*, *T. sarasinorum* and *T. opudi* from different sampling sites revealed a spatial gradient of declining admixture with increasing distance from the river–lake interface (Figure 4c,d). This is suggestive of relatively recent gene flow having happened at the river–lake interface that is diffusing into lake sharpfin populations through migration. One of the samples in our dataset, ‘*T. bonti* 1’, while morphologically identified as *T. bonti*, appears to be a recent *T. bonti* x sharpfin hybrid based on its position in the phylogenetic trees (Figure 2a), in the PCA (Figure 3c), and its high level of heterozygosity (Figure 3e). Admixture analysis assigns this sample almost exactly 50% *T. bonti* and 50% sharpfin-like ancestry as would be expected for an F1 individual (Figure S9).

## 4 | 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. abendani*, which do not show any clear separation into distinct gene pools. Furthermore, river-dwelling *Telmatherina* show within-population variation in their genetic relationships with the lacustrine species groups. We also confirm a mito-nuclear discordance in the sharpfin clade, indicative of introgression between sharpfins and *T. bonti*. Introgression between riverine species and the lacustrine *Telmatherina* is apparent in elevated levels of excess allele sharing, which decline with increasing distance from the lake–river interface.

### 4.1 | New Species Status for Sharpfins and Roundfins

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

The phylogenetic signal for the roundfin clade is clear throughout our analysis and consistent with the described morphospecies, *T. prognatha*, *T. antoniae* ‘small’ and *T. antoniae* ‘large’ (Herder et al. 2008; Pfaender et al. 2011). This result extends previous studies based on fewer genetic markers, which found no (Herder, Nolte, et al. 2006) or only low (Herder et al. 2008) genetic differentiation between the three roundfin species. Based on whole genome sequences, we find *T. antoniae* ‘large’ to be more closely related to *T. prognatha* than to *T. antoniae* ‘small’. We consider *T. antoniae* ‘large’ and *T. antoniae* ‘small’ as separate species and urge for an official taxonomic revision of the roundfin clade, based on our findings regarding the genomic phylogeny and previous findings regarding morphology, adaptive niche differentiation, assortative mating and diet (Herder et al. 2008; Pfaender et al. 2011). The species status of *T. prognatha* has not been under debate in the literature and does also not raise any concerns in our analysis. Even though the overall phylogenetic signal in the roundfins is clear, we found alternative topologies where roundfin species are not monophyletic in some parts of the genome, and the maximum likelihood tree based

on mitochondrial DNA does not clearly distinguish between the three roundfin species. This inconsistency could indicate that reproductive isolation between the species is recent or be a signal of incomplete isolation.

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

### 4.2 | Speciation and Signatures of Past Introgression in *Telmatherina*

In this study, we found strong signals of past introgression of the riverine *T. bonti* into lacustrine sharpfins, shown as elevated values of the  $f_4$ -ratio and  $f$ -branch statistic, as well as an imbalance among alternative gene tree topologies. 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 hybridisation 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). Hybridisation 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, 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 and introgression patterns in *Telmatherina* and related species across the rivers and lakes of the Malili system,

will provide insight into how gene flow has shaped diversity in the sharpfin clade. Additionally, the observed patterns of non-random distribution of introgressed genomic regions should be further investigated in future work.

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

### 4.3 | Recent and Ongoing Gene Flow Between Matano Sharpfins and the Riverine Group

Besides signals of past gene flow, we also detected recent gene flow between the stream-dwelling *T. bonti* and Matano species. In the maximum likelihood (ML) consensus trees, assuming a GTR model, *T. bonti* is positioned as an outgroup to both the sharpfins and roundfins. On the other hand, the NJ tree suggests closer genetic similarity between *T. bonti* and sharpfins relative to roundfins, possibly due to reduced genetic distance resulting from 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, and its shared ancestry in the admixture analysis. 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. An interesting possibility which we could not further explore is that one or both of the two other *T. bonti* samples in our dataset might themselves be backcrosses between *T. bonti* × sharpfin hybrids and *T. bonti*, as they were collected a bit upstream of the hybrid zone. This would suggest 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. More fine grained sampling along the Lawa river will reveal if there is a spatial gradient of proportion of introgressed sharpfin alleles in *T. bonti*.

We detected clear signals of decreased excess allele sharing (lower  $f_4$ -ratios) with increasing distance to the river–lake interface between *T. bonti* and three sharpfin species which occur over the entire lake, *T. sarasinorum*, *T. opudi*, and *T. abendanoni*. This pattern is consistent with a hybrid zone at the interface between the inlet river Lawa and the lake, possibly facilitated by natural selection favouring hybrid individuals in this ecologically intermediate zone. This spatial decline in excess allele sharing is a strong indication of ongoing or recent gene flow, as signals of older gene flow would over time have homogenised over the extent of the lake, erasing spatial signals.

### 4.4 | Signals of Incomplete Speciation

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

## 5 | 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.

### Author Contributions

E.L.R.D.K., F.H., A.B. and H.S. conceived the study. E.L.R.D.K., F.C.J., V.B. and H.S. conducted analysis on resequencing data. S.K., A.D. and G.O. conducted molecular laboratory work. The Wellcome Sanger Institute Tree of life programme organised and conducted fTelBon1.1 reference genome assembly and annotation. A.T. led manual curation of the fTelBon1.1 reference genome. D.W., D.F.M. and F.H. organised and conducted data collection. E.L.R.D.K. and H.S. wrote the manuscript with input from all co-authors.

### Acknowledgements

No samples have been collected for this study. The specimens used in this study were available from museum collections of the Museum Zoologicum Bogoriense (MZB) in Bogor, Indonesia, and the Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany. Original sampling was conducted between 2002 and 2010 with permits from the Indonesian government (Surat Izin Penelitian) 5602/SU/KS/2002 (LIPI) and 0093/SIP/FRP/SM/V/2010 (RISTEK). We thank the Heads of the 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, providing voucher samples re-analysed in the present project. We are grateful for help with analysis and interpretation of results from Ilia Artiushin and other members of the Svardal lab (Universiteit Antwerpen). Access to VSC infrastructure was facilitated by the CalcUA team at Universiteit Antwerpen. E.L.R.D.K. is supported as a postdoctoral fellow in fundamental research by the Flemish Research Foundation (FWO postdoctoral scholarship 12A8423N). F.C.J. is supported as a doctoral fellow in fundamental research by the Flemish Research Foundation (FWO PhD scholarship 11A2P26N). Research and collaboration between Indonesia and Europe were supported by Research Foundation—Flanders (FWO) (project G0A9B24N to H.S. and travel grant V408023N to E.L.R.D.K.) and VLIR-UOS (SI project ID2022SIN349A102 to H.S., GMSR project 54084 to E.L.R.D.K. and SI project ID2025SIN4SEL104 to H.S. and E.L.R.D.K.). We gratefully acknowledge support of the Research Fund of the University of Antwerp (BOF Seal of Excellence and KPBOF project 48620 to E.L.R.D.K.).

### Funding

This work was supported by Fonds Wetenschappelijk Onderzoek, G0A9B24N, V408023N, 12A8423N, 11A2P26N, VLIRUOS, ID2022SIN349A102, ID2025SIN4SEL104, 54084, Universiteit Antwerpen, 48620.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The genome assembly fTelBon1.1 is available on NCBI (<https://www.ncbi.nlm.nih.gov/>) under bioproject PRJEB51590 ([https://www.ncbi.nlm.nih.gov/datasets/genome/GCA\\_933228915.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_933228915.1/)). Raw reads of the resequencing data are available at the European nucleotide archive (ENA; <https://www.ebi.ac.uk/>) under study PRJEB45386.

### References

- Aken, B. L., S. Ayling, D. Barrell, et al. 2016. "The Ensembl Gene Annotation System." *Database: The Journal of Biological Databases and Curation* 2016: baw093.
- Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19: 1655–1664. <http://paperpile.com/b/PTQ89m/q0fV>.
- Allio, R., A. Schomaker-Bastos, J. Romiguier, F. Prosdocimi, B. Nabholz, and F. Delsuc. 2020. "MitoFinder: Efficient Automated Large-Scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics." *Molecular Ecology Resources* 20: 892–905.
- Andrews, S. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data." <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arnold, M. 2006. "Evolution Through Genetic Exchange." <https://academic.oup.com/book/3287>.
- Berner, D., and W. Salzburger. 2015. "The Genomics of Organismal Diversification Illuminated by Adaptive Radiations." *Trends in Genetics* 31: 491–499.
- Brent, S. P., and R. Q. Aaron. 2017. "fast, flexible variant analysis with Python." *Bioinformatics* 33, no. 12: 1867–1869. <https://doi.org/10.1093/bioinformatics/btx057>.
- Cerwenka, A. F., J. D. Wedekind, R. K. Hadiaty, U. K. Schliewen, and F. Herder. 2012. "Alternative Egg-Feeding Tactics in *Telmatherina sarasinorum*, a Trophic Specialist of Lake Matano's Evolving Sailfin Silversides Fish Radiation." *Hydrobiologia* 693: 131–139.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4: 7.
- Chen, Y., Y. Chen, C. Shi, et al. 2018. "SOAPnuke: A MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data." *GigaScience* 7: 1–6.
- Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. "Haplotype-Resolved De Novo Assembly Using Phased Assembly Graphs With Hifiasm." *Nature Methods* 18: 170–175.
- Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10. <https://doi.org/10.1093/gigascience/giab008>.
- Denton, A., G. Oatley, C. Cornwell, M. Quail, and C. Howard. 2023. "Sanger Tree of Life Sample Homogenisation: PowerMash." <https://www.protocols.io/view/sanger-tree-of-life-sample-homogenisation-powermas-cyf8xtrw.html>.

- Diesh, C., G. J. Stevens, P. Xie, et al. 2023. "JBrowse 2: A Modular Genome Browser With Views of Synteny and Structural Variation." *Genome Biology* 24: 74.
- do Amaral, R. J. V., A. Denton, H. Yatsenko, J. Jay, C. Howard, and A. Bates. 2023. "Sanger Tree of Life RNA Extraction: Automated MagMax mirVana." <https://www.protocols.io/view/sanger-tree-of-life-rna-extraction-automated-magma-cxxnpxme.html>.
- Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics* 32: 3047–3048.
- Formenti, G., L. Abueg, A. Brajuka, et al. 2022. "Gfastats: Conversion, Evaluation and Manipulation of Genome Sequences Using Assembly Graphs." *Bioinformatics* 38: 4214–4216.
- Gray, S. M., and J. S. McKinnon. 2006. "A Comparative Description of Mating Behaviour in the Endemic Telmatherinid Fishes of Sulawesi's Malili Lakes." *Environmental Biology of Fishes* 75: 471–482.
- Greenwood, M. P., T. Capblancq, N. Wahlberg, and L. Després. 2025. "Whole Genome Data Confirm Pervasive Gene Discordance in the Evolutionary History of *Coenonympha* (Nymphalidae) Butterflies." *Molecular Phylogenetics and Evolution* 202: 108222.
- Guan, D., S. A. McCarthy, J. Wood, K. Howe, Y. Wang, and R. Durbin. 2020. "Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies." *Bioinformatics* 36: 2896–2898.
- Hadiaty, R. K., and S. Wirjoatmodjo. 2002. "Studi Pendahuluan Biodiversitas dan Distribusikan di Danau Matano, Sulawesi Selatan [Preliminary Study: Biodiversity and Distribution of Fishes in Lake Matano, South Sulawesi]." *Jurnal Ikhtologi Indonesia* 2: 23–29.
- Harry, E. 2022. PretextView (Paired REad TEXTure Viewer): A Desktop Application for Viewing Pretext Contact Maps. <https://github.com/sanger-tol/PretextView>.
- Herder, F., A. W. Nolte, J. Pfaender, J. Schwarzer, R. K. Hadiaty, and U. K. Schliwen. 2006. "Adaptive Radiation and Hybridization in Wallace's Dreamponds: Evidence From Sailfin Silversides in the Malili Lakes of Sulawesi." *Proceedings of the Biological Sciences* 273: 2209–2217.
- Herder, F., J. Pfaender, and U. K. Schliwen. 2008. "Adaptive Sympatric Speciation of Polychromatic "Roundfin" Sailfin Silverside Fish in Lake Matano (Sulawesi)." *Evolution* 62: 2178–2195.
- Herder, F., and U. K. Schliwen. 2010. "Beyond Sympatric Speciation: Radiation of Sailfin Silverside Fishes in the Malili Lakes (Sulawesi)." In *Evolution in Action*, 465–483. Springer Berlin Heidelberg.
- Herder, F., J. Schwarzer, J. Pfaender, R. K. Hadiaty, and U. K. Schliwen. 2006. "Preliminary Checklist of Sailfin Silversides (Teleostei: Telmatherinidae) in the Malili Lakes of Sulawesi (Indonesia), With a Synopsis of Systematics and Threats." *Verhandlungen der Gesellschaft für Ichthyologie* 5: 139–163.
- Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35: 518–522.
- Howe, K., W. Chow, J. Collins, et al. 2021. "Significantly Improving the Quality of Genome Assemblies Through Curation." *GigaScience* 10: gaa153.
- Jay, J., H. Yatsenko, J. P. Narváez-Gómez, et al. 2023. "Sanger Tree of Life Sample Preparation: Triage and Dissection." <https://www.protocols.io/view/sanger-tree-of-life-sample-preparation-triage-and-cztex6je.html>.
- Junier, T., and E. M. Zdobnov. 2010. "The Newick Utilities: High-Throughput Phylogenetic Tree Processing in the UNIX Shell." *Bioinformatics* 26: 1669–1670.
- Kerpedjiev, P., N. Abdennur, F. Lekschas, et al. 2018. "HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps." *Genome Biology* 19: 125.
- Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15: 356.
- Kottelat, M. 1991. "Sailfin Silversides (Pisces: Telmatherinidae) of Lake Matano, Sulawesi, Indonesia, With Descriptions of Six New Species." *Ichthyological Exploration of Freshwaters* 1: 321–344.
- Kozak, K. M., M. Joron, W. O. McMillan, and C. D. Jiggins. 2021. "Rampant Genome-Wide Admixture Across the Heliconius Radiation." *Genome Biology and Evolution* 13: evab099.
- Lamichhaney, S., J. Berglund, M. S. Almén, et al. 2015. "Evolution of Darwin's Finches and Their Beaks Revealed by Genome Sequencing." *Nature* 518: 371–375.
- Li, H. 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM arXiv [q-bio.GN]. <http://arxiv.org/abs/1303.3997>.
- Loh, Y.-H. E., E. Bezault, F. M. Muenzel, et al. 2013. "Origins of Shared Genetic Variation in African Cichlids." *Molecular Biology and Evolution* 30: 906–917.
- Malinsky, M., M. Matschiner, and H. Svardal. 2021. "Dsuite—Fast D-Statistics and Related Admixture Evidence From VCF Files." *Molecular Ecology Resources* 21: 584–595.
- Malinsky, M., H. Svardal, A. M. Tyers, et al. 2018. "Whole-Genome Sequences of Malawi Cichlids Reveal Multiple Radiations Interconnected by Gene Flow." *Nature Ecology & Evolution* 2: 1940–1955.
- Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov. 2021. "BUSCO Update: Novel and Streamlined Workflows Along With Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes." *Molecular Biology and Evolution* 38: 4647–4654.
- Marques, D. A., J. I. Meier, and O. Seehausen. 2019. "A Combinatorial View on Speciation and Adaptive Radiation." *Trends in Ecology & Evolution* 34: 531–544.
- Martin, S. H., and S. M. Van Belleghem. 2017. "Exploring Evolutionary Relationships Across the Genome Using Topology Weighting." *Genetics* 206: 429–438. <http://paperpile.com/b/PTQ89m/ChPo>.
- McLay, T. G. B., R. M. Fowler, P. S. Fahey, et al. 2023. "Phylogenomics Reveals Extreme Gene Tree Discordance in a Lineage of Dominant Trees: Hybridization, Introgression, and Incomplete Lineage Sorting Blur Deep Evolutionary Relationships Despite Clear Species Groupings in Eucalyptus Subgenus Eudesmia." *Molecular Phylogenetics and Evolution* 187: 107869.
- Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. "Ancient Hybridization Fuels Rapid Cichlid Fish Adaptive Radiations." *Nature Communications* 8: 14363.
- Meier, J. I., M. D. McGee, D. A. Marques, et al. 2023. "Cycles of Fusion and Fission Enabled Rapid Parallel Adaptive Radiations in African Cichlids." *Science* 381: eade2833.
- Minh, B. Q., H. A. Schmidt, O. Chernomor, et al. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37: 1530–1534.
- Nolte, A. W., and D. Tautz. 2010. "Understanding the Onset of Hybrid Speciation." *Trends in Genetics* 26: 54–58.
- Peñalba, J. V., A. Runemark, J. I. Meier, et al. 2024. "The Role of Hybridization in Species Formation and Persistence." *Cold Spring Harbor Perspectives in Biology* 16: a041445.
- Pfaender, J., R. K. Hadiaty, U. K. Schliwen, and F. Herder. 2016. "Rugged Adaptive Landscapes Shape a Complex, Sympatric Radiation." *Proceedings of the Royal Society B: Biological Sciences* 283: 20152342.
- Pfaender, J., F. W. Miesen, R. K. Hadiaty, and F. Herder. 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." *Journal of Evolutionary Biology* 24: 2329–2345.
- Pfaender, J., U. K. Schlieven, and F. Herder. 2010. "Phenotypic Traits Meet Patterns of Resource Use in the Radiation of "Sharpfin" Sailfin Silverside Fish in Lake Matano." *Evolutionary Ecology* 24: 957–974.
- Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81: 559–575.
- Rao, S. S. P., M. H. Huntley, N. C. Durand, et al. 2014. "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." *Cell* 159: 1665–1680.
- Rhie, A., B. P. Walenz, S. Koren, and A. M. Phillippy. 2020. "Mercury: Reference-Free Quality, Completeness, and Phasing Assessment for Genome Assemblies." *Genome Biology* 21: 245.
- Rosser, N., F. Seixas, L. M. Queste, et al. 2024. "Hybrid Speciation Driven by Multilocus Introgression of Ecological Traits." *Nature* 628: 811–817.
- Roy, D., M. F. Docker, P. Hehanussa, D. D. Heath, and G. D. Haffner. 2004. "Genetic and Morphological Data Supporting the Hypothesis of Adaptive Radiation in the Endemic Fish of Lake Matano." *Journal of Evolutionary Biology* 17: 1268–1276.
- Schwarzer, J., F. Herder, B. Misof, R. K. Hadiaty, and U. K. Schlieven. 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 in Ecology & Evolution* 19: 198–207.
- Seehausen, O. 2013. "Conditions When Hybridization Might Predispose Populations for Adaptive Radiation." *Journal of Evolutionary Biology* 26: 279–281.
- Seehausen, O., G. Takimoto, D. Roy, and J. Jokela. 2008. "Speciation Reversal and Biodiversity Dynamics With Hybridization in Changing Environments." *Molecular Ecology* 17: 30–44.
- Sheerin, E., F. Sampaio, G. Oatley, et al. 2023. "Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.1." <https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-extraction-automated-m-czh2x38e.html>.
- Stelbrink, B., I. Stöger, R. K. Hadiaty, U. K. Schlieven, and F. Herder. 2014. "Age Estimates for an Adaptive Lake Fish Radiation, Its Mitochondrial Introgression, and an Unexpected Sister Group: Sailfin Silversides of the Malili Lakes System in Sulawesi." *BMC Evolutionary Biology* 14: 94.
- Strickland, M., C. Cornwell, and C. Howard. 2023. "Sanger Tree of Life Fragmented DNA Clean Up: Manual SPRI." <https://www.protocols.io/view/sanger-tree-of-life-fragmented-dna-clean-up-manual-czhkx34w.html>.
- Sukumaran, J., and M. T. Holder. 2010. "DendroPy: A Python Library for Phylogenetic Computing." *Bioinformatics* 26: 1569–1571.
- Svardal, H., F. X. Quah, M. Malinsky, et al. 2020. "Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation." *Molecular Biology and Evolution* 37: 1100–1113.
- Todorovic, M., F. Sampaio, and C. Howard. 2023. "Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor3 for PacBio HiFi." <https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-fragmentation-diagenod-cztmx6k6.html>.
- Uliano-Silva, M., J. G. R. N. Ferreira, K. Krashenninnikova, et al. 2023. "MitoHiFi: A Python Pipeline for Mitochondrial Genome Assembly From PacBio High Fidelity Reads." *BMC Bioinformatics* 24: 288.
- Vaillant, J. J., G. D. Haffner, and M. E. Cristescu. 2011. "The Ancient Lakes of Indonesia: Towards Integrated Research on Speciation." *Integrative and Comparative Biology* 51: 634–643.
- Vasimuddin, M., S. Misra, H. Li, and S. Aluru. 2019. "Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems." In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314–324. IEEE.
- von Rintelen, T., K. von Rintelen, M. Glaubrecht, C. Schubart, and F. Herder. 2012. "Aquatic Biodiversity Hotspots in Wallacea: The Species Flocks in the Ancient Lakes of Sulawesi, Indonesia." In *Biotic Evolution and Environmental Change in Southeast Asia*, edited by D. Gower, K. Johnson, J. Richardson, B. Rosen, L. Ruber, and S. Williams, 290–315. Cambridge University Press.
- von Rintelen, T., A. B. Wilson, A. Meyer, and M. Glaubrecht. 2004. "Escalation and Trophic Specialization Drive Adaptive Radiation of Freshwater Gastropods in Ancient Lakes on Sulawesi, Indonesia." *Proceedings Biological Sciences* 271: 2541–2549. <http://paperpile.com/b/PTQ89m/myTK>.
- Wasiljew, B. D., J. Pfaender, B. Wipfler, et al. 2021. "Sexual Dimorphism in an Adaptive Radiation: Does Intersexual Niche Differentiation Result in Ecological Character Displacement?" *Ecology and Evolution* 11: 14615–14629.
- Wasiljew, B. D., J. Pfaender, B. Wipfler, L. L. Wantania, and F. Herder. 2022. "Functional Morphology of Prey Capture in Stream-Dwelling Sailfin Silversides (Telmatherinidae) Based on High-Speed Video Recordings." *Zoomorphology* 141: 347–359.
- Waskom, M. 2021. "Seaborn: Statistical Data Visualization." *Journal of Open Source Software* 6: 3021.
- Weber, M. 1913. "Neue Beiträge zur Kenntnis der Süßwasserfische von Celebes. Ergebnisse einer Reise von E. C. Abendenon in Celebes." *Bijdragen tot de Dierkunde* 19: 197–213.
- Weber, M., and L. F. de Beaufort. 1922. *The Fishes of the Indo-Australian Archipelago. IV: Heteromi, Soleichthyes, Synentognathi, Percosoces, Labyrinthi, Microcyprini*. Brill, E.J. <https://doi.org/10.5962/bhl.title.12497>.
- Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. "ASTRAL-III: Polynomial Time Species Tree Reconstruction From Partially Resolved Gene Trees." *BMC Bioinformatics* 19: 153.
- Zhou, C., S. A. McCarthy, and R. Durbin. 2023. "YaHS: Yet Another Hi-C Scaffolding Tool." *Bioinformatics* 39: btac808.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Filter settings and number of sequences removed per filter. Area highlighted in red indicates the filter threshold (a) for the nuclear genome and (b) for the mitochondrial DNA. **Figure S2:** Assembly statistics for reference genome assembly of *Telmatherina bontii*. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1000 size-ordered bins around the circumference with each bin representing 0.1% of the 986,039,709bp 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 (48,058,054bp). Orange and pale-orange arcs show the N50 and N90 sequence lengths (41,523,244 and 35,388,860bp), 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>. **Figure S3:** Distribution plot of base coverage in ERR8978456 against position for sequences in assembly CAKOGG01. 100kb windows are coloured by phylum. **Figure S4:** (a) Cumulative sequence length for assembly CAKOGG01. The grey line shows cumulative length for all sequences.

Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrules. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/Telmathetherina%20bonti/dataset/CAKOGG01/cumulative>. (b) Blob plot of base coverage in ERR8978456 against GC proportion for sequences in assembly CAKOGG01. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/Telmathetherina%20bonti/dataset/CAKOGG01/blob>. **Figure S5:** Map of sampling locations. Names of sampling locations correspond to sample information as displayed in Table S2. Inflow river Lawa and outflow river Petea are indicated. **Figure S6:** (a) NJ tree based on distance matrix. Scale is the number of mutations divided by accessible genome size. (b) Astral consensus of NJ trees based local distance trees based on 100 kb windows. All trees are rooted against a sample of the outgroup species *M. ladigesii*. Sharpfin species are highlighted in purple, roundfins in yellow and the *T. bonti* samples are indicated in red background shading. **Figure S7:** Phylogeny based on maximum likelihood (ML) trees with outgroup species *Marosatherina ladigesii*. Consensus phylogeny based on 9150 ML trees in 100 kb windows across the genome. Sharpfin species are highlighted in purple, roundfins in yellow and the *T. bonti* samples are indicated in red background shading. **Figure S8:** Visualisation of the first and second principal components based on the SNP dataset for Lake Matano roundfin species. **Figure S9:** Ancestry composition based on admixture analysis. Each bar is an individual, bars are coloured according to the proportion of ancestry assigned to each of the seven inferred genetic clusters. Diamond shape (◊) indicates *T. bonti* 1, the presumed recent hybrid sample. Sharpfin species are highlighted in purple, roundfins in yellow and the *T. bonti* samples are indicated in red background shading. **Figure S10:** Venn-diagram of shared and private variable sites in sharpfins ( $n = 24$ ), roundfins ( $n = 9$ ) and *T. bonti* ( $n = 2$ ). Outer circles represent the number of variable sites unique to the corresponding group. Overlapping areas indicate the number of variable sites shared between the respective groups, that is, where both alleles were present in both or all three groups. **Figure S11:** Twisst analysis showing frequency of topologies in 100 kb windows over the genome. Topo1, blue: (((sharpfins, roundfins), *T. bonti*), *M. ladigesii*); Topo2, orange: (((sharpfins, *T. bonti*), roundfins), *M. ladigesii*); Topo3, green: (((roundfins, *T. bonti*), sharpfins), *M. ladigesii*). For each window the proportion of subsampled trees supporting each of the topologies is indicated. **Figure S12:** Excess allele sharing (f4-ratio) between *T. bonti* 3 (pink), *T. bonti* 2 (purple), *T. bonti* 1 (blue) (P3) and the sharpfin species indicated on Y axis (P2), relative to the roundfins (P1). **Figure S13:** Heatmap of pairwise f<sub>branch</sub> (fb) statistic, based on SNP dataset. Higher values (darker colour) indicate more excess allele sharing between specimens. **Table S1:** Sample information—sample ID as used in the manuscript and additional information about the samples used in the study. Sampling locations correspond to locations indicated in Figure S3. **Table S2:** Chromosomes of fTelBon1: chromosome size, GC-content in percentage, average coverage and accession number.