Molecular Ecology

MOLECULAR ECOLOGY

ORIGINAL ARTICLE OPEN ACCESS

Haemoglobin Gene Repertoire in Teleost and Cichlid Fishes Shaped by Gene Duplications and Genome Rearrangements

Dmytro Omelchenko[1] | Arnold Roger Bitja-Nyom[2,3] | Michael Matschiner[4] | Milan Malinsky[5,6] | Adrian Indermaur[5] | Walter Salzburger[5] | Oldřich Bartoš[7] | Zuzana Musilova[1]

[1]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic | [2]Department of Aquatic Ecosystems Management, Institute of Fisheries and Aquatic Sciences, University of Douala, Douala, Cameroon | [3]Department of Biological Sciences, University of Ngaoundéré, Ngaoundéré, Cameroon | [4]Natural History Museum, University of Oslo, Oslo, Norway | [5]Zoological Institute, University of Basel, Basel, Switzerland | [6]Institute of Ecology and Evolution, Bern, Switzerland | [7]Military Health Institute, Military Medical Agency, Prague, Czech Republic

Correspondence: Dmytro Omelchenko (omeldima@gmail.com) | Zuzana Musilova (zuzana.musilova@natur.cuni.cz)

Received: 29 April 2024 | **Revised:** 26 August 2024 | **Accepted:** 29 August 2024

Handling Editor: J A H Benzie

Funding: This work was supported by Czech Science Foundation (21-31712S) and Charles University Grant Agency (GAUK 1556119).

Keywords: cichlids | duplication | evolution | haemoglobin | synteny | teleosts

ABSTRACT

Haemoglobin is a key molecule for oxygen transport in vertebrates. It exhibits remarkable gene diversity in teleost fishes, reflecting adaptation to various aquatic environments. In this study, we present the dynamic evolution of haemoglobin subunit genes based on a comparison of high-quality genome assemblies of 24 vertebrate species, including 17 teleosts (of which six are cichlids). Our findings indicate that teleost genomes contain a range of haemoglobin genes, from as few as five in fugu to as many as 43 in salmon, with the latter being the largest repertoire found in vertebrates. We find evidence that the teleost ancestor had at least four Hbα and three or four Hbβ subunit genes, and that the current gene diversity emerged during teleost radiation, driven primarily by (tandem) gene duplications, genome compaction, and rearrangement dynamics. We provide insights into the genomic organisation of haemoglobin clusters in different teleost species. We further show that the evolution of paralogous rhbdf1 genes flanking both teleost clusters (LA and MN) supports the hypothesis for the origin of the LA cluster by rearrangement within teleosts, rather than by the teleost specific whole-genome duplication. We specifically focus on cichlid fishes, where adaptation to low oxygen environment plays role in species diversification. Our analysis of six cichlid genomes, including *Pungu maclareni* from the Barombi Mbo crater lake, for which we sequenced a representative genome, reveals 18–32 copies of the Hb genes, and elevated rates of non-synonymous substitutions compared to other teleosts. Overall, this work facilitates a deeper understanding of how haemoglobin genes contribute to the adaptive potential of teleosts.

1 | Introduction

Haemoglobin is an oxygen-binding metalloprotein responsible for oxygen transport in vertebrates (Storz 2018). In jawed vertebrates (gnathostomes), the haemoglobin molecule has a quaternary structure and consists of four globin polypeptide chains (two Hbα and two Hbβ subunits) attached to the prosthetic heme group (Wells 1999). This tetrameric structure is considered to be of great functional importance since it provides a mechanism for cooperative oxygen-binding and allosteric regulatory control (Storz, Opazo, and Hoffmann 2013). Vertebrates have multiple haemoglobin genes and are known for a developmental switch between juvenile and adult haemoglobins, resulting in most vertebrates expressing different genes during ontogenesis (Chan

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

et al. 1997). This differential expression of distinct haemoglobin isoforms helps to deal with the changing oxygen transport challenges encountered during vertebrate development (Opazo et al. 2013).

Vertebrates have diversified following two rounds of whole-genome duplications in their ancestor (Ohno 1970), and teleost fishes have undergone an additional teleost-specific genome duplication (Kuraku and Meyer 2009 and ref. therein). The evolutionary success of both vertebrates and teleosts has often been associated with these whole-genome duplications thanks to the sufficient genetic “substrate” for selection to act upon (Glasauer and Neuhauss 2014). For example, the Hox gene family (Holland et al. 1994), MHC genes (Abi-Rached et al. 2002) or parvalbumin genes (Mukherjee et al. 2021) still carry the traces of the whole-genome duplications. In general, many gene families have been shaped by the interplay between whole-genome duplications, ancestral or lineage-specific (tandem) gene duplications, and prevalent gene losses (Hughes and Friedman 2003; Rennison, Owens, and Taylor 2012; Steinke et al. 2006; Parey et al. 2022), and that includes also teleost haemoglobins (Quinn et al. 2010; Opazo et al. 2013; Storz 2016).

Haemoglobin genes (Hb) originated together with other globin genes (myoglobin, cytoglobin) after two rounds of whole-genome duplications in the vertebrate ancestor (Hoffmann, Opazo, and Storz 2012). Subsequent subfunctionalization of gas transport and oxygen storage between haemoglobin and myoglobin is sometimes seen as a key innovation of early vertebrates (Storz, Opazo, and Hoffmann 2013). The Hb subunit genes—proto-Hbα and Hbβ have afterwards resulted from an ancient tandem duplication in early jawed vertebrates (Opazo et al. 2013). Haemoglobin subunit genes are found in multi-copy tandem repeats and the ancestral genomic arrangement was probably a single cluster with Hbα and Hbβ subunit genes mixed, as still seen in cartilaginous fish (Marino et al. 2007).

Teleosts have haemoglobins organised in two genomic clusters (labelled as MN and LA—the names of the clusters are acronyms from the names of two flanking genes located upstream to the cluster: lcmt1 and aqp8 in LA, mpg and nprl3 in MN). They both carry mixed Hbα and Hbβ subunits, whereas in tetrapods the Hb genes have been rearranged into a separate Hbα and Hbβ clusters (Hardison 2008; Patel et al. 2008). Homologues of the upstream flanking genes of the teleost MN cluster are also found upstream of the Hbα cluster in tetrapods, whereas the LA cluster seems to be specific for teleosts only (Opazo et al. 2013; Hardison 2012) and it has been suggested that the LA cluster originated from the teleost specific whole-genome duplication (Opazo et al. 2013). Teleost fishes have a greater variation and often a high number of haemoglobin genes compared to air-breathing vertebrates, which can be explained by the fact that variation in oxygen availability is generally much greater in the aquatic environment (Opazo et al. 2013).

In addition to the use of multiple Hb isoforms, haemoglobins diversify and adapt to variation in oxygen availability by sequence evolution. Several studies have shown signatures of positive selection on the amino acid sites in particular Hb gene copies. Specifically, an excess of non-synonymous substitutions in nucleotide sequence of haemoglobins is found in species from extreme environments, such as in lagomorphs and rodents from Qinghai-Tibet plateau (Chen et al. 2016), or two cetacean species (Atlantic bottlenose dolphin and killer whale) where it was hypothesized that it could lead to an adaptation to their prolonged dives (Nery, Arroyo, and Opazo 2013). In teleosts, it has been shown that high-altitude schizothoracine fishes from Qinghai-Tibet plateau have functionally divergent haemoglobin genes with positively selected sites (Lei et al. 2021).

Some teleost families diversify more and occupy a much broader range of niches than others. Cichlid fishes represent one of the most species-rich families of vertebrates, known for their large and rapid adaptive radiations, especially in the tropical lakes of Africa (Seehausen, 2000; Kautt, Machado-Schiaffino, and Meyer 2016; Svardal, Salzburger, and Malinsky 2021), and serve as a model group for research into adaption and diversification (e.g., Musilova et al. 2019; Ronco et al. 2021). Based on five initial short-read genome assemblies, it has been shown that cichlids have elevated numbers of gene duplicates (Brawand et al. 2014), but the particular case of haemoglobin genes has not been investigated. Warm tropical lakes, including the lakes hosting cichlid fish radiations, tend to be permanently stratified with steep oxygen gradients. Nevertheless, deep waters with low oxygen concentrations have repeatedly been colonised by cichlids, and depth adaptation represents one of the most important axes of differentiation in multiple lakes, e.g., Lake Malawi (Malinsky et al. 2018), Lake Masoko (Malinsky et al. 2015), Lake Barombi Mbo (Green, Corbet, and Betney 1973; Musilova et al. 2019), or Lake Tanganyika (Ronco et al. 2021). It has been shown that deep-water adapted cichlid species from Barombi Mbo have highly elevated blood haemoglobin concentrations (Green, Corbet, and Betney 1973) and that six haemoglobin genes contain signs of parallel adaptation in the Lake Malawi deep-water cichlids groups *Diplotaxodon* and ‘deep-benthic’ (Hahn et al. 2017; Malinsky et al. 2018). However, a more comprehensive analysis of selection on haemoglobin genes across multiple cichlid adaptive radiations is lacking.

Haemoglobin gene clusters are highly repetitive, and a correct understanding of their evolutionary dynamics requires accurate genome assemblies. Here we reconstruct the evolution of the haemoglobin subunit genes based on high-quality long-read assemblies of 24 vertebrate species, including 17 teleosts, six of which belong to the cichlid family. Compared with previous analyses, that were based primarily on short-read genome assemblies (e.g., Opazo et al. 2013), we find much expanded Hb repertoires, driven primarily by (tandem) gene duplications. Our reconstruction of Hb cluster evolution reveals dynamics of duplication, rearrangements, and genome compaction (i.e., loss of the genomic elements including paralogous genes), and provides evidence that the teleost LA and MN clusters originated after the teleost-specific genome duplication. We used the comprehensive collection of Hb genes from all 24 genomes to conduct a search for signatures of positive selection on the sequence level, revealing a 27 Hbα and 17 Hbβ genes under positive selection within the cichlid family. These findings contribute valuable insights into the dynamic evolution of the haemoglobin genes in teleosts. Additionally, they establish a foundation for future research into specific haemoglobin and oxygen transport adaptations, such as those found in cichlid species inhabiting hypoxic deep-water lakes (e.g., Barombi Mbo).

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | Material and Methods

2.1 | Sampling and DNA Extraction

In total, we used 24 different species in this study (full list with GenBank accession number in Table 1). For one species, *Pungu maclareni*, we provide the representative genome. Live specimens of *P. maclareni* were collected in the Barombi Mbo crater lake (4°39′41″ N 9°24′9″° E/4.66139° N 9.40250° E) and transported to the aquarium facility at Charles University in Prague in March 2016 (collecting permit: 0000002/MINRESI/B00/C00/C10/C12). One male individual of *Pungu maclareni* was euthanized and liver and spleen tissues were dissected for the DNA extraction. To obtain high-molecular-weight (HMW) DNA, we homogenised the tissues under liquid nitrogen followed by a standard phenol-chloroform DNA extraction (e.g., McKiernan and Danielson 2017). DNA was then quantified using a Qubit fluorometer (Invitrogen) and checked for degradation by standard electrophoresis.

2.2 | Nanopore Sequencing and Assembly

Libraries for Oxford Nanopore Technologies (ONT) sequencing were prepared from a HMW DNA using a Ligation Sequencing Kit (SQK-LSK109). Libraries were sequenced on the ONT GridIONx5 platform using the R9.4.1 chemistry (Flow-Cell). Sequencing data were base-called, i.e. transmission from physical changes in the electric current signal measured by the ONT sequencing device to biologically relevant bases, using Guppy v3.2.10 in "high-quality" settings/mode (e.g., Wick, Judd, and Holt 2019).

Basic characteristics of the ONT libraries were estimated in R 3.6.0 using the NanoR library (R Core Team 2018; Bolognini et al. 2019). In summary, we acquired more than 5 million reads with an average length of over 7000 bp, which in total slightly exceeded 40 Gbp (only reads > 1000 bp were considered; see Table 2). The genome was assembled using the Flye 2.7.1 long-read assembler (Kolmogorov et al. 2019). The primary assembly was then polished with the ONT long-reads alone using the Medaka 1.0.1 software tool (Lee et al. 2021). The assembly was further iteratively (four times) polished by Whole Genome Sequencing (WGS) Illumina reads (GenBank accession number: SRX7645636) using the Pilon software tool (Walker et al. 2014). The quality (completeness) of the assembly was assessed using BUSCO 4.0.6 (actinopterygii_odb10; see Table 3; Simão et al. 2015). The raw Nanopore reads and the resulting genome assembly sequences were deposited into NCBI GenBank repositories (accession number: PRJNA1153940, GCA_041757325.1).

2.3 | Gene Mining and Gene Annotation

We used the generated genomic data (*Pungu maclareni*) complemented by high-quality assemblies from GenBank (see Table 1 for accession numbers) for a set of selected species. Our focus was on a comparative analysis of the haemoglobin subunit genes repertoire and its genomic organisation. In total, we analysed 24 genomes, 17 of teleosts and seven of other vertebrates (see Table 1). We cover all four main teleost lineages (i.e., Elopomorpha, Osteoglossomorpha, Otomorpha and Euteleostei) and specifically focus on cichlids. We used the following bioinformatic tools to mine for haemoglobin genes. Geneious software ver. 9.1 with High sensitivity setting (http://www.geneious.com, Kearse et al. 2012) was used to capture scaffolds/chromosomes carrying Hb genes by mapping the assembled genome against the zebrafish reference composed of one Hbα (NM_131257.3) and one Hbβ gene exons (BC139602.1). If an annotation was available, we extracted all haemoglobin genes from the respective genome. We subsequently manually mapped the single exons of all extracted genes against flanking regions of the identified cluster (i.e., the region between the first Hb gene in the cluster and the first upstream flanking gene, and, similarly, between the last Hb gene and the first downstream gene), and also to the intergenic regions to reveal the entire Hb gene cluster and confirmed the annotation using the Geneious software. In some cases, this approach served to improve the annotation. Furthermore, we performed AUGUSTUS ab initio gene prediction (Stanke and Morgenstern 2005) to reveal haemoglobin genes in the unannotated eel genome (*Anguilla anguilla*) with a Nanopore assembly (accession number PRJEB20018 in the European Nucleotide Archive). We inspected the *Pungu maclareni* genome and we have manually annotated the found Hbα and Hbβ genes.

2.4 | Synteny Analysis

We manually inspected genomic regions that contain haemoglobin subunits genes (both MN and LA clusters) in five cichlid species with publicly available genome assembly: Nile tilapia (*Oreochromis niloticus*), Eastern happy (*Astatotilapia calliptera*) and zebra mbuna (*Maylandia zebra*), which represent African cichlids from rivers and the Lake Malawi, respectively, together with two Neotropical species, Flier cichlid (*Archocentrus centrarchus*), and Midas cichlid (*Amphilophus citrinellus*). Furthermore, we similarly identified haemoglobin clusters in *Pungu maclareni*, the newly sequenced species of the Barombi Mbo radiation, based on our manual gene annotation. Genomic sequences of haemoglobin clusters were extracted from the linkage groups including the whole sequence between flanking genes (LA cluster upstream: lcmt1—aqp8—aqp8, downstream: rhbdf1—foxj1b—ubald1; MN cluster upstream—rhbdf1—mpg—nprl3, downstream: kank2—dock6/7—elav13; Figure 1) and reoriented for further synteny comparison by the aqp8 and nprl3 genes upstream to the haemoglobin subunits in the LA and MN cluster, respectively. We subsequently used the Genome Pair Rapid Dotter tool called Gepard 1.40 (Krumsiek, Arnold, and Rattei 2007) to create comparative dotplots of the haemoglobin clusters and identified traces of gene duplications or inversions.

2.5 | Phylogenetic Analysis of the Haemoglobin Genes and the *rhbdf1* Flanking Gene

To complement the teleost data set, we included also the haemoglobin clusters of six non-teleost fishes and tetrapods (Table 1; Figure 1). Coding sequences of the haemoglobin genes were extracted and used for the construction of a combined gene tree for both Hbα and Hbβ subunits. Cytoglobins from Australian ghost shark, spotted gar and Nile tilapia were used as an outgroup. We

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Table of genomes used for haemoglobin gene mining in 24 vertebrates, including 17 teleosts.

Species	English name	Assembly acc. Num	Assembly name	Hb cluster	Genomic scaffold	Chromosome	Upstream 3	Upstream 2	Upstream 1	Start	Number of haemoglobin genes	End	Downstream 1	Downstream 2	Downstream 3
Callorhinchus milii	Ghost shark	GCA_018977255.1	IMCB_Cmil_1.0	NA	NW_024704757		foxj1b	aanat2	rhbdf1	14,662,718	2	14,623,747	cygb	luc7l	fam234a
Xenopus tropicalis	Clawed frog	GCA_000004195.4	UCB_Xtro_10.0	NA	NC_030685	9	fam234a	luc7l	gby	40,992,113	7	40,846,078	kctd18 + rhbdf1	mpg	nprl3
				MN	NC_030685	9	kctd18 + rhbdf1	mpg	nprl3	40,568,240	9	40,353,331	polr3k	snrnp25	gtfc3c1
Taeniopygia guttata	Zebrafinch	GCA_003957565.4	bTaeGut1.4.pri	DS	NC_044211	1	unchar	o51g2	o51e2	113,747,431	4	113,791,487	fol3	olf52b2	rps11
				MN	NC_044226.2	14	rhbdf1	mpg	nprl3	13,440,000	3	13,461,710	pgap6	mrpl28	axin1
Mus musculus	Mouse	GCA_000001635.9	GRCm39	DS	NC_000073	7	olfr69	olfr68	olfr67	103,441,038	5	103,530,512	olfr66	olfr64	olfr65
				MN	NC_000077	11	rhbdf1	mpg	nprl3	32,217,708	5	32,254,948	sh3pxd2b	ubtd2	efcab9
Homo sapiens	Human	GCA_000001405.29	GRCh38.p14	DS	NC_000011	11	OR52A5	OR52A1	OR51V1	5,200,683	5	5,291,466	OR51B4	OR51B2	OR51B5
				MN	NC_000016.10	16	RHBDF1	MPG	NPRL3	138,674	5	188,989	LUC7L	FAM234A	RGS11
Pan paniscus	Bonobo chimp	GCA_000258655.2	panpan1.1	DS	NC_027879	11	OR52A1	OR52I2	OR51V1	5,018,328	5	5,117,924	OR51B4	OR51B2	OR51B5
				MN	NC_027884	16	RHBDF1	MPG	NPRL3	129,221	5	177,856	LUC7L	FAM234A	RGS11
Lepisosteus oculatus	Spotted gar	GCA_000242695.1	LepOcu1	MN	NC_023191	13	rhbdf1	mpg	nprl3	14,771,739	8	14,827,298	cygb	luc7l	fam234a
Anguilla anguilla	European eel	GCA_013347855.1	fAngAng1	LA	NC_049202.1	2	lcmt1	aqp8	aqp8	51,695,718	1	51,647,385	rhbdf1	aanat2	foxj1b
				MN	NC_049217	17	rhbdf1	mpg	nprl3	11,697,720	8	11,678,314	aqp8	aqp8	arhgrap17b
Paramormyrops kingsleyae	Elephant fish	GCA_002872115.1	PKINGS_0.1	LA	NW_019716143		grid2ip	aghrap17	lcmt1	1,317,361	6	1,422,149	rhbdf1	fkbp10	noxo1b
				MN	NW_019712565		rhbdf1	mpg	nprl3	4,318,013	5	4,347,095	aqp8	aqp8	aghrap17b
Scleropages formosus	Asian arowana	GCA_900964985.2	fSclFor1.1	LA	NC_041825	20	tex47	aghrap17b	lcmt1	12,329,717	9	12,240,083	rhbdf1	foxj1b	foxj1b
				MN	NC_041813	8	rhbdf1	mpg	nprl3	15,039,498	7	15,118,265	aqp8	aqp8	arhgrap17b
Clupea harengus	Atlantic herring	GCA_900700415.2	Ch_v2.0.2	LA	NC_045174	23	lcmt1	aqp8	aqp8	10,999,422	4	11,029,948	rhbdf1	foxj1b	mgrn1a
				MN	NC_045152	1	unchar	wipf2b	nprl3	29,854,559	12	29,916,204	kank2	dock7	elavl3
Danio rerio	Zebrafish	GCA_000002035.4	GRCz11	LA	NC_007123	12	arhgap17a	lcmt1	aqp8	20,355,138	4	20,303,433	rhbdf1	foxj1b	ubald1a
				MN	NC_007114	3	rhbdf1	mpg	nprl3	55,079,750	13	55,161,626	kank2	dock6	magfa
Esox lucius	Northern pike	GCA_011004845.1	fEsoLuc1	LA	NC_047573	5	aghrap17a	lcmt1	aqp8	17,843,032	1	17,824,101	rhbdf1	foxj1b	ubald1
				MN	NC_047579	11	rhbdf1	mpg	nprl3	46,448,666	26	46,624,438	kank2	dock7	elavl3
				MN	NC_047579	11	dnmt1	eif3g	p2ry11	47,051,188	10	47,145,485	p2ry11	ppan	angptl6

(Continues)

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | (Continued)

Species	English name	Assembly acc. Num	Assembly name	Hb cluster	Genomic scaffold	Chromosome	Upstream 3	Upstream 2	Upstream 1	Start	Number of haemoglobin genes	End	Downstream 1	Downstream 2	Downstream 3
Salmo salar	Atlantic salmon	GCA_905237065.2	Ssal_v3.1	LA1	NC_059469	28	lcmt1	aqp8	aqp8	15,149,397	3	15,111,249	rhbdf1	foxj1b	ubald1a
				LA2	NC_059442	1	aghrap17b	aqp8	aqp8	89,407,904	1	89,339,065	rhbl4	foxj1b	ubald1a
				MN1	NC_059447	6	rhbdf1	mpg	nprl3	37,696,574	22	37,545,604	kank2	dock6	elavl3
				MN2	NC_059444	3	rhbdf1	mpg	nprl3	64,048,038	16	64,211,380	kank2	dock7	elavl3
Gadus morhua	Atlantic cod	GCA_902167395.1	gadMor3.0	LA	NC_044065	18	lcmt1	aqp8	aqp8	12,989,901	5	12,964,384	rhbdf1	foxj1b	ubald1a
				MN	NC_044049	2	rhbdf1	mpg	nprl3	8,596,052	4	8,586,401	kank2	dock7	elavl3
Takifugu rubripes	Fugu	GCA_901000725.2	fTakRub1.2	LA	NC_042285.	1	grid2ip	aghrap17b	lcmt1	18,867,272	3	18,877,285	rhbdf1	foxj1b	ubald1a
				MN	NC_042289	5	rhbdf1	mpg	nprl3	4,101,781	2	4,108,121	kank2	dock7	elavl3
Oryzias latipes	Medaka	GCA_002234675.1	ASM223467v1	LA	NC_019877	19	aghrap17b	lcmt1	aqp8	22,442,584	3	22,454,916	rhbdf1	yeats2	unchar
				MN	NC_019866	8	rhbdf1	mpg	nprl3	8,207,525	11	8,243,375	kank2	dock7	elavl3
Nothobranchius furzeri	Killifish	GCA_027789165.1	Nfu_20140520	LA	NC_029660	12	aghrap17b	lcmt1	aqp8	35,832,040	2	35,824,547	ZNF146	zinc finger protein 664-like	nuf2
				MN	NC_029653	5	rhbdf1	mpg	nprl3	66,571,704	12	66,628,604	kank2	dock7	elavl3
Amphilophus citrinellus	Midas cichlid	GCA_013435755.1	fMPI-2012	LA	contig_243	NA	lcmt1	aqp8	aqp8	5,484,908	3	5,500,504	rhbdf1	foxj1b	ubald1a
				MN	contig_728	NA	rhbdf1	mpg	nprl3	434,307	15	379,542	kank2	dock7	elavl3
Archocentrus centrarchus	Flier cihlid	GCA_007364275.2	fArcCen1	LA	NC_044364	19	lcmt1	aqp8	aqp8	7,822,921	3	7,848,866	rhbdf1	foxj1b	ubald1a
				MN	NC_044353	8	rhbdf1	mpg	nprl3	6,139,626	20	6,206,746	kank2	dock7	elavl3
Oreochromis niloticus	Nile tilapia	GCA_001858045.3	O_niloticus_UMD	LA	NC_031973	8	lcmt1	aqp8	aqp8	21,334,461	3	21,353,715	rhbdf1	foxj1b	ubald1a
				MN	NC_031969	4	rhbdf1	mpg	nprl3	24,434,332	29	24,348,606	kank2	dock7	elavl3
Maylandia zebra	Zebra mbuna	GCA_000238955.5	M_zebra_UMD2a	LA	NC_036787	8	lcmt1	aqp8	aqp8	17,452,769	3	17,414,339	rhbdf1	foxj1b	ubald1a
				MN	NC_036783	4	rhbdf1	mpg	nprl3	10,548,069	18	10,630,330	kank2	dock7	elavl3
Astatotilapia calliptera	Eastern happy	GCA_900246225.3	fAstCal1.2	LA	NC_039309	8	lcmt1	aqp8	aqp8	16,527,459	3	16,548,723	rhbdf1	foxj1b	ubald1a
				MN	NC_039305	4	rhbdf1	mpg	nprl3	6,518,619	23	6,497,049	kank2	dock7	elavl3
Pungu maclareni	Pungu	GCA_041757325.1	ASM4175732v1	LA	Contig_1931	NA	lcmt1	aqp8	aqp8	2,291,871	3	2,306,248	rhbdf1	foxj1b	ubald1a
				MN	Contig_791	NA	rhbdf1	mpg	nprl3	1,776,322	19	1,720,355	kank2	dock7	elavl3

Note: Table contains genomic coordinates, names of three flanking genes upstream and downstream of the haemoglobin clusters and the exact number of haemoglobin genes in each cluster.

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Basic overall statistics of sequenced ONT library.

Mean length [bp]	7231.8
Median length [bp]	6173
N50 [bp]	8743
N10 [bp]	28,037
N25 [bp]	13,004
N75 [bp]	6187
Number of reads	5,571,289
Bases [Gbp]	40.291

Note: Only reads > 1000 bp were included in analysis.

TABLE 3 | Results of BUSCO analysis.

Complete BUSCOs (C)	3567
Complete and single-copy BUSCOs (S)	3533
Complete and duplicated BUSCOs (D)	34
Fragmented BUSCOs (F)	7
Missing BUSCOs (M)	66
Total BUSCO groups searched	3640

also performed the additional *rhbdf1* gene analysis on the selected species with one (human, mouse, zebrafinch, clawed frog) and two paralogs (eel, elephant fish, zebrafish, pike, salmon, cod, medaka, tilapia) and using the ghost shark as an outgroup. The nucleotide sequences of both data sets were aligned with MAFFT ver. 7.222 (Katoh and Standley 2013). The best fitting substitution model (GTR + I + G) was selected by jModeltest2 ver. 2.1.6 (Darriba et al. 2012). We performed Bayesian inference (for Hb genes and rhbdf1 gene) and Maximum Likelihood (only for Hb genes) phylogenetic analysis. Phylogenetic analyses were performed on the CIPRES Science Gateway portal (Miller, Pfeiffer, and Schwartz 2010) using MrBayes ver. 3.2.6 (Ronquist et al. 2012) for 20 million generations and 12.5% burnin. We used IQ-TREE (Nguyen et al. 2014) to infer phylogenetic trees by maximum likelihood using the SH-aLRT test with substitution model selection by ModelFinder (Kalyaanamoorthy et al. 2017) and the ultrafast bootstrap approximation (Hoang et al. 2017) with 1000 replicates.

2.6 | Positive Selection Test

Codon alignments were used for the selection test using CODEML package as a part of the PAML software (Yang 2007). Hb genes that contained in-frame stop codons were removed from the analysis. We performed a codon alignment in the PAL2NAL software (Suyama, Torrents, and Bork 2006). We ran analyses separately for the Hbα and Hbβ subunits to reduce computational time. We used a branch-site model with heterogeneous omega ratios among sites allowed and estimated the ratio between the non-synonymous and synonymous substitutions for background branches. We subsequently produced final trees with branches coloured accordingly to the dN/dS ratio using the *treeio* R package (Wang et al. 2020).

3 | Results

3.1 | Haemoglobin Gene Repertoire and Cluster Synteny

We analysed haemoglobin genes in the high-quality genomes of 24 vertebrate species, one cartilaginous fish, five tetrapods, one non-teleost and 17 teleosts, of which six were cichlids (Figure 1). Our cichlid collection included both African and Neotropical cichlids, and to complement the existing cichlid data set, we sequenced the representative genome of *Pungu maclareni*, an endemic cichlid species from crater lake Barombi Mbo (Cameroon, Africa) through the Oxford Nanopore technology. The assembled genome with 2,443 scaffolds (> 1000 bp) has been deposited on GenBank (acc. no.: *tba*) and has a N50 of 4,250,295 bp (and N90 = 516,595 bp) with both haemoglobin clusters recovered intact on separate scaffolds.

Our results show that teleost fishes generally have a larger haemoglobin gene repertoire compared to other vertebrates. Teleosts possess between 5 and 43 haemoglobin genes in their genome (Figure 1). The Japanese pufferfish, fugu (*Takifugu rubripes*), has the lowest number of haemoglobin genes (four Hbα subunit genes and one Hbβ subunit gene; Figure 1), whereas the Atlantic salmon has the highest number of haemoglobin subunit genes found among teleosts or gnathostomes (43 Hb genes in total; 20 of Hbα and 23 of Hbβ subunit) followed by the northern pike (*Esox lucius*) with 38 Hb genes (18 + 20). The Nile tilapia (*Oreochromis niloticus*), with the repertoire of 32 haemoglobin subunit genes (17 Hbα and 15 Hbβ subunits) has the highest number of haemoglobin genes among cichlids.

Most of the analysed species possess two haemoglobin clusters, with the exception of the spotted gar (*Lepisosteus oculatus*) and the Australian ghostshark (*Callorhinchus milii*), which both have a single haemoglobin cluster (Figure 1). In addition, the northern pike (*Esox lucius*) and the Atlantic salmon (*Salmo salar*) with three and four haemoglobin clusters, respectively, demonstrate a variability in genomic organisation of haemoglobin genes even within a single evolutionary lineage (Figure 1). The flanking genes upstream and downstream of each cluster further help to shed light on evolution of the whole paralogons. We found that synteny of the haemoglobin clusters and flanking genes is highly conserved across major evolutionary groups of teleosts, yet with certain deviations (Figure 1).

3.2 | Evolutionary History of the Haemoglobin Subunit Genes

In total, we identified and extracted the sequences of 197 haemoglobin alpha (Hbα) subunit genes and 185 haemoglobin beta (Hbβ) subunit genes from the 24 analysed species (see Table 1). We reconstructed the gene trees (Figures 2, S1 and S2) and identified two major clades confirming the Hbα and Hbβ subunits in both approaches. For both Hbα and Hbβ subunit genes, the phylogenetic signal revealed several

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

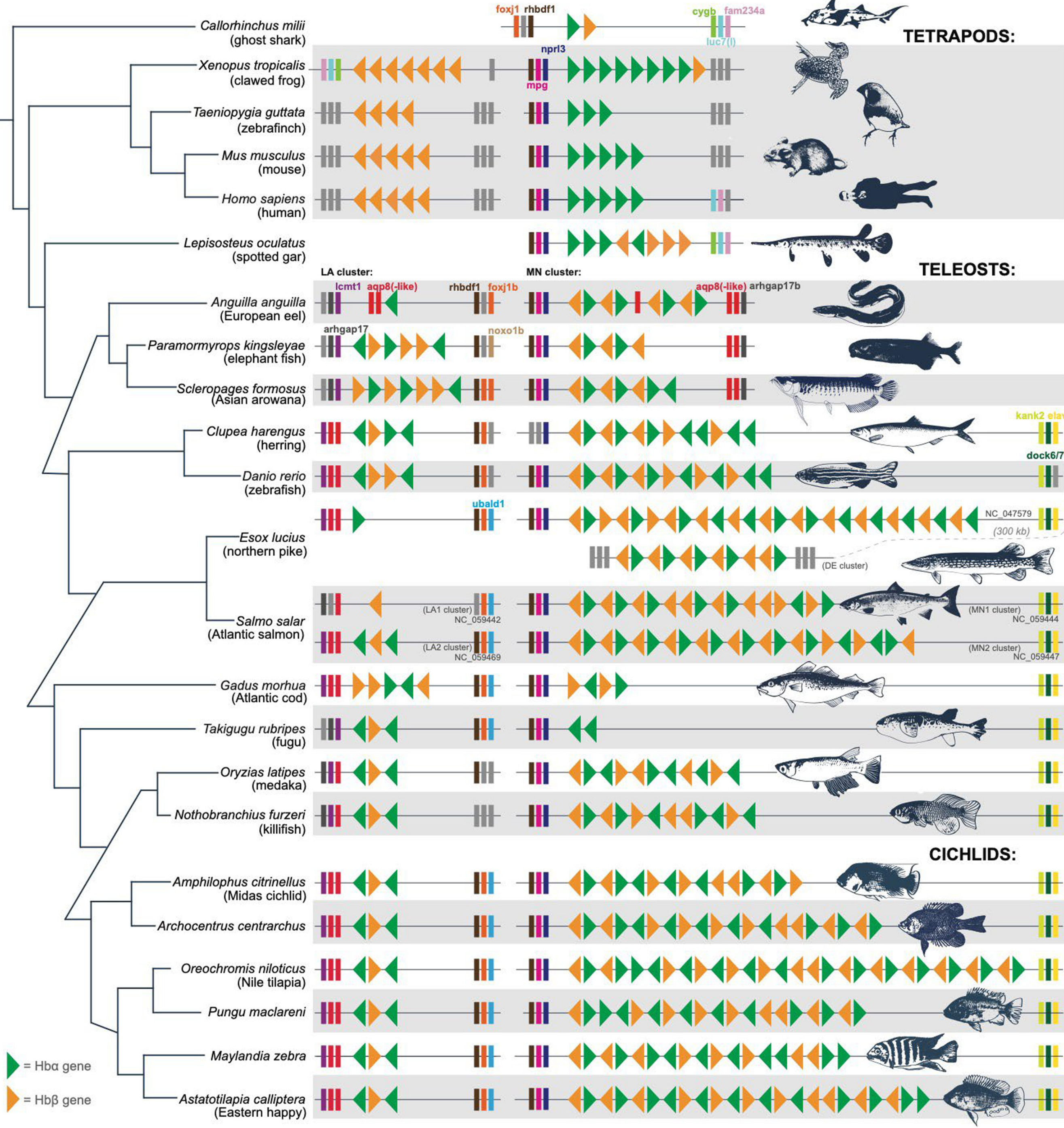

FIGURE 1 | Conserved synteny and genomic organisation of the haemoglobin clusters in major linages of gnathostomes, teleosts and cichlids. Haemoglobin genes shown as triangles with the forward or reverse orientation, Hbα subunit genes shown in green, Hbβ subunit genes shown in orange. Flanking genes shown as rectangles either having a corresponding colour code (for the genes frequent in different evolutionary groups) or in grey if unique. We present the clusters in the same orientation for comparative purposes. The MN cluster refers to a larger and more variable cluster with the mpg and nprl3 upstream flanking genes, whereas the LA cluster is more conserved and named after lcmt1 and aqp-8 flanking genes. Note that in Elopomorpha (*Anguilla*) and Osteoglossomorpha (*Paramormyrops*, *Scleropages*) the MN cluster has sometimes been referred to as NA cluster due to the presence of the aquaporin (aqp8) genes. The tree topology is based on relationships after Betancur-R et al. (2017). Data for the Bonobo chimpanzee genome (acc. no. GCA_000258655.2) not shown because the clusters and flanking genes are identical to human. For more details on the genomes, coordinates and accession numbers, refer to Table 1.

independent gene clades that have diversified in the teleost ancestor. Tetrapod genes correspondingly branch at the base of the tree, whereas teleost Hbα and Hbβ subunits are organised in four and three-to-four main clades each, respectively. The clades partially correspond to the genomic location either in LA or MN cluster, but not for all genes. The overall

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

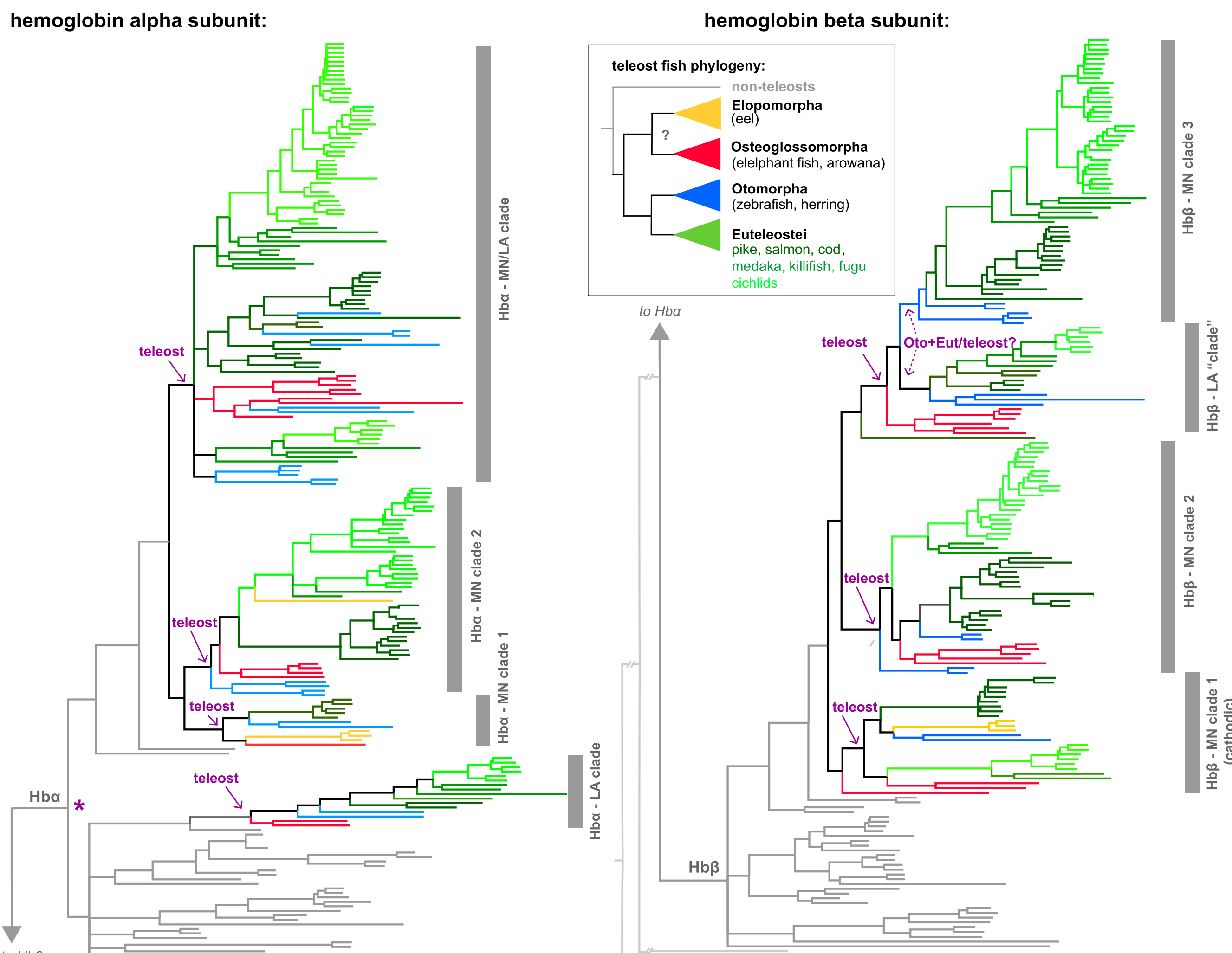

FIGURE 2 | Gene trees of the Hbα and Hbβ haemoglobins subunits reconstructed by Bayesian analysis. Four main teleost lineages are highlighted by different colours (see inset for the teleost phylogeny) or black to elucidate Hb gene diversification in the teleost ancestor, as well as after the teleost radiation. Non-teleost fish and tetrapod Hb genes are coloured as grey. The tree topology suggests that the teleost ancestor had at least four Hbα genes and three (or four) Hbβ genes before the teleost diversification corresponding to the recovered clades marked and labelled by purple arrow. Note that the duplication involving the LA clade Hbα genes and the rest Hbα genes possibly happened before teleost diversification (marked by a star symbol), hence teleost haemoglobins are found in two different clades of the tree. For the full-size gene trees with the copy names refer to Figures S1 and S2.

phylogenetic signal is more robust for the Hbβ subunit genes, whereas less clear for the Hbα subunit gene tree, which shows some polytomy and unresolved topology within one of the clades with mixed MN/LA genes (Figures 2, S1 and S2). One of the clades within the Hbβ subunits is formed predominantly by the so-called cathodic subunits, the highly conserved teleost haemoglobins whose name refers to their position on the isoelectric focusing gel. Overall, the observed phylogenetic pattern suggests a presence of at least four putative copies of Hbα subunits and three or four Hbβ subunits in the ancestor of teleost fishes (Figure 2).

3.3 | Comparative Genomics of Haemoglobin Clusters in Cichlids

We focused on the genomic architecture of the haemoglobin clusters in six cichlid species. Five of the high-quality genomes produced with the PacBio sequencing technology (*A. citrinellus*, *A. centrarchus*, *A. calliptera*, *M. zebra*, *O. niloticus*) have been complemented by one species sequenced by the Oxford Nanopore (*Pungu maclareni*; this study). All cichlid species have two haemoglobin clusters (LA and MN). The LA cluster is conserved in all cichlid species and contains 3 haemoglobin genes in the same orientation flanked with lcmt1 + aqp8 + aqp8 upstream of the cluster and rhbdf1b + foxj1b + ubald1a in the downstream region (Figure 1). The MN cluster differs in number and orientation of the haemoglobin genes—from 15 haemoglobin subunit genes in *Amphilophus citrinellus*, 18 in *Maylandia zebra*, 19 in *Pungu maclareni*, 20 in *Archocentrus centrarchus*, 23 in *Astatotilapia calliptera* to, finally, *Oreochromis niloticus* with 29 haemoglobin genes, the species with the highest number so far identified among cichlids (Figure 1). The cichlid comparison revealed that all species carry signals of gene duplications mainly within the two gene-rich regions of the MN cluster. The overall synteny in cichlids seems to be conserved among species, yet

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

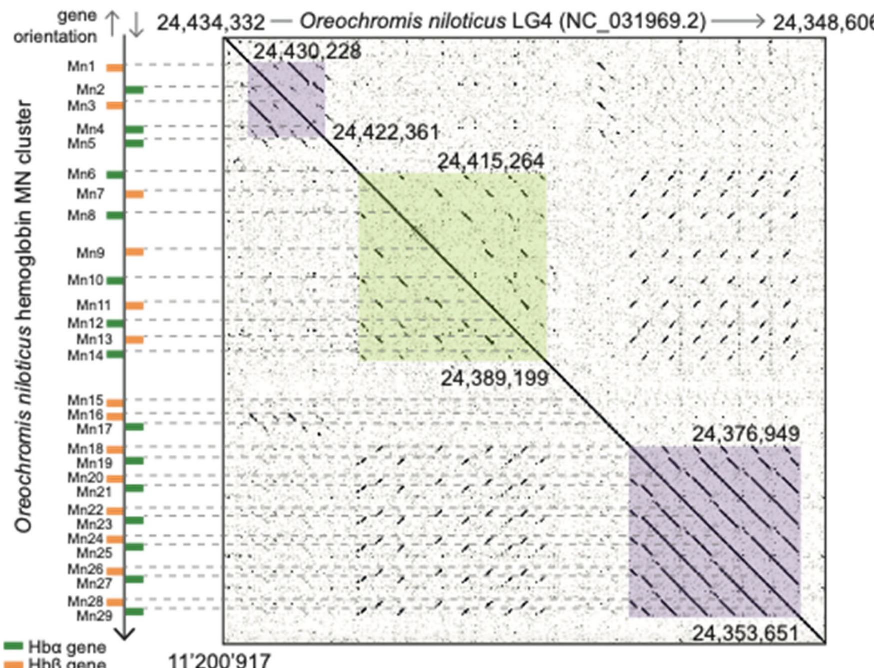

FIGURE 3 | Genome pair dotplot (Gepard) of the haemoglobin MN cluster of Nile tilapia (*Oreochromis niloticus*). The haemoglobin MN cluster (located at chromosome LG4) is plotted against itself to reveal high-similarity regions (i.e., lines formed by black dots). Synteny of the haemoglobin genes in the cluster is shown on the left. Hbα subunit genes are coloured green, whereas Hbβ subunits are marked in orange. Horizontal lines represent the first (for forward orientation) or the last (for reverse) codon of each haemoglobin gene. The duplication region (positions 24,415,264–24,389,199) marked by a green rectangle is evolutionary older and is shared among African cichlids. For other species, see Figures 5 and S4. The two regions (positions 24,430,228–24,422,361 and 24,376,949–24,353,651) marked by a purple rectangle are a result of more recent duplication events (specific for tilapia or an oreochromine ancestor; Figure 4). The suggested evolutionary scenario of haemoglobin gene evolution is presented in Figure 4. Genomic coordinates for the LG4 (NCBI accession number NC_031969.2).

there are differences suggesting ongoing genome rearrangements, as African and Neotropical cichlids differ slightly in the composition of their clusters (Figures 5 and S4). By similarity comparison of the MN gene cluster sequence (plotted against itself), we searched for signs of the major duplications, i.e. involving multiple genes. In Nile tilapia (*O. niloticus*), there are signs of three such duplicated regions, one of an older evolutionary origin (shared with other African species), and two more recent located in the beginning (four Hb genes) and at the end of the gene cluster (12 Hb genes; Figures 3 and 4). Pungu (*P. maclareni*), the closest relative of the Nile tilapia, lacks one of the recently duplicated regions and has only two Hbα and two Hbβ genes duplicated in the second region, and its genome carries just a major signal of one older duplicated region typical for African cichlids (Figure 5). In other cichlids (Figures 5 and S4), the pattern also slightly differs. Zebra mbuna (*M. zebra*) has two duplicated regions, but unlike in Nile tilapia, its major recently duplicated region is located in the beginning of the gene cluster. The Eastern happy (*A. calliptera*) has three duplicated regions, of which one is again of the older evolutionary origin, and two are more recently duplicated regions, one located in the beginning of the cluster (with the genes corresponding to the zebra mbuna and Nile tilapia), and one at the end of the gene cluster (with the duplicated genes different from Nile tilapia; based on gene trees in Figures S1 and S2). Two species of Neotropical cichlids do not show signs of the older duplication event typical for the African cichlids, instead Flier cichlid (*A. centrarchus*) rather shows signature of one recent duplication region and the Midas cichlid (*A. citrinellus*) with the lowest number of genes does not express any signs of major duplications (Figures 5 and S4).

3.4 | The Haemoglobin Gene Repertoire in Nile Tilapia as a Model for Haemoglobin Evolution

The Nile tilapia (*Oreochromis niloticus*), similarly to other cichlids, has two haemoglobin clusters, the conserved LA cluster and the MN cluster, the latter of which demonstrates greater evolutionary dynamics. We used the Nile tilapia haemoglobins as a model to reconstruct more detailed evolutionary scenario (Figure 4) and with the subsequent addition of the Hbα and Hbβ gene trees (Figures S1 and S2) we interpret the following sequence of events. The ancestral Hbα and Hbβ subunit genes have diversified in teleosts. The evolutionarily oldest (and putatively functionally most divergent) are the cathodic Hbβ (Mn15 in tilapia) and LA cluster Hbα (La3) subunit genes, which have diversified in the teleost ancestor (or even earlier in case of LA; Figures 2 and 4). Furthermore, the haemoglobin cluster underwent downstream and upstream expansion in gene numbers most likely due to the multiple tandem duplication events that happened in the teleost ancestor. Cichlids seem to further have diversified their haemoglobin gene repertoire namely in the MN cluster with many genes (also see Figures 3 and 5). Nile tilapia also carries traces of a recent gene expansion involving 12 Hb genes, as noticed by the Mn18–29 genes. Being partially shared with *Pungu maclareni* (Figures 5, S1 and S2), these multiplications most likely started in the oreochromine ancestor (i.e., of

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

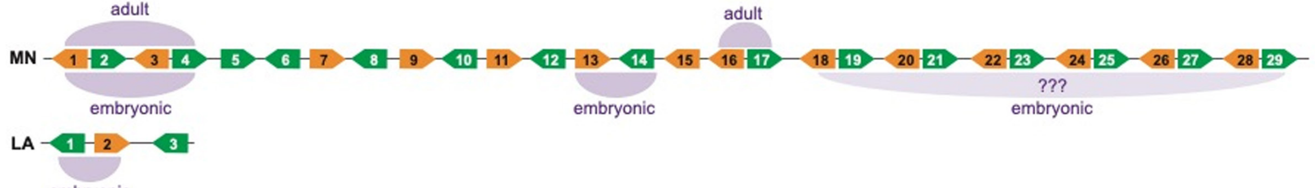

FIGURE 4 | Evolutionary scenario of the Nile tilapia (*Oreochromis niloticus*) haemoglobin gene clusters. (A) The conserved LA cluster is composed of three Hb genes, whereas the more dynamic MN cluster is composed of 29 Hb genes, the highest number among cichlids. Adult and embryonic expression highlighted based on the gills (SAMN05171417, SAMN05171418, SAMN05171419, SAMN03246846, SAMN03246849 and SAMN03246855) and embryo (SAMN00767863) transcriptomes. The ??? symbol marks recently duplicated genes previously referred to as "embryonic"; confirmation on multiple developmental stages is needed. Note that the highest expressed Hb genes in the embryo are located in the LA cluster, while the genes expressed in the adults are found in the MN cluster. (B) Evolutionary scenario showing the series of events as reconstructed and interpreted from (C) and Figures 2, S1, S2, and 3. The arrow symbol is used for gene duplications. The gene numbers mark the first appearance of the recent Hb gene after which no more subsequent duplications have occurred. The main events occurred in the pre-teleost, teleost, cichlid, or oreochromine/*Oreochromis* ancestor (timing interpreted from the topology of the gene trees in Figures S1 and S2). Note that the genes currently located at the LA cluster have an older origin (pre-teleost and teleost), whereas the genes within the MN cluster proliferated by further gene duplications. (C) The haemoglobin gene tree of the Nile tilapia extracted from the overall gene tree (i.e., Hbα and Hbβ subunit gene trees in Figures 2, S1 and S2). Hbα subunit genes in green, Hbβ subunit genes in orange.

both *Oreochromis* and *Pungu*) and further continued exclusively in *Oreochromis*.

3.5 | Positive Selection Test

We performed a test for positive selection on 186 and 175 full-length non-pseudogenized sequences of the Hbα and Hbβ haemoglobin subunit genes. Both Hbα and Hbβ subunits share similar pattern of selective pressures acting on them and often the genes with high signatures of positive selection belong to cichlids (Figures S5 and S6). We found an excess of non-synonymous mutations in the cichlid haemoglobin subunit genes that appeared as a result of lineage-specific tandem duplications, such as the recent duplicated regions, or the typical block of evolutionary older duplicated region typical

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

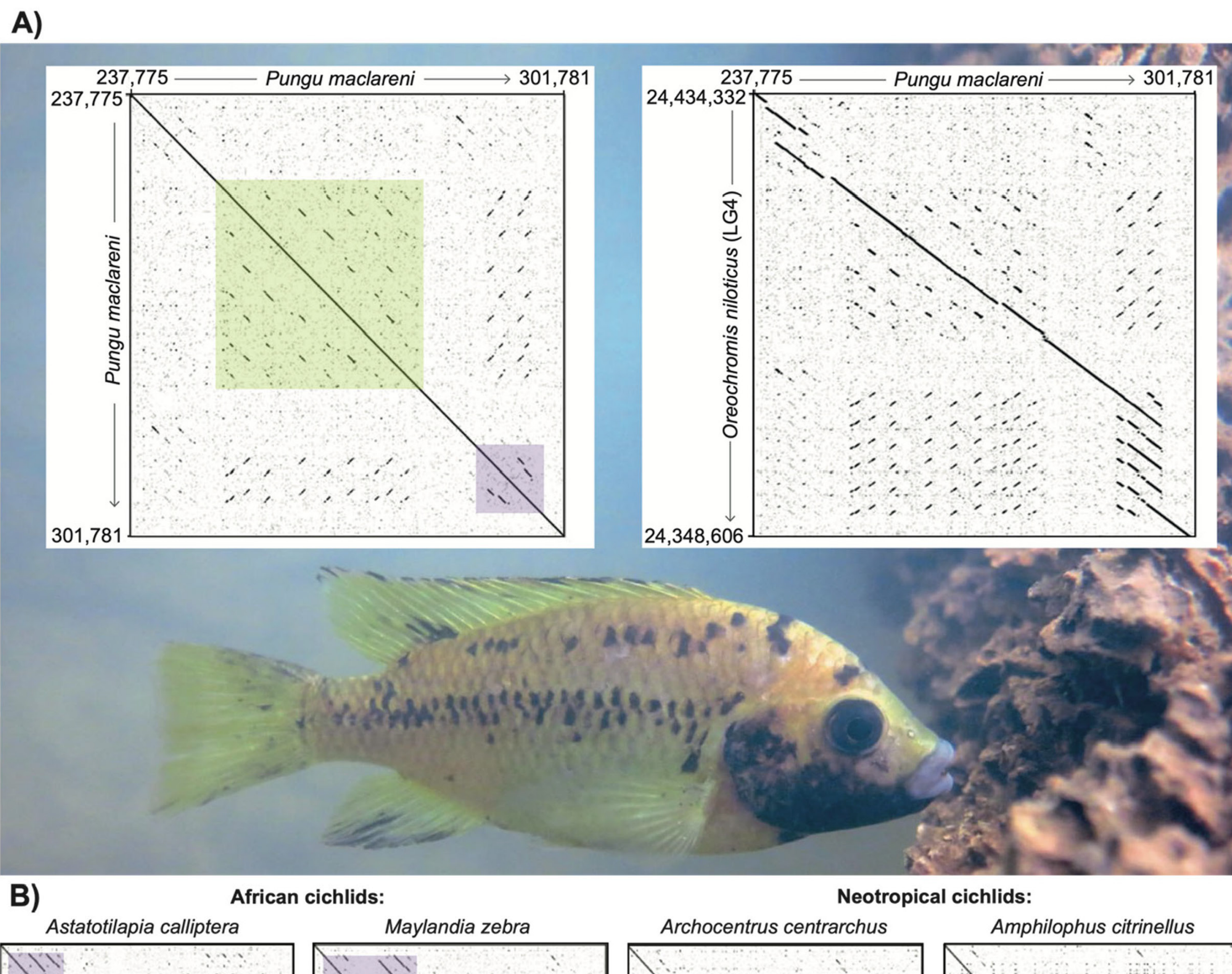

FIGURE 5 | Comparison of the cichlid haemoglobin gene MN cluster. (A) Gepard dotplots reveal similar regions of the MN cluster nucleotide sequences from pungu (*Pungu maclareni*; contig_791) plotted against itself (left) and against the MN cluster of Nile tilapia (*Oreochromis niloticus*; right). Photo of pungu (*Pungu maclareni*), the species from the Barombi Mbo crater lake sequenced within this study. (B) Gepard plots of the MN cluster of four other cichlid species in this study (each plotted against itself). Green rectangle highlights a region of the older major gene duplication event specific for African cichlids, purple squares mark regions of the recent duplications (which have independently happened in different species). See Figure 3 for the Nile tilapia gepard plot (plotted against itself), and Figure S4 for more details (such as genomic coordinates) of the gepard plots of the four cichlid species.

for African cichlids1. This pattern can be observed for example in recently duplicated genes, such as Hbβ subunits Mn1 and Mn3 together with tandemly duplicated Mn18, Mn20, Mn22, Mn24, Mn26 and Mn28 in *Oreochromis niloticus*, Mn1 Hbβ + Mn2 Hbα and Mn5 Hbβ + Mn6 Hbα in *Maylandia zebra*, Hbα: Mn2, Mn4, Mn6 and Hbβ: Mn17, Mn19 and Mn21 in *Astatotilapia calliptera*, Hbα: Mn7, Mn11, Mn5, Mn18, Mn19, Hbβ: Mn6 and Mn8 in *Archocentrus centrarchus*, and, finally, Mn16 and Mn18 Hbβ in *Pungu maclareni*. In several of these genes, the burst of positive selection was detected on the branch before their duplication (Figures S5 and S6). In some cases, similar patterns were also found in non-cichlids with expanded Hb gene repertoire, such as pike (De1 Hbα and De8 Hbα; DE is a haemoglobin cluster specific for Northern pike

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

flanked by dnmt1 and eif3g genes) and salmon (Mn1.10 Hbα and Mn2.13 Hbα). Signs of elevated positive selection are also found in the cathodic Hbβ subunit genes of African cichlids (Figures S5 and S6).

4 | Discussion

Over evolutionary time, teleost fishes have undergone remarkable genetic changes in response to selective pressures associated with their aquatic habitats. Understanding the molecular evolution of fish haemoglobins not only elucidates the genetic basis for their functional diversity but also offers insights into interplay between genes, environment, and adaptation in the aquatic realm. In this study, we applied comparative genomics to trace the evolutionary history of haemoglobin genes across different teleost (and particularly cichlid) fish species, revealing patterns of gene duplications, losses, and genome rearrangements.

Although haemoglobins are well studied and understood from many perspectives, the molecular basis and effects of single mutations are still less known than for example for rhodopsins (Yokoyama 2008; Musilova, Salzburger, and Cortesi 2021). Major steps in evolution of haemoglobins, such as emergence of tetramers from the monomers and via dimers, has been described at the molecular level only recently (Pillai et al. 2020) by identification of the key amino acid substitutions, and this enables opportunities for subsequent research.

4.1 | Haemoglobin Gene Repertoire Organised in Clusters and the Whole-Genome Duplication(s)

Teleost fishes exhibit an overwhelming morphological, ecological and taxonomical diversity, and, similarly, we also observe diversity in the genomic organisation of the haemoglobin gene clusters. Although the ancestral vertebrate whole-genome duplications played a crucial role in the origin of haemoglobins (Hoffmann, Opazo, and Storz 2012), we report that the haemoglobin diversity has been mostly driven by gene (mostly tandem) duplications contributing possibly to the adaptive potential of haemoglobins (Storz 2016). The ancestor of jawed vertebrates most likely had a single haemoglobin gene pair, i.e., one Hbα and one Hbβ subunit gene (Figure 4; Opazo et al. 2013). An organisation of Hb genes in a single cluster is also found in a non-teleost ray-finned fish (i.e., gar—*Lepisosteus oculatus*) and in cartilaginous fishes (Figure 1), whereas tetrapods and teleosts possess two clusters in their genome (most likely of independent origin since there are no similarities in the flanking genes nor the Hb gene sequence; Figure 1). The origin of the second, shorter, cluster in teleosts (labelled as LA cluster) has been previously associated with the teleost-specific whole genome duplication (TSGD) (Opazo et al. 2013). Our data does not support this hypothesis. One of the Hbα genes (sometimes also referred to as Hbα subunits D-like due to their similarity to the haemoglobin D; Ahmed, Ghatge, and Safo 2020; Figures 2 and S1) located in the LA cluster seems to have an older evolutionary origin, i.e., diversifying prior to the emergence of teleosts, [i.e., the teleost LA Hbα genes group together with one gar (non-teleost) Hbα gene, whereas other teleost Hbαs are more related to other two gar Hbα haemoglobins (Figures 2, 4 and S1)]. On the other hand, the paralogous flanking gene *rhbdf1* is found in both LA and MN clusters in all teleost lineages, suggesting the possible origin of the cluster by the TSGD. However, to test this we have reconstructed the gene tree of the *rhbdf1* genes from our data set (Figure S3) and found that the common origin of the two rhbdf1 paralogs is evolutionarily younger, in the Otomorpha + Euteleostei ancestor. There are clear gene clades of the LA and MN *rhbdf1* of otomorphs (e.g., zebrafish) + euteleosts (e.g., tilapia), whereas the elopomorph (eel) and osteoglossomorph (elephant fish) MN and LA copies form a separate clade (Figure S3). As a conclusion, we suggest that the evolution of the Hb genes most likely preceded their organisation into clusters, i.e., some of the existing genes (of pre-teleost origin) have been reorganised into newly emerged clusters later (post-teleost diversification). This is also supported by the grouping of teleost LA Hbα genes with one gar Hbα gene in the gene tree (Figure S1), whereas the physical location of this gene in the gar genome is within a single gene cluster together with all other haemoglobins (Figure 1). Interestingly, reorganisation of Hb genes seems to be a common feature also among tetrapods, some of which have experienced massive haemoglobin gene reorganisation into separate Hbα and Hbβ clusters (Figure 1; Hardison 2012). In sharks, gars and teleosts, on the other hand, the Hbα and Hbβ subunit genes are mixed in the haemoglobin clusters mainly as the Hbα—Hbβ couples.

4.2 | Function of Multiple Haemoglobin Genes in Teleosts

Teleosts seem to have the highest number of haemoglobin genes among vertebrates so far, ranging from five (fugu) to 43 found in salmon (Figure 1) compared to two haemoglobins in shark, seven in zebra finch, ten in mouse and human, and 16 in clawed frog. All teleosts have at least two haemoglobin clusters, the LA and MN clusters. We found that northern pike and Atlantic salmon have an increased number of haemoglobin clusters and genes. Pike has 38 Hb genes located in three haemoglobin clusters (LA, MN and DE; the latter two on the same chromosome but in two separate regions). Salmon has 43 Hb genes in four haemoglobin clusters thanks to the *Salmoniformes*-specific whole genome duplication 12 million years ago (Koop et al. 2008). Many studies focus on function of fish haemoglobins, yet just a few of them connect genotype (haemoglobin genes) and phenotype (e.g., physiological properties), such as those focused on the cathodic haemoglobins (Fago et al. 1995) of on fishes from extreme environments (Mazzarella et al. 2006). In general, more gene copies (duplicates) do not automatically mean an enhanced function, although higher number of paralogous genes provides more substrate for selection and sub/neofunctionalization events in evolution, and, therefore, it may contribute to adaptive potential. Certain association with the environment has been found in codfishes (Gadiformes), with higher number of Hb copies in the shallow-dwelling species, whereas deep-water species have somewhat limited haemoglobin repertoire (Baalsrud et al. 2017). This is most likely because shallow-water habitats are often variable in the oxygen availability, whereas the deep-water zone is more stable. Another reason for multiple Hb gene copies is alternative gene expression during ontogeny

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and the dynamics of developmental changes might be enormous. Developmental switches of haemoglobin expression are actually known from many vertebrate groups, and have most likely evolved independently (Rohlfing et al. 2016). In the Nile tilapia, for example, the highest embryonic expression involves two haemoglobin genes from the LA cluster (La1 and La2) followed by the expression of the Mn13 and Mn14 genes, whereas the adult expression sticks mainly to three Hbα (Mn2, Mn4, Mn14) and three Hbβ (Mn1, Mn3, Mn13) genes from the MN cluster (Figure 4). Interestingly, the most recently duplicated genes (i.e., Mn18-Mn29) are labelled as "embryonic" in the GenBank genome annotation as well as in previous literature (Opazo et al. 2013; Baalsrud et al. 2017), albeit the embryo transcriptome (SAMN00767863) does not support this. A proper independent confirmation based on different developmental stages is still lacking for tilapia haemoglobins (Figure 4). Generally, in case of an urgent need of functional haemoglobin at the same time, the higher number of identical genes increases the efficiency of transcription and proteosynthesis, and as such it can be beneficial to have multiple copies even if they have identical function (Kondrashov 2012). Accordingly, the embryonic and juvenile-specific haemoglobin genes have been identified in the Atlantic cod (Baalsrud et al. 2017 and references therein) and zebrafish (Brownlie et al. 2003).

Interestingly, in the species with the more complex genome, such as in salmon, some LA genes have an MN origin, bringing further evidence for common and recent rearrangements of the Hb genes between clusters (Quinn et al. 2010). Contrarily, the pufferfishes (*Tetraodontiformes*) are known for a genome compaction (Brenner et al. 1993), which also affected haemoglobin genes in the MN cluster. Fugu has the lowest number of Hb genes (5) of all known teleosts, possessing only two haemoglobin subunit genes in the MN cluster (while the LA cluster remains conserved with three genes as in most teleosts).

Our findings for Hb gene numbers are most likely not the final ranges for teleosts given that we focused only on 24 selected vertebrate species. We believe that future research including more species may reveal even higher numbers of Hb copies. It has been impossible to infer correct genomic structure of complex loci, such as haemoglobin clusters without high-quality genome sequencing with reads long enough to cover the high similarity regions (e.g., Malinsky et al. 2018). The upcoming routine application of the high-quality genome sequencing will certainly lead to future findings regarding the haemoglobin gene clusters. This may be especially relevant for the species with high numbers of Hb genes. Noteworthy, while we analysed the most recent version of the pike and salmon genomes (with 38 and 43 Hb genes respectively), there were only 29 and 33 identified genes in the previous versions of these genomes. On the other hand, we are aware of possible assembly errors and overestimation of the gene copies in the repetitive gene clusters (Ko et al. 2022). Lastly, we also note that there might be intraspecific variability among different individuals of the same species, especially in the number of identical genes (such as Hb18-29) in tilapia. Despite that, we believe our study still provides a valuable reference for future research, and general comparison of Hb gene repertoires, even though the exact numbers might vary or change in the future.

4.3 | Diversification of the Haemoglobin Genes in Teleosts

The origin of Hbα and Hbβ haemoglobin subunit genes is a result of an ancient duplication event that likely happened ca 450 million years ago in the ancestor of gnathostome vertebrates (Czelusniak et al. 1982). Our phylogeny confirms this, as Hbα and Hbβ subunits form two well supported clades in the gene tree (Figures 2, S1 and S2).

We found three major clades within the Hbβ subunit gene tree (Figures 2 and S2) and our results are in accord with a previous fish haemoglobin gene tree (Opazo et al. 2013). Unlike the Hbα subunits, all Hbβ subunit genes seem to have diversified after the teleost ancestor (Figures 2 and S2). The earliest branching genes are the cathodic haemoglobins (labelled as Hbβ MN15 in tilapia) located within the MN cluster. Cathodic haemoglobin gene (β subunit) is found as a single copy gene in most of the studied species, with the exception of pike (*Esox lucius*), Atlantic cod (*Gadus morhua*) and Japanese pufferfish (*Takifugu rubripes*) where it has been lost, and in European eel and Atlantic salmon, where it has been duplicated into multiple copies (Figure S2).

The teleost ancestor is likely to have had at least four Hbα genes, whereas the signal is not that clear for Hbβ. The inferred phylogenetic pattern indicates the presence of at least three putative Hbβ copies in the teleost ancestor, which have emerged by gene duplications. Most likely, there was another duplication leading to four ancestral teleost Hbβ genes (and lost in some lineages), or that the last duplications happened later in evolution, in the ancestor of Otomorpha+Euteleostei. Our gene trees (Figures 2 and S2) suggest the latter, whereas the reconstructed scenario and the physical location in LA versus MN clusters speaks rather for the former option. Additional analyses with more species are needed to resolve this uncertainty.

Our Hbα subunit gene tree is partially discordant with previously reported phylogenies, and the LA clade 2 (as in Opazo et al. 2013) could not be easily delineated; instead the genes are grouped within the combined MN/LA clade. Unlike most other clades within the gene tree, this clade is poorly supported, which could have been influenced by factors such as gene conversion or strong selection driving convergent evolution. Gene conversion and subsequent non-reciprocal exchange of the genetic material could in fact obscure the phylogenetic inference in paralogous genes (Archibald and Roger 2002; Ratnakumar et al. 2010; Cortesi et al. 2015) although no conversion has been recently detected by a targeted analysis in ray-finned fish haemoglobin genes (Mao et al. 2023). Strong positive selection has also been found in the mixed Hbα clade (see below) and could partially explain the poor resolution of—and within—this clade (Figures S5 and S6). These factors are, however, unlikely to compromise the overall strong phylogenetic signal, as we were generally able to recover well supported clades (with the aforementioned exception).

4.4 | Evolution of Haemoglobins in Cichlid Fishes

Cichlids are characterised by a relatively high number of haemoglobin subunit genes compared to other studied teleosts

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Figure 1), ranging from 18 haemoglobin genes in *Amphilophus citrinellus* up to 32 in *Oreochromis niloticus*. Only two teleost species (pike, salmon) have a greater haemoglobin gene repertoire (38 and 43, respectively). All studied cichlids share the same synteny of the LA cluster including haemoglobins and three flanking genes upstream and downstream of the cluster. We report variation in the gene copies solely within the MN cluster of cichlids and this is probably a result of tandem duplications. Some Hb genes from African and Neotropical cichlids form sister clades suggesting emergence of many copies prior to cichlid diversification. Substantial proportion of the gene diversity was apparently driven by lineage-specific or species-specific duplication events that happened after the split of the African and Neotropical lineages throughout cichlid evolution.

It has previously been shown that cichlids exhibit elevated numbers of duplicated genes in general (Brawand et al. 2014, Berner and Salzburger 2015), and possibly this might have contributed to their evolutionary success—resulting in large numbers of species and adaptive radiations that enabled them to conquer various ecological and trophic niches. More specifically, Victoria cichlids are known to switch between haemoglobin types to cope with hypoxia (Rutjes et al. 2007; van den Thillart et al. 2018), and signs of positive selection have been reported in four haemoglobin genes from the MN cluster in deep-water Malawi cichlid from *Diplotaxodon* genus (Malinsky et al. 2018). Since haemoglobins are putative candidates for adaptive evolution, we perform the analysis of positive selection on the haemoglobin genes. We found an excess in non-synonymous-to synonymous mutations in some (but not all) duplicated genes of cichlid species, however the pattern is not clear. Most of the genes under positive selection result from recent duplications, might have experienced bursts of positive selection, namely the tilapia recent duplicates MN1 Hbβ and MN3 Hbβ subunit genes (with elevated positive selection signal). Interestingly, Nile tilapia itself is not a part of any cichlid radiation although it has the highest number of Hb genes so far. However, it is a species quite tolerant to hypoxia (Bergstedt, Pfalzgraff, and Skov 2021), hence, there may be a biological relevance or traces of positive selection speaking for episodes of recent selection.

5 | Conclusion

We apply comparative genomics to reveal a great diversity of the haemoglobin gene repertoires. Gene duplications, gene losses and genomic rearrangements have all contributed to the genomic architecture of haemoglobin subunit genes in teleost fishes. We conclude that tandem duplications played a major role in the dynamic increase of haemoglobin gene copies within the MN cluster of cichlid fishes. Furthermore, our results demonstrate that some genes within the LA cluster had a pre-teleost origin, whereas the LA cluster itself may be evolutionarily younger, speaking for genome rearrangements. Contrarily, the MN cluster underwent a very dynamic evolution—from a dramatic genome compaction in fugu with only two genes preserved in the MN cluster to the expansion in cichlid lineage (up to 29 genes), or furthermore, to duplication of both clusters in salmonid fishes. The teleost ancestor had at least four Hbα genes and three or four Hbβ genes, and the extant diversity of the haemoglobin gene repertoires has arisen from these ancestral copies. This study aims to serve as a valuable resource for the future research aiming to integrate observation from the wild with physiological experiments, targeted mutagenesis and in vitro protein engineering to understand the mechanisms of haemoglobin molecular evolution and their function.

Author Contributions

D.O. and Z.M. conceived the study, D.O. performed the analyses, figure drafts and revisions, writing, proofreading. Z.M. did conceptualization, figure drafts and revisions, writing, proofreading. M.Mal. helped to process genomic data, M.Mat. contributed to the detection of positive selection signatures. A.R.B.-N., A.I., D.O. and Z.M. performed and supported fieldwork. O.B. was responsible for Nanopore sequencing and Pungu maclareni genome assembly. Z.M. and D.O. together with W.S., M.Mal., M.Mat. revised the manuscript.

Acknowledgements

We would like to express our thanks to Veronika Truhlarova for a technical support, Helle Tessand Baalsrud, Sissel Jentoft and Monica Hongrø Solbakken for the help with the entire first steps in the bioinformatic analyses, Demian Burguera for his bioinformatic insight and Kristyna Eliasova for additional help with bioinformatics, Daniel Elleder for help with the long-fragment DNA extraction and Petr Pajer for his valuable help and performing the Nanopore sequencing protocol. We further thank local people of the Barombi village to allow us to fish in the Barombi Mbo lake, the Ministry of Scientific Research and Innovation in Cameroon to grant us research permits (for D.O., A.I., A.R.B.-N., and Z.M.). Special thanks to Herman Igor Kittio for help with the fieldwork. We would also like to extend our thanks to Karyna Pahuta for her guidance on graphic production. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic. The project was funded by the Czech Science Foundation (21-31712S) and Charles University Grant Agency (GAUK 1556119).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The genomic data of the de novo sequenced *Pungu maclareni* in this study has been deposited on GenBank under the Accession number: GCA_041757325.1. The alignment of the haemoglobin genes is included as. nex format in the Data S1.

Benefit Sharing Statement

The results of this research generate a benefit from the sharing our results on public available databases as described above. A local scientist has been a valuable member of our research team and one of the coauthors. More broadly, our group is committed to international scientific partnerships, as well as institutional capacity building (we have repeatedly hosted our collaborator and his PhD student in our laboratories for research visits).

References

Abi-Rached, L., A. Gilles, T. Shiina, P. Pontarotti, and H. Inoko. 2002. "Evidence of En Bloc Duplication in Vertebrate Genomes." *Nature Genetics* 31, no. 1: 100–105.

Ahmed, M. H., M. S. Ghatge, and M. K. Safo. 2020. "Hemoglobin: Structure, Function and Allostery." *Sub-Cellular Biochemistry* 94: 345–382.

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Archibald, J. M., and A. J. Roger. 2002. "Gene Conversion and the Evolution of Euryarchaeal Chaperonins: A Maximum Likelihood-Based Method for Detecting Conflicting Phylogenetic Signals." *Journal of Molecular Evolution* 55, no. 2: 232–245.

Baalsrud, H. T., K. L. Voje, O. K. Tørresen, et al. 2017. "Evolution of Hemoglobin Genes in Codfishes Influenced by Ocean Depth." *Scientific Reports* 7, no. 1: 7956.

Bergstedt, J. H., T. Pfalzgraff, and P. V. Skov. 2021. "Hypoxia Tolerance and Metabolic Coping Strategies in *Oreochromis niloticus*." *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology* 257: 110956.

Berner, D., and W. Salzburger. 2015. "The Genomics of Organismal Diversification Illuminated by Adaptive Radiations." *Trends in Genetics* 31, no. 9: 491–499.

Betancur-R, R., E. O. Wiley, G. Arratia, et al. 2017. "Phylogenetic Classification of Bony Fishes." *BMC Evolutionary Biology* 17: 1–40.

Bolognini, D., N. Bartalucci, A. Mingrino, A. M. Vannucchi, and A. Magi. 2019. "NanoR: A User-Friendly R Package to Analyze and Compare Nanopore Sequencing Data." *PLoS One* 14, no. 5: e0216471.

Brawand, D., C. E. Wagner, Y. I. Li, et al. 2014. "The Genomic Substrate for Adaptive Radiation in African Cichlid Fish." *Nature* 513, no. 7518: 375–381.

Brenner, S., G. Elgar, R. Sanford, A. Macrae, B. Venkatesh, and S. Aparicio. 1993. "Characterization of the Pufferfish (Fugu) Genome as a Compact Model Vertebrate Genome." *Nature* 366, no. 6452: 265–268.

Brownlie, A., C. Hersey, A. C. Oates, et al. 2003. "Characterization of Embryonic Globin Genes of the Zebrafish." *Developmental Biology* 255, no. 1: 48–61.

Chan, F. Y., J. Robinson, A. Brownlie, et al. 1997. "Characterization of Adult α-and β-Globin Genes in the Zebrafish." *Blood* 89, no. 2: 688–700.

Chen, Z., F. Qiao, Y. He, H. Xie, and D. Qi. 2016. "Evidence for Positive Selection on α and β Globin Genes in Pikas and Zokor From the Qinghai-Tibetan Plateau." *Gene and Translational Bioinformatics* 2: 1–10.

Cortesi, F., Z. Musilová, S. M. Stieb, et al. 2015. "Ancestral Duplications and Highly Dynamic Opsin Gene Evolution in Percomorph Fishes." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 5: 1493–1498.

Czelusniak, J., M. Goodman, D. Hewett-Emmett, M. L. Weiss, P. J. Venta, and R. E. Tashian. 1982. "Phylogenetic Origins and Adaptive Evolution of Avian and Mammalian Haemoglobin Genes." *Nature* 298, no. 5871: 297–300.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. "jModelTest 2: More Models, New Heuristics and Parallel Computing." *Nature Methods* 9, no. 8: 772.

Fago, A., V. Carratore, G. di Prisco, R. J. Feuerlein, L. Sottrup-Jensen, and R. E. Weber. 1995. "The Cathodic Hemoglobin of *Anguilla anguilla*: Amino Acid Sequence and Oxygen Equilibria of a Reverse Bohr Effect Hemoglobin With High Oxygen Affinity and High Phosphate Sensitivity (*)." *Journal of Biological Chemistry* 270, no. 32: 18897–18902.

Glasauer, S. M., and S. C. Neuhauss. 2014. "Whole-Genome Duplication in Teleost Fishes and Its Evolutionary Consequences." *Molecular Genetics and Genomics* 289, no. 6: 1045–1060.

Green, J., S. A. Corbet, and E. Betney. 1973. "Ecological Studies on Crater Lakes in West Cameroon the Blood of Endemic Cichlids in Barombi Mbo in Relation to Stratification and Their Feeding Habits." *Journal of Zoology* 170: 299–308.

Hahn, C., M. J. Genner, G. F. Turner, and D. A. Joyce. 2017. "The Genomic Basis of Cichlid Fish Adaptation Within the Deepwater "Twilight Zone" of Lake Malawi." *Evolution Letters* 1, no. 4: 184–198.

Hardison, R. C. 2008. "Globin Genes on the Move." *Journal of Biology* 7, no. 9: 35.

Hardison, R. C. 2012. "Evolution of Hemoglobin and Its Genes." *Cold Spring Harbor Perspectives in Medicine* 2, no. 12: a011627.

Hoang, D. T., O. Chernomor, A. Von Haeseler, B. Q. Minh, and L. S. Vinh. 2017. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35, no. 2: 518–522.

Hoffmann, F. G., J. C. Opazo, and J. F. Storz. 2012. "Whole-Genome Duplications Spurred the Functional Diversification of the Globin Gene Superfamily in Vertebrates." *Molecular Biology and Evolution* 29, no. 1: 303–312.

Holland, P. W., J. Garcia-Fernàndez, N. A. Williams, and A. Sidow. 1994. "Gene Duplications and the Origins of Vertebrate Development." *Development* 1994, no. Supplement: 125–133.

Hughes, A. L., and R. Friedman. 2003. "2R or Not 2R: Testing Hypotheses of Genome Duplication in Early Vertebrates." In *Genome Evolution*, edited by A. Meyer and Y. van de Peer, 85–93. Dordrecht: Springer.

Kalyaanamoorthy, S., B. Q. Minh, T. K. Wong, A. von Haeseler, and L. S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14, no. 6: 587–589.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780.

Kautt, A. F., G. Machado-Schiaffino, and A. Meyer. 2016. "Multispecies Outcomes of Sympatric Speciation After Admixture With the Source Population in Two Radiations of Nicaraguan Crater Lake Cichlids." *PLoS Genetics* 12, no. 6: e1006157.

Kearse, M., R. Moir, A. Wilson, et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28, no. 12: 1647–1649.

Ko, B. J., C. Lee, J. Kim, et al. 2022. "Widespread False Gene Gains Caused by Duplication Errors in Genome Assemblies." *Genome Biology* 23, no. 1: 205.

Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37, no. 5: 540–546.

Kondrashov, F. A. 2012. "Gene Duplication as a Mechanism of Genomic Adaptation to a Changing Environment." *Proceedings of the Royal Society B: Biological Sciences* 279, no. 1749: 5048–5057.

Koop, B. F., K. R. Von Schalburg, J. Leong, et al. 2008. "A Salmonid EST Genomic Study: Genes, Duplications, Phylogeny and Microarrays." *BMC Genomics* 9, no. 1: 545.

Krumsiek, J., R. Arnold, and T. Rattei. 2007. "Gepard: A Rapid and Sensitive Tool for Creating Dotplots on Genome Scale." *Bioinformatics* 23, no. 8: 1026–1028.

Kuraku, S., and A. Meyer. 2009. "The Evolution and Maintenance of Hox Gen in Vertebrates and the Teleost-Specific Genome Duplication." *International Journal of Developmental Biology* 53: 765–773.

Lee, J. Y., M. Kong, J. Oh, et al. 2021. "Comparative Evaluation of Nanopore Polishing Tools for Microbial Genome Assembly and Polishing Strategies for Downstream Analysis." *Scientific Reports* 11, no. 1: 20740.

Lei, Y., L. Yang, Y. Zhou, et al. 2021. "Hb Adaptation to Hypoxia in High-Altitude Fishes: Fresh Evidence From Schizothoracinae Fishes in the Qinghai-TiHbβn Plateau." *International Journal of Biological Macromolecules* 185: 471–484.

Malinsky, M., R. Challis, A. M. Tyers, et al. 2015. "Genomic Islands of Speciation Separate Cichlid Ecomorphs in an East African Crater Lake." *Science* 350, no. 6267: 1493–1498.

Malinsky, M., H. Svardal, A. M. Tyers, et al. 2018. "Whole-Genome Sequences of Malawi Cichlids Reveal Multiple Radiations

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Interconnected by Gene Flow." *Nature Ecology and Evolution* 2: 1940–1955.

Mao, Y., T. Peng, F. Shao, Q. Zhao, and Z. Peng. 2023. "Molecular Evolution of the Hemoglobin Gene Family Across Vertebrates." *Genetica* 151, no. 3: 201–213.

Marino, K., L. Boschetto, D. de Pascale, and E. Cocca. 2007. "Organisation of the Hb 1 Genes of the Antarctic Skate Bathyraja Eatonii: New Insights Into the Evolution of Globin Genes." *Gene* 406, no. 1–2: 199–208.

Mazzarella, L., G. Bonomi, M. C. Lubrano, et al. 2006. "Minimal Structural Requirements for Root Effect: Crystal Structure of the Cathodic Hemoglobin Isolated From the Antarctic Fish *Trematomus newnesi*." *Proteins: Structure, Function, and Bioinformatics* 62, no. 2: 316–321.

McKiernan, H. E., and P. B. Danielson. 2017. "Molecular Diagnostic Applications in Forensic Science." In *Molecular diagnostics*, 371–394. San Diego, CA: Academic Press.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees." In *In 2010 Gateway Computing Environments Workshop (GCE)*, 1–8. New Orleans, LA: IEEE.

Mukherjee, S., O. Bartoš, K. Zdeňková, P. Hanák, P. Horká, and Z. Musilova. 2021. "Evolution of the Parvalbumin Genes in Teleost Fishes After the Whole-Genome Duplication." *Fishes* 6, no. 4: 70.

Musilova, Z., A. Indermaur, A. R. Bitja-Nyom, et al. 2019. "Evolution of the Visual Sensory System in Cichlid Fishes From Crater Lake Barombi Mbo in Cameroon." *Molecular Ecology* 28, no. 23: 5010–5031.

Musilova, Z., W. Salzburger, and F. Cortesi. 2021. "The Visual Opsin Gene Repertoires of Teleost Fishes: Evolution, Ecology, and Function." *Annual Review of Cell and Developmental Biology* 37: 441–468.

Nery, M. F., J. I. Arroyo, and J. C. Opazo. 2013. "Genomic Organization and Differential Signature of Positive Selection in the Hbα and Hbβ Globin Gene Clusters in Two Cetacean Species." *Genome Biology and Evolution* 5, no. 12: 2359–2367.

Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2014. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32, no. 1: 268–274.

Ohno, S. 1970. *In Evolution by Gene Duplication*, 132–139. New York: Springer Verlag.

Opazo, J. C., G. T. Butts, M. F. Nery, J. F. Storz, and F. G. Hoffmann. 2013. "Whole-Genome Duplication and the Functional Diversification of Teleost Fish Hemoglobins." *Molecular Biology and Evolution* 30, no. 1: 140–153.

Parey, E., A. Louis, J. Montfort, Y. Guiguen, H. R. Crollius, and C. Berthelot. 2022. "An Atlas of Fish Genome Evolution Reveals Delayed Rediploidization Following the Teleost Whole-Genome Duplication." *Genome Research* 32, no. 9: 1685–1697.

Patel, V. S., S. J. Cooper, J. E. Deakin, et al. 2008. "Platypus Globin Genes and Flanking Loci Suggest a New Insertional Model for Hbβ-Globin Evolution in Birds and Mammals." *BMC Biology* 6, no. 1: 34.

Pillai, A. S., S. A. Chandler, Y. Liu, et al. 2020. "Origin of Complexity in Haemoglobin Evolution." *Nature* 581, no. 7809: 480–485.

Quinn, N. L., K. A. Boroevich, K. P. Lubieniecki, et al. 2010. "Genomic Organization and Evolution of the Atlantic Salmon Hemoglobin Repertoire." *BMC Genomics* 11: 1–17.

R Core Team. 2018. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org/.

Ratnakumar, A., S. Mousset, S. Glémin, et al. 2010. "Detecting Positive Selection Within Genomes: The Problem of Biased Gene Conversion." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, no. 1552: 2571–2580.

Rennison, D. J., G. L. Owens, and J. S. Taylor. 2012. "Opsin Gene Duplication and Divergence in Ray-Finned Fish." *Molecular Phylogenetics and Evolution* 62, no. 3: 986–1008.

Rohlfing, K., F. Stuhlmann, M. F. Docker, and T. Burmester. 2016. "Convergent Evolution of Hemoglobin Switching in Jawed and Jawless Vertebrates." *BMC Evolutionary Biology* 16: 30.

Ronco, F., M. Matschiner, A. Böhne, et al. 2021. "Drivers and Dynamics of a Massive Adaptive Radiation in Cichlid Fishes." *Nature* 589, no. 7840: 76–81.

Ronquist, F., M. Teslenko, P. Van Der Mark, et al. 2012. "MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space." *Systematic Biology* 61, no. 3: 539–542.

Rutjes, H. A., M. C. Nieveen, R. E. Weber, F. Witte, and G. E. E. J. M. Van den Thillart. 2007. "Multiple Strategies of Lake Victoria Cichlids to Cope With Lifelong Hypoxia Include Hemoglobin Switching." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 293, no. 3: R1376–R1383.

Seehausen, O. 2000. "Explosive Speciation Rates and Unusual Species Richness in Haplochromine Cichlid Fishes: Effects of Sexual Selection." *Advances in Ecological Research* 31: 237–274.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 3210–3212.

Stanke, M., and B. Morgenstern. 2005. "AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints." *Nucleic Acids Research* 33, no. suppl_2: W465–W467.

Steinke, D., W. Salzburger, I. Braasch, and A. Meyer. 2006. "Many Genes in Fish Have Species-Specific Asymmetric Rates of Molecular Evolution." *BMC Genomics* 7: 20.

Storz, J. F. 2016. "Gene Duplication and Evolutionary Innovations in Hemoglobin-Oxygen Transport." *Physiology* 31, no. 3: 223–232.

Storz, J. F. 2018. *Hemoglobin: Insights Into Protein Structure, Function, and Evolution*. Oxford: Oxford University Press.

Storz, J. F., J. C. Opazo, and F. G. Hoffmann. 2013. "Gene Duplication, Genome Duplication, and the Functional Diversification of Vertebrate Globins." *Molecular Phylogenetics and Evolution* 66, no. 2: 469–478.

Suyama, M., D. Torrents, and P. Bork. 2006. "PAL2NAL: Robust Conversion of Protein Sequence Alignments Into the Corresponding Codon Alignments." *Nucleic Acids Research* 34, no. suppl_2: W609–W612.

Svardal, H., W. Salzburger, and M. Malinsky. 2021. "Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes." *Annual Review of Animal Biosciences* 9: 55–79.

van den Thillart, G., I. Wilms, M. Nieveen, R. E. Weber, and F. Witte. 2018. "Hypoxia-Induced Changes in Hemoglobins of Lake Victoria Cichlids." *Journal of Experimental Biology* 221, no. 17: jeb177832.

Walker, B. J., T. Abeel, T. Shea, et al. 2014. "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." *PLoS One* 9, no. 11: e112963.

Wang, L. G., T. T. Y. Lam, S. Xu, et al. 2020. "Treeio: An R Package for Phylogenetic Tree Input and Output With Richly Annotated and Associated Data." *Molecular Biology and Evolution* 37, no. 2: 599–603.

Wells, R. M. 1999. "Evolution of Haemoglobin Function: Molecular Adaptations to Environment." *Clinical and Experimental Pharmacology and Physiology* 26, no. 8: 591–595.

Wick, R. R., L. M. Judd, and K. E. Holt. 2019. "Performance of Neural Network Basecalling Tools for Oxford Nanopore Sequencing." *Genome Biology* 20: 129.

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Yang, Z. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24, no. 8: 1586–1591.

Yokoyama, S. 2008. "Evolution of Dim-Light and Color Vision Pigments." *Annual Review of Genomics and Human Genetics* 9: 259–282.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

1365294x, 2024, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17559 by Schweizerische Akademie Der, Wiley Online Library on [27/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License