Journal of Great Lakes Research 50 (2024) 102326

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

The genome organization of the Lake Magadi tilapia, *Oreochromis Alcolapia grahami*, a cichlid extremophile

Giacomo Bernardi [a,*], Geraldine D. Kavembe [b], Harold L. Bergman [c], Giuseppe Bucciarelli [a], Chris M. Wood [d,e,f]

[a] Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, the United States of America
[b] School of Dryland Agriculture Science and Technology, South Eastern Kenya University, 90200 Kitui, Kenya
[c] Department of Zoology and Physiology, University of Wyoming, Laramie, WY 82071, the United States of America
[d] Department of Zoology, University of British Columbia, Vancouver, BC, Canada
[e] Dept. of Biology, McMaster University, Hamilton, ON, Canada
[f] Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami, Miami, FL, the United States of America

ARTICLE INFO

Communicated by Tomasz Mamos

Keywords:
Oreochromis Alcolapia grahami, Lake Magadi,
Isochores
GC-rich genome
Nanopore
Illumina

ABSTRACT

The genome of vertebrates is made of a mosaic of long stretches of DNA, called isochores, which are compositionally uniform, and belong to a few families of GC-poor (L1 and L2) and GC-rich (H1, H2, and H3) components. Poikilotherms tend to have GC-poor genomes, while endotherms comprise both GC-poor and GC-rich isochores. The thermal theory claimed that temperature and natural selection played an active role in favoring GC-rich genomic regions, yet empirical evidence was difficult to obtain. Early work based on cesium chloride ultracentrifugation gradients showed that the Lake Magadi tilapia, a hot-water adapted fish species, displayed GC-rich regions that were absent from a close relative that lives in colder water. The goal of this study was to revisit the original study using full genome sequencing. We found that the original GC-rich regions are indeed present, that they are interspersed in the genome. Indeed, when comparing Lake Magadi tilapia with the temperate Nile tilapia, we found that 59.3 % of the genome of Lake Magadi tilapia had a base composition higher than 40 %GC, as opposed to 55.3 % of the genome of the Nile tilapia having a base composition higher than 40 % GC. We also found that their genomes comprised similar amounts of repetitive elements (20 % and 19.5 %, respectively) indicating that the shifts in base composition might not be due to repetitive elements. Further work on repetitive element analyses, protein coding genes and additional hot-water adapted fishes will provide clues as to the origin of GC-rich isochores in Lake Magadi tilapia.

1. Introduction

The genome of vertebrates is characterized by long stretches of compositionally homogeneous regions of DNA called isochores (Bernardi, 2000; Bernardi et al., 1985). Those regions have specific GC contents and have been classified in broad categories including GC-poor (light, L1, L2) and GC-rich (heavy, H1, H2, H3) families. These families have different characteristics, with GC-rich regions comprising higher density of genes, and higher levels of gene expression (Bernardi, 2021). Several decades ago, major differences in genome organization were observed, with homeothermic vertebrates (comprising mammals and birds) displaying heterogenous genomes, as they contained both GC-rich and GC-poor families of isochores, and poikilothermic vertebrates (comprising amphibians, reptiles, and fishes) displaying more homogeneous genomes, where GC-poor isochores formed the majority of the genome. This gave rise to the thermal hypothesis, where temperature was suggested to play a role in shaping the compositional organization of the genome by increasing the amount of GC-rich isochores in vertebrates that have higher body temperatures (Bernardi and Bernardi, 1986). This hypothesis was consistent with the fact that GC-rich regions are more thermally stable, compared to AT-rich regions, and that GC-rich codons tend to encode for amino acids that help stabilize proteins (Bernardi and Bernardi, 1986), and in turn making a case for natural selection as the ultimate mechanism for the formation of GC-rich isochores. Indeed, early work, based on ultracentrifugation in cesium chloride (CsCl) density gradients, showed that the genomes of fish that

* Corresponding author.
E-mail address: bernardi@ucsc.edu (G. Bernardi).

https://doi.org/10.1016/j.jglr.2024.102326
Received 28 March 2023; Accepted 12 February 2024
Available online 10 March 2024

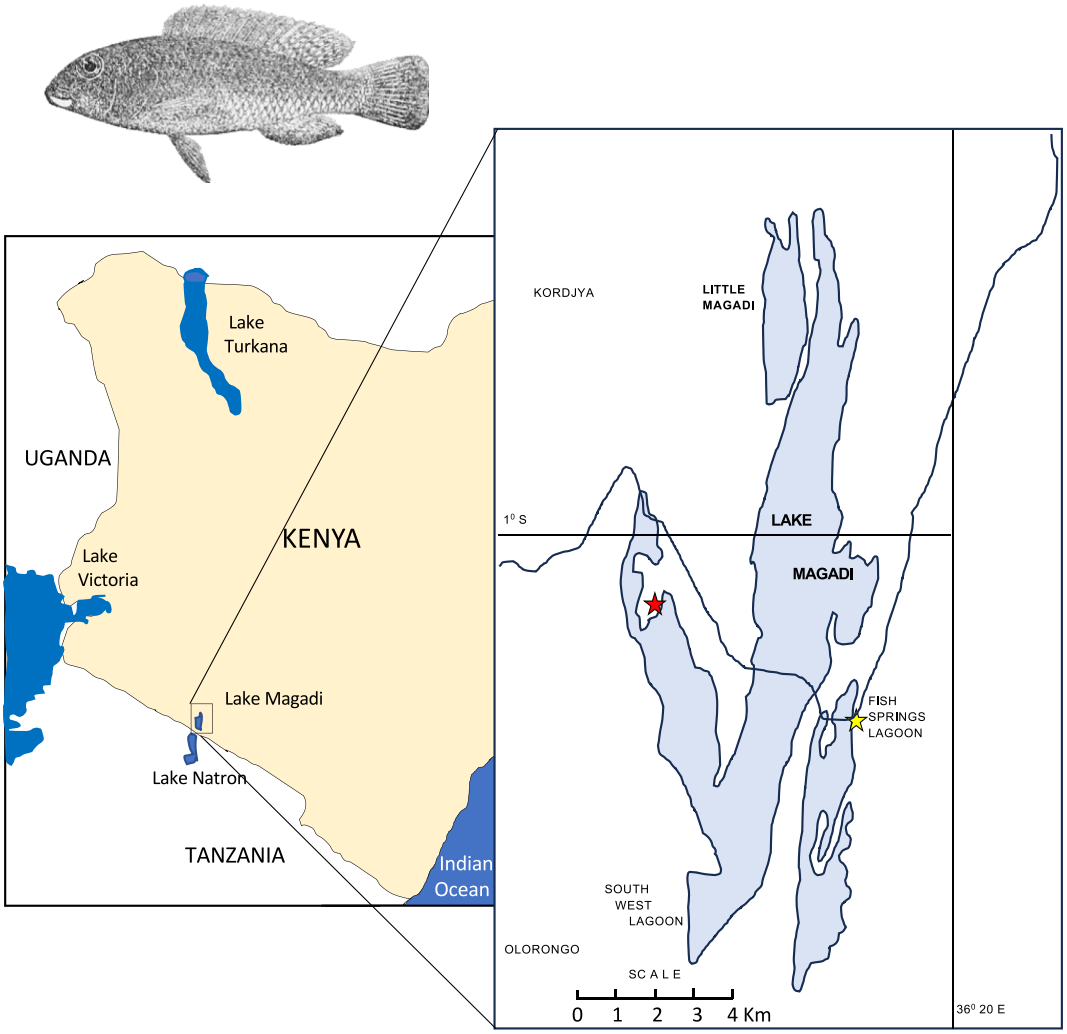

Fig. 1. Lake Magadi, Kenya and Oreochromis Alcolapia grahami. Sampling site for this study is indicated with a red star. Fish Springs Lagoon is indicated with a yellow star. Redrawn from Coe (1966).

live at high temperature exhibit GC-rich genomic domains that are absent in closely related species that live at cooler temperatures (Bernardi and Bernardi, 1986). In that study, CsCl density gradients were analyzed for the genome of the warm water adapted *Oreochromis (Alcolapia) grahami* collected in Lake Magadi, Kenya (Fig. 1). The Lake Magadi tilapia, *Oreochromis* (*Alcolapia) grahami* is a tilapiine cichlid species that lives at high temperatures, commonly up to 42 °C (Coe, 1966), with populations that show an upper critical temperature (Ct_{max}) for lethality of 44.5 °C, one of the highest recorded temperature for a fish (Wood et al., 2016). *Oreochromis (Alcolapia) grahami* showed GC-rich components that were absent in the close relative, *Oreochromis aureus*, that lives in cooler temperatures, ranging from 8 °C to 30 °C (Trewavas, 1983) (Fig. 2). At that time, it was unclear if these GC-rich regions were coding, non-coding, or repetitive sequences. Data, however, were tantalizing in their support of the thermal hypothesis. Importantly, a number of alternative hypotheses have been put forth, that generally rest on neutral processes such as biased gene conversion (Duret and Galtier, 2009; Galtier et al., 2001) and mutation bias (Eyre-Walker and Hurst, 2001). The main basis for neutral hypotheses rest on a primary neutral mechanism, for example, a mutation bias that would randomly increase GC levels, or a biased recombination that mainly affects GC-rich regions and rapidly increases their abundance. This neutral mechanism would then be followed by phylogenetic inertia, where branches leading to birds and mammals would maintain GC-rich isochores by neutral processes that do not require, nor expect, any changes in GC levels. Thus, poikilotherms would ancestrally not exhibit high GC-levels, while the derived branches of mammals and birds would. With the recent advent of rapid and affordable genomic sequencing, it is now possible to revisit the question using one of the original species, the Lake Magadi tilapia, *O. A. grahami*.

Lake Magadi is an ancient alkaline Rift Lake located in southern Kenya that is seasonally flooded and located approximately 600 m above sea level (Owen et al., 2018) (Fig. 1). It is part of a greater complex of alkaline lakes, Lake Natron in Tanzania, to the south, and an additional small lake, Little Magadi, to the northwest, that were originally part of the larger paleolake Orolonga, that probably completed their isolation about 7,000 years ago (Butzer et al., 1972; Tichy and Seegers, 1999). Together, these lakes harbor a species flock of tilapiine cichlid fishes (Ford et al., 2019a; Seegers et al., 1999; Seegers and Tichy, 1999; Trewavas, 1983). This flock comprises species in a genus formerly named *Alcolapia.* Recent molecular work has shown that the genus *Alcolapia* is embedded within the tilapiine genus *Oreochromis,* and as such, should be regarded as a subgenus (Ford et al., 2019a). Species in the subgenus *Alcolapia* are closely related to an *Oreochromis* species that is also adapted to survive in alkaline conditions, the Manyara tilapia, *O. amphimelas*, suggesting that the adaptation arose in this entire group (*O. amphimelas* + *Alcolapia*). *Alcolapia* species, which are adapted to high pH (commonly exceeding pH = 9) and high temperature (up to 42 °C) form a species flock comprising *Oreochromis Alcolapia alcalica, O. A. latilabris,* and *O. A. ndalalani,* that live in Lake Natron, Tanzania, and

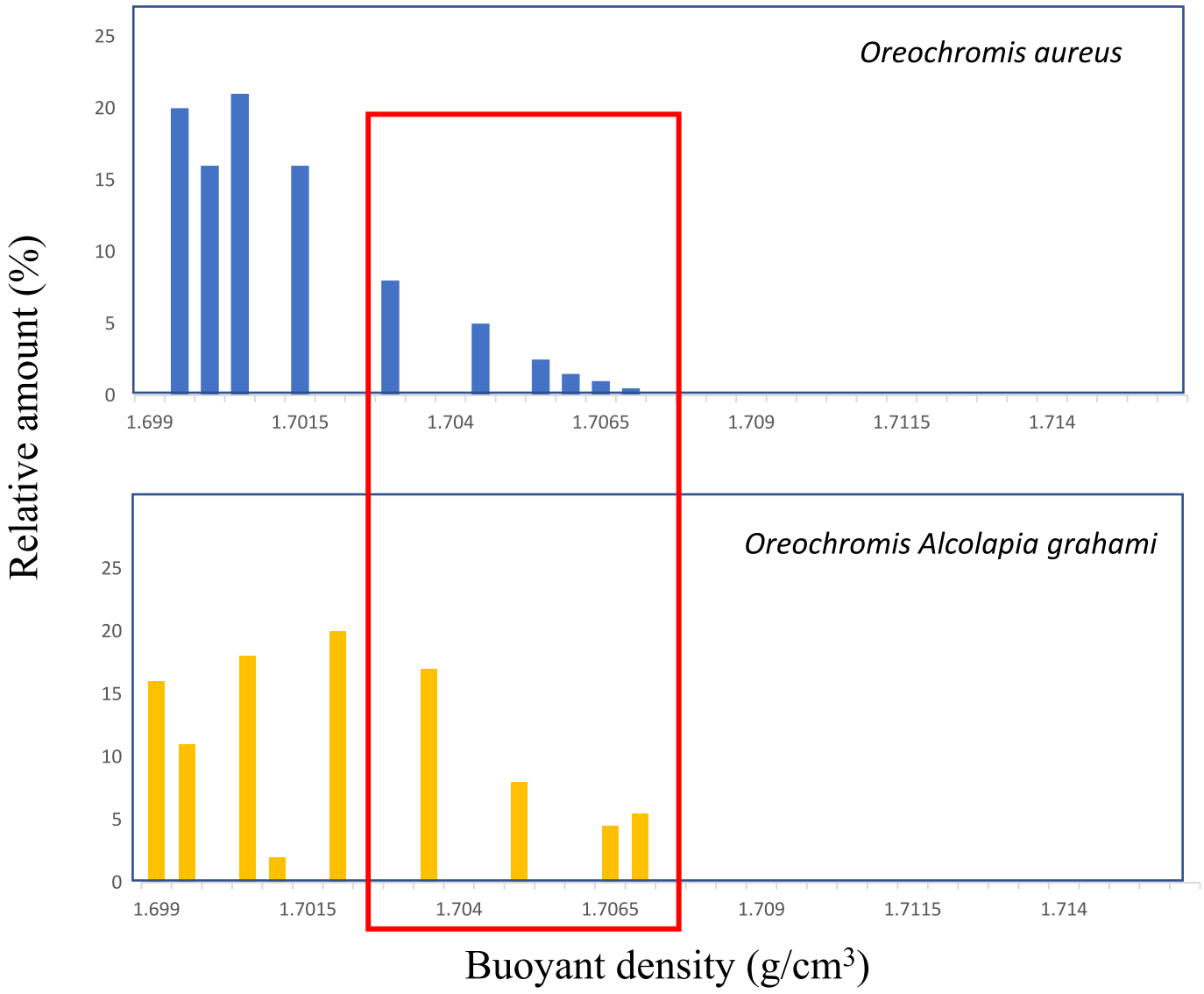

Fig. 2. Bar charts showing the relative amounts and Cesium Chloride (CsCl) buoyant densities of DNA fractions for Oreochromis aureus and Oreochromis Alcolapia grahami. The red rectangle highlights the GC-rich region of the genome for both species. Redrawn from Bernardi and Bernardi (1986).

O. A. grahami, that is restricted to Lake Magadi (Ford et al., 2019). *Oreochromis Alcolapia grahami* is a species that is found in shallow pools of warm and alkaline waters of Lake Magadi and adjacent Little Magadi (Coe, 1966) (Fig. 1). Its taxonomy has been modified over the years, moving across nomenclatures several times, including *Tilapia grahami, Sarotherodon alcalicus grahami, Oreochromis grahami, Alcolapia grahami,* and finally *Oreochromis* (*Alcolapia) grahami*). It has been the subject of studies mostly focused on its unique ureotelic and extreme environmental adaptation physiology (Johannsson et al., 2014; Kavembe et al., 2016, 2015; Wood et al., 2016, 2013, 2012), but has also been the topic of a few studies that have dealt with its population genetics (Ford et al., 2019a,b; Johannsson et al., 2014; Kavembe et al., 2015; Wilson et al., 2004, 2000).

The goal of this study was to sequence the genome of *O. A. grahami* to elucidate its compositional organization, by looking at its isochore structure, and repetitive sequences. Its genome was compared to the published genome of *Oreochromis niloticus*, a species that lives at more moderate temperatures, naturally below 30 °C (Nivelle et al., 2019).

2. Materials and methods

2.1. Sample collection and DNA extraction

Fin, liver, and gill clips were taken from eight individuals that were collected in Lake Magadi, Kenya (1° 51.589S, 36° 13.826E, Fig. 1) with hand nets. Samples were immediately placed in 95 % ethanol and stored at −20 °C upon reaching the laboratory. DNA was extracted using a DNeasy Blood and Tissue kit according to the manufacturer's protocol (Qiagen, Hilden, Germany).

2.2. Mitochondrial DNA sequencing, phylogenetic analysis, and mitochondrial genome

The mitochondrial control region (D-loop) was PCR amplified following published protocols (Domingues et al., 2005). To ascertain the identity of our samples, sequences of the eight individuals were placed in a phylogenetic context to ascertain that the genome of the correct species was to be sequenced. The mitochondrial D-loop of our eight samples and an additional 72 *Alcolapia* individuals from GenBank, were aligned using Geneious (v10.2.6), and then used to reconstruct

phylogenetic trees using Neighbor-Joining approaches with the APE (Analyses of Phylogenetics and Evolution) R-package (Paradis and Schliep, 2019). Finally, the complete mitochondrial genome of *Oreochromis Alcolapia grahami* was extracted from the genomic sequencing runs and assembled using the MINIMAP2 and SAMTOOLS packages (Danecek et al., 2021; Li, 2018), to control again that the sequenced genome was of the correct species.

2.3. Whole-genome library construction and sequencing

From the mitochondrial DNA analysis mentioned above, a single individual (that belonged to the *O. A. grahami* clade and had the best DNA quality) was selected to be used for the genomic sequencing (individual OAG_MAG_041103, Electronic Supplementary Material (ESM) Fig. S1). DNA was first sheared using Covaris g-TUBE's following the manufacturer's protocol for 10 Kb fragments. Four individual Oxford Nanopore Technologies (ONT, Oxford, UK) libraries were prepared with 1.5 mg of DNA using the SQK-LSK109 library preparation protocol according to the manufacturer's protocol (Oxford Nanopore Technologies, Oxford, UK). One library was sequenced on a R9.4 flow cell, and three libraries were sequenced on R10.3 flow cells using the MinIon DNA sequencer (Mk1B). Maximum run time ranged between 48 and 72 h. Raw data were basecalled separately using Guppy 3.3 basecaller on a GPU-based high-performance computer cluster server.

In addition, we prepared an Illumina library with 250 ng of the same DNA using the Kapa Hyperplus Library Preparation Kit with only one third of the volume reactions as described in the manufacturer's protocol (Kapa Biosystems, Wilmington, MA). The total fragmentation volume was 16.6 ml and was incubated at 37 °C for 7:45 min. The incubation parameters were previously optimized to target fragments of ~ 500 bp. Post-ligation purification was done using a 0.8X KAPA Pure bead cleanup. Library amplification was carried out with a total PCR reaction volume of 16.6 ml for 8 PCR thermal cycles. Finally, we did a double size-selection post-amplification cleanup with SPRIselect beads using a 0.56X upper and 0.72X lower selection ratio (Beckman Coulter, Inc). The final Illumina library was sequenced with a HiSeq4000 (150 bp paired-end) (Novogene Corporation Inc.).

GenomeScope (Vurture et al. 2017) was used to estimate genome size, repeat content, and heterozygosity across all k-mers (k = 21) previously detected using Jellyfish v2.2.10 (Marçais and Kingsford 2011) to help choose parameters for downstream analysis.

2.4. Genome assembly

Long reads obtained from the ONT were concatenated into one large fastq file and trimmed with Porechop v. 0.2.3 (https://github.com/rrwick/Porechop). Nanofilt v. 2.5.0 (https://github.com/wdecoster/nanofilt) was used to create two different filtered data sets to help the contiguity of the final assembly. The first filtered data set was used to keep the longest reads and perform an initial more contiguous assembly (Nanofilt parameters -q 3; -l 1000). The second was explicitly used for downstream assembly polishing (-q 5 and -l 500). The former sequences were assembled using Wtdbg2 v2.5 (Ruan and Li, 2020), setting a minimum sequence length of 1000 bp (-L 1000). In order to improve the draft assembly, two rounds of consensus correction were performed using the -q 5 filtered ONT reads by mapping reads to the draft genome with Minimap2 v. 2.17 and polishing with Racon v. 1.4.7.

Short accurate Illumina reads were used to further polish the ONT genome. Raw sequences were adapter-trimmed with Trimmomatic v. 0.39 (Bolger et al., 2014), using only sequences with phred scores of 33, and with the settings LEADING:2 TRAILING:2 MINLEN:25; and quality checked before and after trimming using FastQC v 0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Two rounds of polishing were carried out by mapping the trimmed short reads to the assembly using BWA v 0.7.17 (Li et al., 2009) with default settings, sorted and indexed with Samtools v 1.9 (Li et al., 2009), and consensus corrected using Pilon v 1.23 (Walker et al., 2014).

Finally, the assembly was screened for sequences of bacteria, viruses, and plasmids using Kraken 2.0.9 (Wood and Salzberg, 2014), resulting in the removal of 0.5 % of the assembly. Genome completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO v3.0.2) (Simão et al., 2015; Waterhouse et al., 2018) by comparing the sequenced genome to the Actinopterygii (n = 4,584) ortholog gene datasets. Assembly statistics and BUSCO completeness were assessed after the initial draft assembly, and subsequently, after each polishing iteration.

2.5. Genome comparisons

The genome of *O. A. grahami* was compared to the genome of the Nile tilapia, *Oreochromis niloticus,* obtained from GenBank (Accession number GCA_000188235.2) (Brawand et al., 2014). Specifically, we used the UCSC Genome Browser assembly ID: oreNil2. Genomes were aligned using D-GENIES (Cabanettes and Klopp, 2018). Isochores were visualized for both genomes using the 'draw_chromosome_gc.pl' PERL package (Pačes et al., 2004). Following recommendations for eukaryotic (i.e. large) genomes, we used GC levels of each 100 kb non-overlapping sliding window. Color coded GC levels vary from < 32.5 %GC to > 55 %GC. Isochore boundaries have been defined as follows: L1 < 38 % GC, L2 38–42 % GC, H1 42–47 % GC, H2 47–52 % GC, and H3 > 52 % GC. Frequencies of genome regions according to their GC contents were obtained using the R-package SequinR (Charif and Lobry, 2007). Additionally, repetitive elements were predicted by running RepeatMasker (open-4.0.6, Smit et al., 2015) with the Teleostei database to identify repetitive elements in the genome and soft-mask the assembly. RepeatMasker.out was converted to GFF (General Feature Format) with RepeatMasker script rmOutToGFF3.pl.

3. Results

3.1. Mitochondrial DNA

Mitochondrial control regions of *Oreochromis Alcolapia* species from Genbank cluster in four clades with most samples clustering in two main clades that correspond to *O. A. grahami* and *O. A. alcalicus* (ESM Fig. S1). There are a few exceptions in the placement of described individuals in given clades, due to the complicated nature of the species complex, its young evolutionary age, and the molecular marker used here. A proper treatment of the genomics of the species complex goes beyond the scope of this paper, and has been discussed elsewhere (Seegers et al., 1999). Sequences from the eight samples collected from this study clustered with *O. A. grahami* (seven samples) and *O. A. alcalicus* (one sample) (ESM Fig. S1). We sequenced the genome of the OAG_MAG_041103 sample from the *O. A. grahami clade*. Once genomic sequencing was complete, the mitochondrial genome of this sample was bioinformatically extracted and deposited in GenBank (accession number: NC_070106.1). As expected, the control region sequence from the extracted mitochondrial genome is identical to the direct PCR amplified sequence obtained prior to the genomic sequencing (ESM Fig. S1).

3.2. The genome of Oreochromis Alcalicus grahami and its compositional characteristics

A total of 20.98 Gb (N50: 6560 bp, longest read: 474 205 bp) were generated on the Oxford Nanopore MinIon device. A total of 95.6 Gb, with an average cleaned read of 149 bp, were generated by Illumina genomic sequencing. The genome of *Oreochromis Alcalicus grahami* presented here (GenBank, JAQSNX000000000; BioSample SAMN33214940) had a contig N50 (the sequence length of the shortest contig at 50 % of the total assembly length) of 176.7 Kb and was fairly complete with a Complete BUSCO score of 94.9 % (Complete and Single copy BUSCOs, S = 93.9 %; Complete and duplicated BUSCOs, D = 1.0 %;

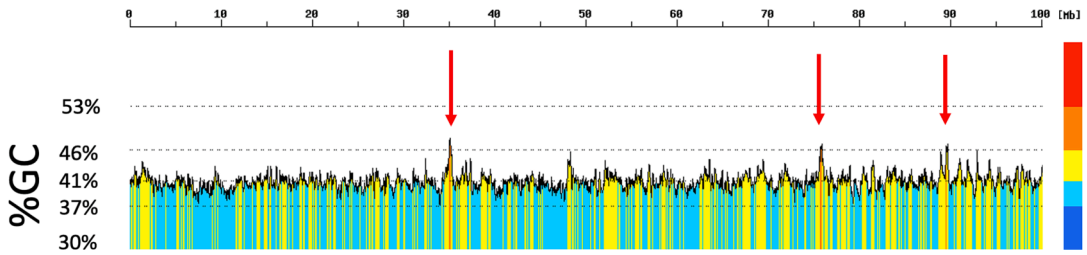

Fig. 3. Isochores of the genome of *Oreochromis Alcolapia grahami* using a 100 kb non-overlapping sliding window. This figure depicts the first 100 Mb of the genome, Electronic Supplementary Material Figure S3 shows the entire genome. Isochores (GC levels) are divided in five different colors from dark blue (32.5 % GC) to red (>55 % GC) (color bar is at the bottom right) using the method of Paces et al. (2004). Examples of the GC-richest isochores are indicated by red arrows.

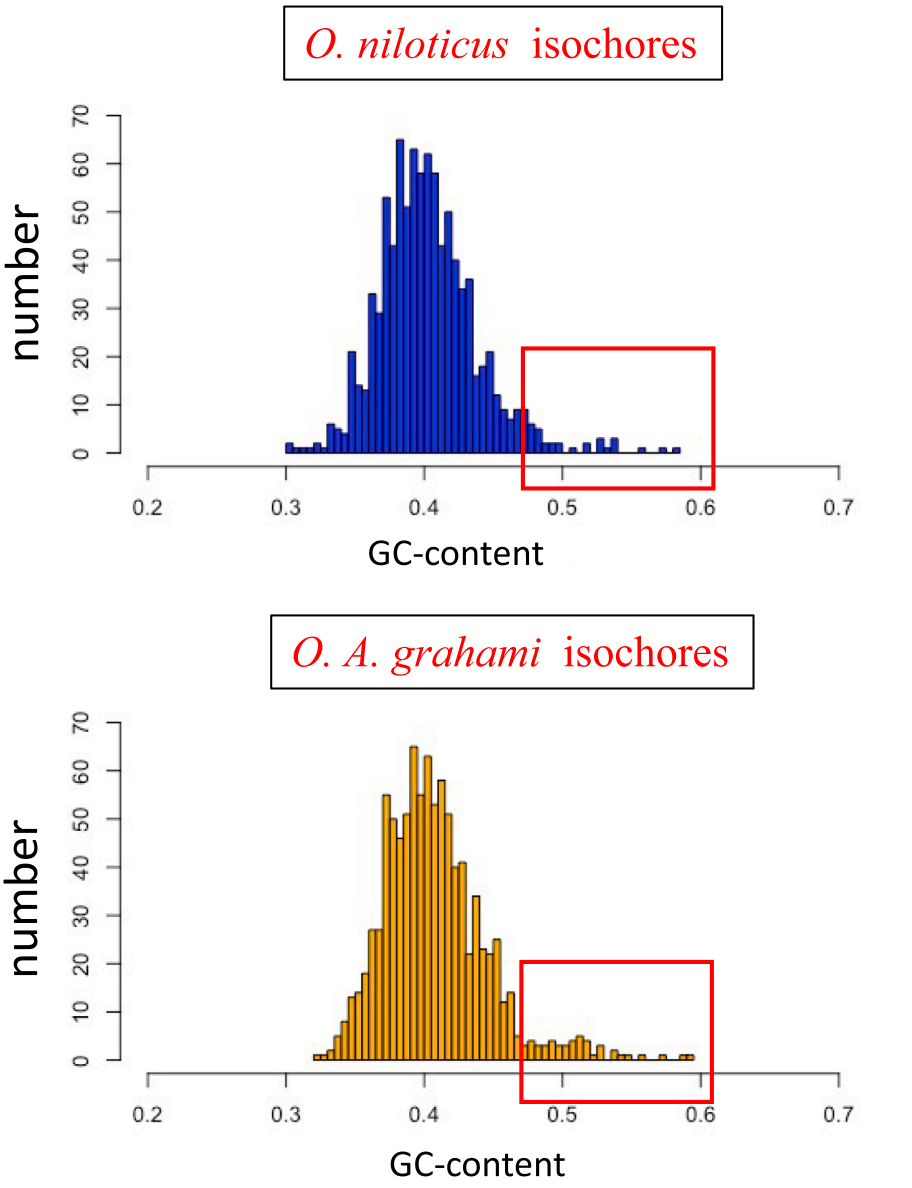

Fig. 4. Histograms representing the cumulative amounts of isochores of the genomes of *Oreochromis niloticus* and *Oreochromis Alcolapia grahami.* The red rectangles highlight the GC-rich region of the genome for both species.

Fragmented BUSCOs, F = 1.3 %; Missing BUSCOs, M = 3.8 %) (N50 = 176,105 bp, n = 1444, largest contig 1,444,179 bp).

The genome size of *O. A. grahami* was estimated to have a length of 892 Mb with a heterozygosity level of 0.43 %. Considering a genome size of 892 Mb, the output of 20.98 Gb of Oxford Nanopore Technologies and 95.6 Gb of Illumina reads represented a total of 23.5X and 107X coverage respectively, based on the size of our final genome assembly. The assembled genome size was estimated to be 892.4 Mb, which is

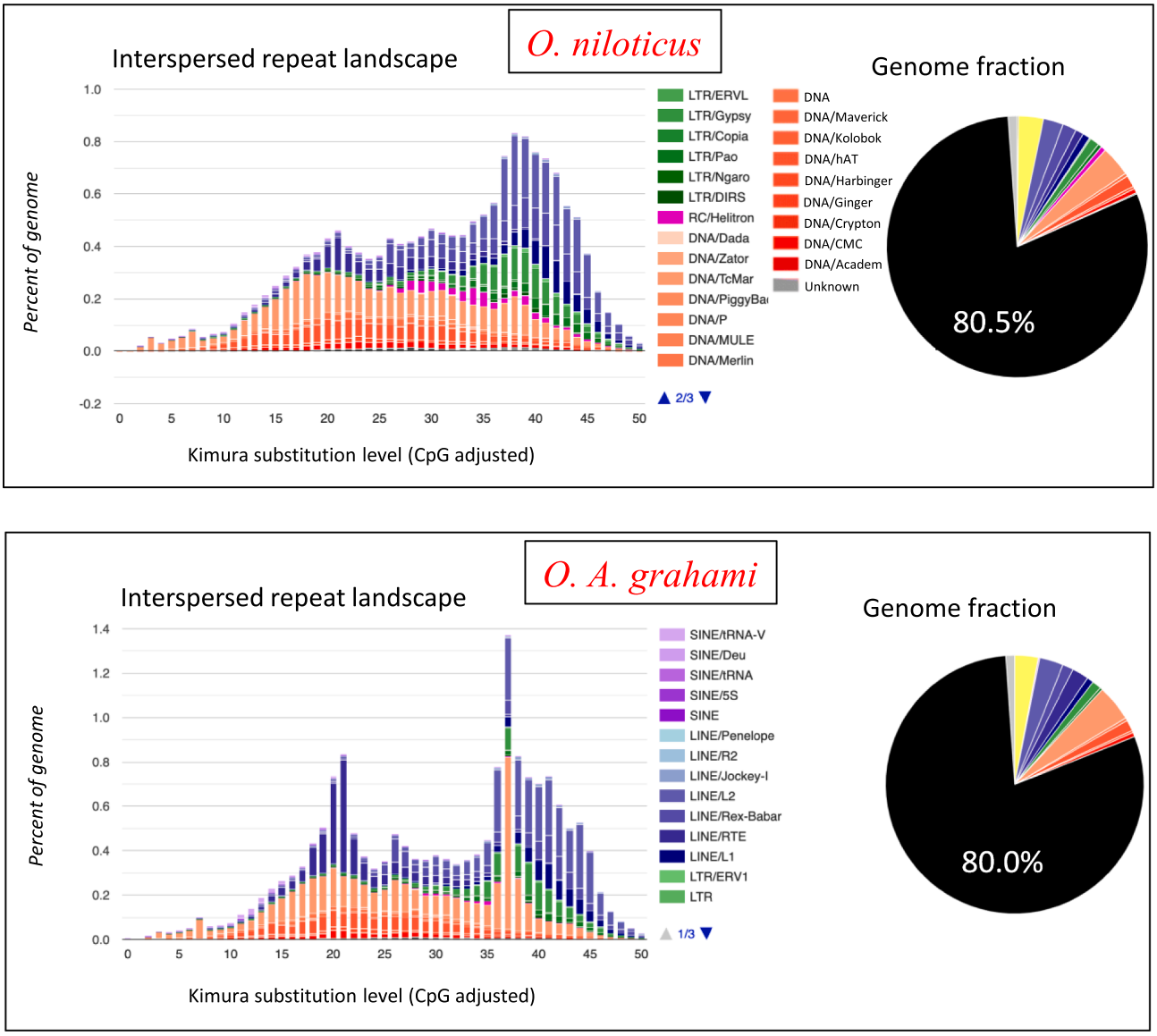

Fig. 5. Repeat Landscape comparison of *Oreochromis niloticus* and *Oreochromis Alcolapia grahami*. The percentage of both the *O. niloticus* and *O. A. grahami* and assemblies that each Transposable Element (TE) family is represented at in particular substitution levels analogous to the age of TEs (Kimura substitution level – CpG adjusted).

similar to the size of the genomes of close relatives, *O. aureus* (920 Mb) and *O. niloticus* (1010 Mb) (Bian et al., 2019). The DNA base composition was 41.07 % GC, which is similar to what had been predicted based on CsCl buoyant density gradients (42 % GC) (Bucciarelli et al., 2002), and similar again to *O. aureus* (41.6 % GC) and *O. niloticus* (41.7 % GC) (Bucciarelli et al., 2002).

The genomes of *O. A. grahami* and *O. niloticus* aligned very well (ESM Fig. S2) and the compositional organization was obtained for both species (detailed map of *O. A. grahami* is shown in Fig. 3 and ESM Fig. S3). As in previous results obtained using CsCl gradients, the GC rich portion of the genome was greater in *O. A. grahami* than in *O. niloticus* (Fig. 4). Indeed, 59.3 % of the genome of *O. A. grahami* had a base composition higher than 40 %GC, as opposed to 55.3 % of the genome being higher than 40 %GC in *O. niloticus*. This was also reflected at higher GC values, where 12.8 % and 3.3 % of the genome of *O. A. grahami* was higher than 45 % GC and 50 % GC, respectively, while these values were 10.2 % and 1.6 % for *O. niloticus*. In other words, the genome of *O. A. grahami* was found to be more skewed toward GC rich regions than the genome of *O. niloticus* (Fig. 4). At a larger scale, GC-poor isochore families L1 + L2 (comprising DNA < 42 % GC) represented 65.67 % of the genome of *O. niloticus* and 51.53 % of the genome of *O. A. grahami*, while the GC-rich isochore families H1 + H2 (comprising DNA between 42 % and 52 % GC) represented 34.33 % and 48.47 % of the genomes of *O. niloticus* and *O. A. grahami,* respectively.

3.3. Repetitive sequences

RepeatMasker estimated that 20 % of the genome of *Oreochromis Alcalicus grahami* and 19.5 % of the genome of *O. niloticus* consisted of repetitive sequences (Fig. 5). These were primarily LINEs, and DNA transposons (Fig. 5). The number of repetitive sequences was therefore very similar in *O. A. grahami* and *O. niloticus*, and the timing of their evolutionary origin (Kimura substitution level) was also similar (Fig. 5). One outlier was present in *O. A. grahami,* where an increase of DNA transposons (Tc1/mariner) that account for 0.8 % of the genome are found at Kimura substitution level 37 (Fig. 5).

4. Discussion

While the compositional organization of the genome of vertebrates in isochore families has been known for several decades (Thiery et al., 1976), its significance and evolution remains an open question. Empirical evidence has been accumulating (Bernardi, 2021), yet the mechanisms and causes to explain the difference between poikilotherms and endotherms are still debated. In fishes, where most species show homogeneous AT-rich genomes, the presence of species that live at high temperature may provide a window into the genesis of GC-rich isochores.

In this study, we revisited early findings that were based on CsCl density gradients of *Oreochromis Alcolapia grahami*, a hot-water adapted species that holds the world record for temperature tolerance (Ctmax = 45.6 °C) in a teleost fish (Wood et al., 2016). CsCl gradients showed that some portion of the genome was GC-rich, a novelty for fishes and in particular in comparison with a 'cooler' closely related species. Findings presented here based on full genome sequencing were consistent with the early findings, where GC-rich regions, present in the genome of *O. A. grahami* are absent from the genome of the closely related and cooler living *O. niloticus.* The genome of *O. A. grahami* was found to be more skewed toward GC-rich regions than the genome of *O. niloticus* (Fig. 4). This was also originally observed with CsCl gradients, where skewness is identified as genome heterogeneity (*O. A. grahami* H = 3.6*, O.niloticus* H = 2.0) (Bucciarelli et al., 2002). The genomic sequence of *O. A. grahami* provided additional information that was not available in the original CsCl studies, however. First, the novel GC-rich regions are unlikely to be due to a mere increase in repetitive sequences. The genome of *O. A. grahami* is in fact slightly smaller than the genome of *O. niloticus,* and repetitive sequences in those two species account for a similar proportion of the genome (approximately 20 % of the genome). Importantly, however, we do not know the base composition of the repetitive sequences, so it might be possible that repetitive sequences in Lake Magadi tilapia are consistently more GC rich than the cool-water *O. niloticus* counterpart. Also, while the amount of repetitive sequences is the same, their relative contribution to GC richness might be different. Second, the GC-richest portions of the genome of *O. A. grahami* are interspersed throughout the genome (Fig. 3), thus ruling out the presence of a small GC-rich region of the genome or the presence of a GC-rich mini-chromosomes.

GC-rich regions of mammalian and bird genomes have been shown to encompass intergenic, intronic, as well as protein coding regions (Bernardi, 2021). At this point, it is not known if the GC-rich regions in the genome of *O. A. grahami* correspond exclusively to non-coding regions, or if they comprise protein coding genes. Additional analyses of the genome, together with the sequencing of more genomes from warm adapted fishes such as the Death Valley pupfishes (genus *Cyprinodon*), warm adapted cichlids (genus *Danakilia*), or intertidal gobies (genus *Gillichthys*) are likely to provide essential information as to the evolution and origin of GC-rich isochores.

CRediT authorship contribution statement

Giacomo Bernardi: Conceptualization, Data curation, Writing – original draft, Writing – review and editing, Formal Analysis, Methodology. **Geraldine D. Kavembe:** Conceptualization, Writing – review and editing. **Harold L. Bergman:** Conceptualization, Writing – review and editing. **Giuseppe Bucciarelli:** Conceptualization, Writing – review and editing. **Chris M. Wood:** Conceptualization, Writing – review and editing.

Acknowledgements

We would like to dedicate this work to the memory of Giorgio Bernardi, a pioneer in isochore research, who first proposed the thermal theory of isochore evolution. CMW's work is funded by the NSERC (Canada) Discovery Grant Program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jglr.2024.102326.

References

Bernardi, G., 2000. Isochores and the evolutionary genomics of vertebrates. Gene 241, 3–17. https://doi.org/10.1016/S0378-1119(99)00485-0.
Bernardi, G., 2021. The "genomic code": DNA pervasively moulds chromatin structures leaving no room for "junk". Life 11, 342. https://doi.org/10.3390/life11040342.
Bernardi, G., Bernardi, G., 1986. Compositional constraints and genome evolution. Journal of Molecular Evolution 24, 1–11.
Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., Rodier, F., 1985. The mosaic genome of warm-blooded vertebrates. Science, New Series 228, 953–958.
Bian, C., Li, J., Lin, X., Chen, X., Yi, Y., You, X., Zhang, Y., Lv, Y., Shi, Q., 2019. Whole genome sequencing of the blue tilapia (Oreochromis aureus) provides a valuable genetic resource for biomedical research on tilapias. Marine Drugs 17, 386. https://doi.org/10.3390/md17070386.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.-F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O., Di Palma, F., 2014. The genomic substrate for adaptive radiation in african cichlid fish. Nature 513, 375–381. https://doi.org/10.1038/nature13726.
Bucciarelli, G., Bernardi, G., Bernardi, G., 2002. An ultracentrifugation analysis of two hundred fish genomes. Gene 295, 153–162. https://doi.org/10.1016/S0378-1119(02)00733-3.
Butzer, K.W., Isaac, G.L., Richardson, J.L., Washbourn-Kamau, C., 1972. Radiocarbon dating of east african Lake levels: new observations provide fresh insights into late Quaternary paleoclimates. Science 175, 1069–1076. https://doi.org/10.1126/science.175.4026.1069.
Cabanettes, F., Klopp, C., 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6, e4958.
Charif, D., Lobry, J., 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Structural Approaches to Sequence Evolution: Molecules, Networks, Populations, Biological and Medical Physics, Biomedical Engineering. Springer Verlag, New York, pp. 207–232.
Coe, M.J., 1966. The biology of "Tilapia grahami" boulenger in Lake magadi. Kenya. https://doi.org/10.5169/SEALS-311342.
Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008.
Domingues, V.S., Bucciarelli, G., Almada, V.C., Bernardi, G., 2005. Historical colonization and demography of the Mediterranean damselfish, CHROMIS CHROMIS: CHROMIS CHROMIS SPECIATION. Molecular Ecology 14, 4051–4063. https://doi.org/10.1111/j.1365-294X.2005.02723.x.
Duret, L., Galtier, N., 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genom. Hum. Genet. 10, 285–311. https://doi.org/10.1146/annurev-genom-082908-150001.
Eyre-Walker, A., Hurst, L.D., 2001. The evolution of isochores. Nat Rev Genet 2, 549–555. https://doi.org/10.1038/35080577.
Ford, A.G.P., Bullen, T.R., Pang, L., Genner, M.J., Bills, R., Flouri, T., Ngatunga, B.P., Rüber, L., Schliewen, U.K., Seehausen, O., Shechonge, A., Stiassny, M.L.J., Turner, G. F., Day, J.J., 2019. Molecular phylogeny of oreochromis (cichlidae: oreochromini) reveals Mito-nuclear discordance and multiple colonisation of adverse aquatic environments. Molecular Phylogenetics and Evolution 136, 215–226. https://doi.org/10.1016/j.ympev.2019.04.008.
Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159, 907–911. https://doi.org/10.1093/genetics/159.2.907.
Johannsson, O.E., Bergman, H.L., Wood, C.M., Laurent, P., Kavembe, D.G., Bianchini, A., Maina, J.N., Chevalier, C., Bianchini, L.F., Papah, M.B., Ojoo, R.O., 2014. Air breathing in magadi tilapia Alcolapia grahami, under normoxic and hyperoxic conditions, and the association with sunlight and reactive oxygen species: air breathing in alcolapia grahami. J Fish Biol 84, 844–863. https://doi.org/10.1111/jfb.12289.

Kavembe, G.D., Franchini, P., Irisarri, I., Machado-Schiaffino, G., Meyer, A., 2015. Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a cichlid fish from one of earth's Most extreme environments, the hypersaline soda Lake magadi in Kenya, East Africa. J Mol Evol 81, 90–109. https://doi.org/10.1007/s00239-015-9696-6.

Kavembe, G.D., Kautt, A.F., Machado-Schiaffino, G., Meyer, A., 2016. Eco-morphological differentiation in Lake magadi tilapia, an extremophile cichlid fish living in hot, alkaline and hypersaline lakes in East Africa. Mol Ecol 25, 1610–1625. https://doi.org/10.1111/mec.13461.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Nivelle, R., Gennotte, V., Kalala, E.J.K., Ngoc, N.B., Muller, M., Mélard, C., Rougeot, C., 2019. Temperature preference of Nile tilapia (Oreochromis niloticus) juveniles induces spontaneous sex reversal. PLoS ONE 14, e0212504.

Owen, R.B., Muiruri, V.M., Lowenstein, T.K., Renaut, R.W., Rabideaux, N., Luo, S., Deino, A.L., Sier, M.J., Dupont-Nivet, G., McNulty, E.P., Leet, K., Cohen, A., Campisano, C., Deocampo, D., Shen, C.-C., Billingsley, A., Mbuthia, A., 2018. Progressive aridification in East Africa over the last half million years and implications for human evolution. Proc. Natl. Acad. Sci. u.s.a. 115, 11174–11179. https://doi.org/10.1073/pnas.1801357115.

Paĉes, J., Zíka, R., Paĉes, V., Pavlíĉek, A., Clay, O., Bernardi, G., 2004. Representing GC variation along eukaryotic chromosomes. Gene 333, 135–141. https://doi.org/10.1016/j.gene.2004.02.041.

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Ruan, J., Li, H., 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods 17, 155–158. https://doi.org/10.1038/s41592-019-0669-3.

Seegers, L., Sonnenberg, R., Yamamoto, R., 1999. Molecular analysis of the alcolapia flock from lakes Natron and magadi, Tanzania and Kenya (teleostei: cichlidae), and implications for their systematics and evolution. Ichthyological Exploration of Freshwaters 10, 175–199.

Seegers, L., Tichy, H., 1999. The oreochromis alcalicus flock (teleostei: cichlidae) from lakes Natron and magadi, Tanzania and Kenya, with descriptions of two new species. Ichthyological Exploration of Freshwaters 10, 97–146.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Smit, AFA, Hubley, R & Green, P. 2013-2015. *RepeatMasker Open-4.0*. <http://www.repeatmasker.org>.

Thiery, J.-P., Bernardi, G., Macaya, G., 1976. An analysis of eukaryotic genomes by density gradient centrffugation. Journal of Molecular Biology 108, 219–235.

Tichy, H., Seegers, L., 1999. The oreochromis alcalicus flock (teleostei: cichlidae) from lakes Natron and magadi, Tanzania and Kenya: a model for the evolution of "new" species flocks in historical times? Ichthyological Exploration of Freshwaters 10, 147–174.

Trewavas, E., 1983. Tilapiine fishes of the genera sarotherodon, oreochromis, and danakilia. british museum. Natural History.

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

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution 35, 543–548. https://doi.org/10.1093/molbev/msx319.

Wilson, P.J., Wood, C.M., Maina, J.N., White, B.N., 2000. Genetic structure of Lake magadi tilapia populations. Journal of Fish Biology 56, 590–603. https://doi.org/10.1111/j.1095-8649.2000.tb00757.x.

Wilson, P.J., Wood, C.M., Walsh, P.J., Bergman, A.N., Bergman, H.L., Laurent, P., White, B.N., 2004. Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake magadi tilapia. Physiological and Biochemical Zoology 77, 537–555. https://doi.org/10.1086/422054.

Wood, C.M., Bergman, H.L., Bianchini, A., Laurent, P., Maina, J., Johannsson, O.E., Bianchini, L.F., Chevalier, C., Kavembe, G.D., Papah, M.B., Ojoo, R.O., 2012. Transepithelial potential in the magadi tilapia, a fish living in extreme alkalinity. J Comp Physiol B 182, 247–258. https://doi.org/10.1007/s00360-011-0614-y.

Wood, C.M., Nawata, C.M., Wilson, J.M., Laurent, P., Chevalier, C., Bergman, H.L., Bianchini, A., Maina, J.N., Johannsson, O.E., Bianchini, L.F., Kavembe, G.D., Papah, M.B., Ojoo, R.O., 2013. Rh proteins and NH4+-activated Na+-ATPase in the magadi tilapia (Alcolapia grahami), a 100% ureotelic teleost fish. Journal of Experimental Biology 216, 2998–3007. https://doi.org/10.1242/jeb.078634.

Wood, C.M., Brix, K.V., De Boeck, G., Bergman, H.L., Bianchini, A., Bianchini, L.F., Maina, J.N., Johannsson, O.E., Kavembe, G.D., Papah, M.B., Letura, K.M., Ojoo, R. O., 2016. Mammalian metabolic rates in the hottest fish on earth. Sci Rep 6, 26990. https://doi.org/10.1038/srep26990.

Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15, R46. https://doi.org/10.1186/gb-2014-15-3-r46.