A multi-marker perspective on the evolutionary history of East African cichlid fishes

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Britta Silke Meyer
aus
Hamburg, Deutschland

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Walter Salzburger, Prof. Dr. Markus Pfenninger
(Mitglieder des Dissertationskomitees: Fakultätsverantwortlicher, Korreferent)

Basel, den
10.12.2013
(Datum der Genehmigung durch die Fakultät)

Prof. Dr. Jörg Schibler
(Dekan)

Creative Commons

Creative Commons License Deed

Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

This is a human-readable summary of (and not a substitute for) the license.
Disclaimer

You are free to:

Share — copy and redistribute the material in any medium or format

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

The applicable mediation rules will be designated in the copyright notice published with the work, or if none then in the request for mediation. Unless otherwise designated in a copyright notice attached to the work, the UNCITRAL Arbitration Rules apply to any arbitration.

Für meine Liebsten

Contents

Introduction

The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth. The green and budding twigs may represent existing species; and those produced during each former year may represent the long succession of extinct species. At each period of growth all the growing twigs have tried to branch out on all sides, and to overtop and kill the surrounding twigs and branches, in the same manner as species and groups of species have tried to overmaster other species in the great battle for life. The limbs divided into great branches, and these into lesser and lesser branches, were themselves once, when the tree was small, budding twigs; and this connexion of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups. Of the many twigs which flourished when the tree was a mere bush, only two or three, now grown into great branches, yet survive and bear all the other branches; so with the species which lived during long-past geological periods, very few now have living and modified descendants. From the first growth of the tree, many a limb and branch has decayed and dropped off; and these lost branches of various sizes may represent those whole orders, families, and genera which have now no living representatives, and which are known to us only from having been found in a fossil state. As we here and there see a thin straggling branch springing from a fork low down in a tree, and which by some chance has been favoured and is still alive on its summit, so we occasionally see an animal like the Ornithorhynchus or Lepidosiren, which in some small degree connects by its affinities two large branches of life, and which has apparently been saved from fatal competition by having inhabited a protected station. As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications.

On the Origin of Species, Charles Darwin

Charles Darwin's book "On the Origin of Species" changed the vision on the diversity of life[1]. He poetically described the "great Tree of Life", formed by all extant and extinct species. His ideas about the processes of diversification within and among different lineages and their common ancestry provided a new framework to taxonomic and systematic studies – that of evolution – ultimately leading to the field of phylogenetics, which focuses on the relationships between different taxa on the basis of heritable factors. In molecular phylogenetics, molecular characters are used to draw the twigs and branches of the tree of life. The source of this information is manifold and can originate from amino acids within proteins or from nucleotides within mitochondrial or nuclear DNA and from different RNA molecules. Variability within these molecules is used to infer similarity and thus common ancestry. Besides the sequence information of the character states itself, specific length variations such as indels or duplications and rearrangements within genomes can be used to infer relationships, as these events are supposed to be rare. Other commonly used genetic markers, which can differ between taxa and thus are valuable for phylogenetic inferences, are for example

allozymes[2], single nucleotide polymorphisms (SNPs)[2, 3], amplified fragment length polymorphisms (AFLPs)[2] or restriction site associated DNA (RAD) markers [4].

The available methods for phylogenetic reconstruction are among the most powerful tools in biological research and widespread in nearly all fields of biology. Molecular phylogenetics is an integral part in systematics and taxonomic research investigating the relationships within the tree of life[5-8]; in phylogeography, which investigates the distribution of biological entities in the context of geography[3, 9, 10]; in the field of molecular evolution exploring the history of genomes and gene-families or single genes thereby detecting signals of selection[11-13]; in the field of evolutionary developmental biology helping to determine homologous characters, to estimate ancestral states and thus to trace back character evolution[14]; in palaeobiology reconstructing the history of ancient life forms and molecules[11, 15-17]; and also, in medicine when it comes to the understanding of the behavior of certain cell lineages and tumor development[18-22] or to the search of the sources of epidemiological outbreaks[23, 24]. Strong and reliable phylogenetic hypotheses are thus crucial for various biological questions, especially those asked by comparative and evolutionary biologists.

Besides the application of phylogenetic approaches, phylogenies themselves are a fascinating and rapidly developing research area. In the last three decades, the algorithms and software packages have co-evolved with the kind of questions asked and with the increasing amount of data available[25-28]. Among the fields' recent (methodological) trends are: (i) reducing alignment and assembly errors[29-32], (ii) improving the strategies of accessing appropriate substitution models for the data and the golden way of its partitioning[33-35]; and (iii) the enhancement of the "traditional" phylogenetic inference methods as the Maximum likelihood-based[36, 37] and the Bayesian approaches[38, 39]. The relatively young era of phylogenomics, a research field at the intersection of genomics and phylogenetics, which arose with the availability of genome or transcriptome data, led to a new range of questions and insights. Along with the long-lasting debates of which markers to use came the awareness of the discordance among gene trees and the necessity to acquire several (to hundreds) of markers to reliably infer evolutionary history[40-43]. A range of software packages for species tree estimation from sequence or biallelic markers[38, 40, 44-46] as well as from gene tree estimates[47, 48] were developed, besides using a concatenated supermatrix and standard phylogenetic inferences[36-39]. In the concatenation approach, a supermatrix is generated from several concatenated gene alignments and thereupon treated as a single gene, which is than used to infer the species tree. Combining concordant gene trees, this produces very accurate results. Whereas the first-mentioned species tree approaches are based on coalescent theory[38, 40, 44-46]. Yet, as phylogenomics is still in its infancy, it remains unclear how many and which loci should be analyzed to adequately

reflect the evolutionary history of a taxon and to examine how the individual signals from the genes and alleles influence the (species) tree estimate[41, 49, 50]. In addition, it is largely unknown whether or not the concatenation approach is always capable to redraw the species history[50, 51], how evolutionary processes such as incomplete lineage sorting and hybridization can be integrated [52], and how non-phylogenetic signals, originating from the uncertainty in homology, errors in the alignments and the inappropriate model choice, should be handled[49, 53, 54]. Systems with short branch lengths (in coalescent units) and big population sizes are particularly challenging, as they encompass a high degree of incomplete lineage sorting and, consequently, a profound amount of gene tree discordance, making them prone to inconsistencies in phylogenetic inferences. Ancient and recent adaptive radiations represent examples where genes splits drag behind species splits[43, 55, 56].

The East African cichlids, more precisely the cichlid fishes of Lake Tanganyika, Lake Victoria and Lake Malawi, are among the most famous textbook examples of adaptive radiations[57-62]. Both hybridization and incomplete lineage sorting explain the high degree of shared gene lineages within these species-flocks[63-68].

Considerable effort has been put into the understanding of the relationships between and among the main lineages as this is essential to establish the phylogenetic backbone of the East African cichlid radiations, which in turn is crucial to formulate and validate hypotheses about the patterns and processes underlying this unparalleled species diversification. So far, we know that the radiations within the three Great Lakes started at different time points and with a different set of seeding lineages; the radiations thus differ with respect to the number and diversity of species. The oldest of the lakes, Lake Tanganyika, harbors around 250 species from 12 to 16 different cichlid lineages, whereas the younger Lakes Malawi and Victoria contain a recent radiation of one lineage, namely the haplochromines[58, 66, 69-77]. These two lakes are home to more than 1,200 cichlid species. Lake Tanganyika, although comprising less variety in number of species, is considered as the source of the East African cichlid radiation, as the modern haplochromines in Lake Malawi and Lake Victoria are derived from the Lake Tanganyikan haplochromines. Hence, Lake Tanganyika cichlids play a crucial role for the onset of the two other enormous radiations[70]. However, the mainly mitochondrial DNA based phylogenies for the Lake Tanganyika cichlids lack resolution for some lineages as support values are low and phylogenetic positions are not consistent within the different tree estimates[58]. It is unknown if this lack of resolution can be only found on the level of single gene trees or if it is reflected on the species level as a real biological polytomy. In my thesis I aimed to generate a solid phylogenetic framework for the cichlid tribes of Lake Tanganyika and hence to get insights to macroevolutionary processes. I developed a broad range of primers and

applied next generation sequencing methods, resulting in the most comprehensive sequence-based multi-marker set for Lake Tanganyikan cichlids so far. Further I applied phylogenetic inferences and other methods based on the theory of molecular evolution to phylogeographic and other evolutionary questions (e.g. gene diversity).

The first two chapters of this thesis were driven by the question if more information from different genes would help to resolve the phylogenetic relationship of the major tribes within Lake Tanganyika. In the first chapter "A novel primer set for multilocus phylogenetic inference in East African cichlid fishes"[78] I describe a new PCR primer designing strategy for multi-locus phylogenetic inferences in East African cichlids and make available 24 of such primers. I further demonstrate the successful amplification of these markers with Sanger sequencing in a wide range of cichlid species and provide first insights regarding the power of resolution for this marker set with respect to phylogenetic inference. The designing phase and experiments in the laboratory were conducted at the time when no cichlid genomes were available. I later extended the marker set from the first chapter with the same requirements to 45 markers. However, instead of Sanger sequencing, I generated bar-coded fusion primers for Roche`s 454 pyrosequencing technology (see Appendix 1) for multiplexing reactions[79]. These primers were mixed in groups of eight or respectively ten primer pairs (Appendix 2) and then used in multiplex polymerase chain reactions (Appendix 3) for about eight to sixteen individuals per species. These amplicons were then sequenced unidirectional starting from the forward primer.

In the second chapter "A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach" I applied the nuclear marker set from the first study[78] and took a subset of sequences of the large pyrosequencing experiment described in the paragraph above. I picked one individual per species, for which most of the markers were available. In total this data set consisted of 42 loci from 45 species. These 1890 sequences of a length of nearly 18,000 bp were used to conduct phylogenetic analyses, both in a concatenated dataset including all markers and with Bayesian concordance analysis. The first method assumes that all gene trees are reflected in one primary history, the species tree, whereas the latter approach accounts for possible gene tree discordance. Based on this unparalleled dataset we could propose new hypotheses for the evolutionary history of the East African cichlids.

In the third chapter "Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes" we report the occurrence of a new cichlid species, found in Lake Tanganyika. Using two nuclear and two mitochondrial markers and conducting different phylogenetic analyses, we detected its genetic affiliation to the haplochromine cichlids of the Lake Victoria superflock. Until recently it was

thought that the Great East African Lakes diversified independently after the seeding of these radiations. However, a large SNP-screen in East African cichlids conducted by Loh et al.[77] showed an astonishing amount of shared genetic polymorphisms among the Lakes. It was hypothesized that this high level of shared SNPs arose due to convergent mutations and also by "riverine transporter" taxa, that migrated between lakes. With the discovery of this modern haplochromine species within Lake Tanganyika, we provide the first case of a recent invasion of a species belonging to a lineage associated to one of the other Great Lakes and thus strengthen the hypotheses described by Loh et al.[77].

In the fourth chapter, "The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change", I contributed to the study of the molecular basis of the egg spots, which are a novel phenotypic color trait within the most species-rich lineage of the haplochromines[60]. To investigate the underlying genetics of the egg spots it was crucial to have a proper phylogenetic hypothesis at hand. To this end, I generated a new phylogeny of the haplochromines, thus providing a solid basis for the assumptions of the trait emergence and consequently the experiments. I used a subset of the sequences from the second chapter (9 nuclear markers from 12 species) and other new sequences from *Thoracochromis brauschi*, *Serranochromis macrocephalus* and *Astatoreochromis alluaudi*, in addition to one mitochondrial marker and sequences from the genomes of *Maylandia zebra*, *Oreochromis niloticus* and *Neolamprologus brichardi*.

In the last chapter I present ongoing work on "The role of parasites and the immune system in the adaptive radiation of Lake Tanganyika cichlids". We explore the covariations of trophic morphology, trophic level, diet, body shape, macro-parasitism and MHC genes to investigate the contribution of parasite-mediated selection in this adaptive radiation. Besides parasitological screening, one barcoded primer pair was used to amplify MHC class II loci (both intron 1 and exon 2) in several individuals of 39 taxa. We showed that the trophic-morphological axis of diversification in Lake Tanganyika cichlids is strongly correlated with infection levels of metazoan macroparasites. We further detect correlation between these parasites and MHC constitution. This gives insights to the potential influence of parasitism and immunogenetic adaptations to the Lake Tanganyika cichlid radiation.

Finally, in the last section, I discuss the results obtained and briefly suggest future directions.

1. Darwin, C. The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life (John Murray, London, UK, 1859).

2. Avise, J.C. Molecular Markers: Natural History and Evolution (Chapman & Hall, 1994).

3. Brito, P. & Edwards, S. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135, 439-455 (2009).

4. Rubin, B.E.R., Ree, R.H. & Moreau, C.S. Inferring Phylogenies from RAD Sequence Data. PLoS ONE 7, e33394 (2012).

5. Engel, P. et al. Parallel Evolution of a Type IV Secretion System in Radiating Lineages of the Host-Restricted Bacterial Pathogen Bartonella. PLoS Genet 7, e1001296 (2011).

6. Betancur, R.R. et al. The tree of life and a new classification of bony fishes. PLoS Curr 5 (2013).

7. McCormack, J.E. et al. A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment and High-Throughput Sequencing. PLoS ONE 8, e54848 (2013).

8. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature advance online publication (2013).

9. Ronquist, F. & Sanmartín, I. Phylogenetic Methods in Biogeography. Annual Review of Ecology, Evolution, and Systematics 42, 441-464 (2011).

10. Mateus, C.S. et al. Strong genome-wide divergence between sympatric European river and brook lampreys. Current Biology 23, R649-R650 (2013).

11. Chang, B.S. Ancestral gene reconstruction and synthesis of ancient rhodopsins in the laboratory. Integr Comp Biol 43, 500-7 (2003).

12. Diepeveen, E.T., Kim, F.D. & Salzburger, W. Sequence analyses of the distal-less homeobox gene family in East African cichlid fishes reveal signatures of positive selection. BMC Evol Biol 13, 153 (2013).

13. Böhne, A., Heule, C., Boileau, N. & Salzburger, W. Expression and sequence evolution of aromatase cyp19a1 and other sexual development genes in East African cichlid fishes. Mol Biol Evol 30, 2268-85 (2013).

14. Telford, M.J. & Budd, G.E. The place of phylogeny and cladistics in Evo-Devo research. Int J Dev Biol 47, 479-90 (2003).

15. Lorenzen, E.D. et al. Species-specific responses of Late Quaternary megafauna to climate and humans. Nature 479, 359-64 (2011).

16. Orlando, L. et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499, 74-78 (2013).

17. Foote, A.D. et al. Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. Nat Commun 4, 1677 (2013).

18. Salipante, S.J. & Horwitz, M.S. A phylogenetic approach to mapping cell fate. Curr Top Dev Biol 79, 157-84 (2007).

19. Frank, S.A. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proc Natl Acad Sci U S A 107 Suppl 1, 1725-30 (2010).

20. Schwartz, R. & Shackney, S.E. Applying unmixing to gene expression data for tumor phylogeny inference. BMC Bioinformatics 11, 42 (2010).

21. Peng, W. Tumor phylogenetics. Nat Genet 44, 368-368 (2012).

22. Bernard, H.U., Calleja-Macias, I.E. & Dunn, S.T. Genome variation of human papillomavirus types: phylogenetic and medical implications. Int J Cancer 118, 1071-6 (2006).

23. Smith, G.J. et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459, 1122-5 (2009).

24. Sharp, P.M. & Hahn, B.H. The evolution of HIV-1 and the origin of AIDS. Philos Trans R Soc Lond B Biol Sci 365, 2487-94 (2010).

25. Suárez-Díaz, E. & Anaya-Muñoz, V.H. History, objectivity, and the construction of molecular phylogenies. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 39, 451-468 (2008).

26. McCormack, J., Hird, S. & Zellmer..., A. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular phylogenetics ... (2013).

27. Rannala, B. & Yang, Z. Phylogenetic inference using whole genomes. Annual review of genomics and human genetics 9, 217-231 (2008).

28. Lemmon, E.M. & Lemmon, A.R. High-Throughput Genomic Data in Systematics and Phylogenetics. Annual Review of Ecology, Evolution, and Systematics (2012).

29. Katoh, K. & Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30, 772-80 (2013).

30. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589-595 (2010).

31. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat Meth 9, 357-359 (2012).

32. Wallace, I.M., O'Sullivan, O., Higgins, D.G. & Notredame, C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Research 34, 1692-1699 (2006).

33. Posada, D. jModelTest: phylogenetic model averaging. Mol Biol Evol 25, 1253-6 (2008).

34. Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29, 1695-701 (2012).

35. Brown, J.M. & Lemmon, A.R. The Importance of Data Partitioning and the Utility of Bayes Factors in Bayesian Phylogenetics. Systematic Biology 56, 643-655 (2007).

36. Stamatakis, A., Ludwig, T. & Meier, H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21, 456-63 (2005).

37. Zwickl, D.J. (The University of Texas at Austin, 2006).

38. Heled, J. & Drummond, A.J. Bayesian Inference of Species Trees from Multilocus Data. Molecular Biology and Evolution 27, 570-580 (2010).

39. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61, 539-42 (2012).

40. Edwards, S.V. Is a new and general theory of molecular systematics emerging? Evolution 63, 1-19 (2009).

41. Leigh, J.W., Susko, E., Baumgartner, M. & Roger, A.J. Testing congruence in phylogenomic analysis. Systematic Biology 57, 104-15 (2008).

42. Baum, D. Concordance Trees, Concordance Factors, and the Exploration of Reticulate Genealogy. Taxon 56, 417-426 (2007).

43. Degnan, J.H. & Rosenberg, N.A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol 24, 332-40 (2009).

44. Liu, L. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542-3 (2008).

45. Larget, B.R., Kotha, S.K., Dewey, C.N. & Ane, C. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26, 2910-1 (2010).

46. Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Molecular Biology and Evolution (2012).

47. Kubatko, L.S., Carstens, B.C. & Knowles, L.L. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971-3 (2009).

48. Liu, L., Yu, L. & Edwards, S.V. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol Biol 10, 302 (2010).

49. Kumar, S., Filipski, A.J., Battistuzzi, F.U., Kosakovsky Pond, S.L. & Tamura, K. Statistics and truth in phylogenomics. Mol Biol Evol 29, 457-72 (2012).

50. Weisrock, D.W. et al. Concatenation and Concordance in the Reconstruction of Mouse Lemur Phylogeny: An Empirical Demonstration of the Effect of Allele Sampling in Phylogenetics. Molecular Biology and Evolution 29, 1615-1630 (2012).

51. Kubatko, L.S. & Degnan, J.H. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol 56, 17-24 (2007).

52. Maddison, W.P. & Knowles, L.L. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology 55, 21 (2006).

53. Philippe, H. et al. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. PLoS Biol (2011).

54. Rodriguez-Ezpeleta, N. et al. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol 56, 389-99 (2007).

55. Degnan, J.H. & Rosenberg, N.A. Discordance of Species Trees with Their Most Likely Gene Trees. PLoS Genet 2, e68 (2006).

56. Glor, R.E. Phylogenetic Insights on Adaptive Radiation Annual Review of Ecology, Evolution, and Systematics 41, 251-270 (2010).

57. Seehausen, O. African cichlid fish: a model system in adaptive radiation research. Proceedings of the Royal Society of London - Series B: Biological Sciences 273, 1987-98 (2006).

58. Koblmüller, S. & Sefc..., K. The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. Hydrobiologia (2008).

59. Kocher, T.D. Adaptive evolution and explosive speciation: the cichlid fish model. Nature Reviews Genetics (2004).

60. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Molecular Ecology 18, 169-85 (2009).

61. Santos, M.E. & Salzburger, W. Evolution. How cichlids diversify. Science 338, 619-21 (2012).

62. Danley, P.D. et al. The impact of the geologic history and paleoclimate on the diversification of East african cichlids. Int J Evol Biol 2012, 574851 (2012).

63. Takahashi, K., Terai, Y., Nishida, M. & Okada, N. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Molecular Biology and Evolution 18, 2057-2066 (2001).

64. Moran, P. & Kornfield, I. Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. Mol Biol Evol 10, 1015 (1993).

65. Mims, M.C., Darrin Hulsey, C., Fitzpatrick, B.M. & Todd Streelman, J. Geography disentangles

introgression from ancestral polymorphism in Lake Malawi cichlids. Mol Ecol (2010).

66. Nagl, S. et al. The origin and age of haplochromine fishes in Lake Victoria, East Africa. Proceedings of the Royal Society B: Biological Sciences 267, 1049-1061 (2000).

67. Koblmüller, S., Egger, B., Sturmbauer, C. & Sefc, K.M. Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. Molecular Phylogenetics and Evolution 55, 318-34 (2010).

68. Seehausen, O. et al. Speciation through sensory drive in cichlid fish. Nature 455, 620-6 (2008).

69. Verheyen, E., Salzburger..., W. & Snoeks, J. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science (2003).

70. Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol Biol 5, 17 (2005).

71. Salzburger, W., Meyer, A., Baric, S., Verheyen, E. & Sturmbauer, C. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. Systematic Biology 51, 113-35 (2002).

72. Day, J.J., Cotton, J.A. & Barraclough, T.G. Tempo and mode of diversification of lake Tanganyika cichlid fishes. PLoS ONE 3, e1730 (2008).

73. Day, J.J., Santini, S. & Garcia-Moreno, J. Phylogenetic relationships of the Lake Tanganyika cichlid tribe Lamprologini: the story from mitochondrial DNA. Mol Phylogenet Evol 45, 629-42 (2007).

74. Wagner, C.E. et al. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol Ecol 22, 787-98 (2013).

75. Kocher, T.D., Conroy, J.A., McKaye, K.R., Stauffer, J.R. & Lockwood, S.F. Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. Molecular Phylogenetics and Evolution 4, 420-32 (1995).

76. Joyce, D.A. et al. Repeated colonization and hybridization in Lake Malawi cichlids. Current biology : CB 21, R108-R109 (2011).

77. Loh, Y.-H.E. et al. Origins of Shared Genetic Variation in African Cichlids. Molecular Biology and Evolution (2012).

78. Meyer, B.S. & Salzburger, W. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. Mol Ecol Resour 12, 1097-104 (2012).

79. Binladen, J. et al. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. PLoS ONE 2, e197 (2007).

Chapter 1

A novel primer set for multilocus phylogenetic inference in East African cichlid fishes

Britta S. Meyer & Walter Salzburger

Molecular Ecology Resources (2012) 12, 1097–11,
doi: 10.1111/j.1755-0998.2012.03169.x

BSM developed the primer pairs, conducted the laboratory work, analyzed the resulting sequence data and drafted the manuscript. WS designed the study and helped with the data analyses and manuscript drafting.

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) **12**, 1097–1104 doi: 10.1111/j.1755-0998.2012.03169.x

A novel primer set for multilocus phylogenetic inference in East African cichlid fishes

BRITTA S. MEYER and WALTER SALZBURGER
Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

Abstract

The cichlid fishes in the East African Great Lakes are a prime model system for the study of adaptive radiation. Therefore, the availability of an elaborate phylogenetic framework is an important prerequisite. Previous phylogenetic hypotheses on East African cichlids are mainly based on mitochondrial and/or fragment-based markers, and, to date, no taxon-rich phylogeny exists that is based on multilocus DNA sequence data. Here, we present the design of an extensive new primer set (24 nuclear makers) for East African cichlids that will be used for multilocus phylogenetic analyses in the future. The primers are designed to work for both Sanger sequencing and next-generation sequencing with the 454 technology. As a proof of principle, we validate these primers in a phylogenetically representative set of 16 cichlid species from Lake Tanganyika and main river systems in the area and provide a basic evaluation of the markers with respect to marker length and diversity indices.

Keywords: adaptive radiation, cichlid species flocks, nuclear markers, organismal diversification

Received 20 December 2011; revision received 8 June 2012; accepted 12 June 2012

Introduction

Cichlid fishes (Cichlidae) belong to one of the most species-rich families of vertebrates, with a distribution range from Africa including Madagascar, to Central and South America and South India. The Great Lakes in the East African Rift Valley harbour the largest and most diverse species flocks of cichlid fishes (Snoeks 2000; Turner *et al.* 2001) and are regarded as prime model systems to study evolutionary processes (reviewed in Kocher 2004; Salzburger 2009; Seehausen 2006). Among the three main species flocks, that of Lake Victoria, Lake Malawi and Lake Tanganyika, the latter is the morphologically, behaviourally, ecologically and genetically most diverse (Sturmbauer & Meyer 1992; Salzburger *et al.* 2002b; Young *et al.* 2009). This is due to the greater age of the lake, estimated between nine and 12 Ma (Cohen *et al.* 1993), and, consequently, the greater age of the radiation itself (Genner *et al.* 2007; Koblmüller *et al.* 2008; Schwarzer *et al.* 2009). Moreover, and unlike the species flocks of Lake Victoria and Lake Malawi, which exclusively consist of species of the haplochromine sublineage ('tribe'), the cichlid assemblage in Lake Tanganyika consists of 12–16 tribes, of which the haplochromines are but one (Poll 1986; Salzburger *et al.* 2002b; Takahashi 2003; Clabaut *et al.* 2005).

Correspondence: Walter Salzburger, Fax: +41-61-267-0301;
E-mail: walter.salzburger@unibas.ch

Since their discovery at the turn of the nineteenth century, the species flocks of cichlids in East Africa have been in the centre of empirical and theoretical research. The main focus has always been on speciation in general, and in particular, on the tempo and mode of diversification, the possible triggers and the progress of adaptive radiations, the respective role of sexual and natural selection and the role of evolutionary key innovations (Verheyen *et al.* 2003; Salzburger *et al.* 2005, 2007; Day *et al.* 2008; Seehausen *et al.* 2008; Salzburger 2009). Importantly, most of this research depends on phylogenetic hypotheses, which appear difficult to obtain in the rapidly evolving assemblages of cichlids in East Africa (Kocher 2003). Especially in the comparably young cichlid radiations of lakes Malawi and Victoria, there is only limited genetic variation in mitochondrial markers between both species and genera, and haplotype sharing is a common phenomenon (Meyer *et al.* 1990; Parker & Kornfield 1997; Shaw *et al.* 2000; Verheyen *et al.* 2003). The fragment-based amplified fragment length polymorphism (AFLP) method provided better resolution here (Albertson *et al.* 1999; Allender *et al.* 2003; Joyce *et al.* 2011), although comprehensive phylogenies are still lacking for cichlids from lakes Malawi and Victoria.

A more extensive phylogenetic framework is available for the cichlid species flock of Lake Tanganyika, which also includes analyses of its sublineages ('tribes'). Most of the available phylogenetic hypotheses are based on

mitochondrial markers (e.g. Cyprichromini: Brandstätter *et al.* 2005; Lamprologini: Day *et al.* 2007; Limnochromini: Duftner *et al.* 2005; Bathybatini: Koblmüller *et al.* 2005; Ectodini: Koblmüller *et al.* 2004; Haplochromini: Salzburger *et al.* 2005; Koblmüller *et al.* 2008). Fewer studies used a combination of sequence-based nuclear and mitochondrial markers (Salzburger *et al.* 2002a; Clabaut *et al.* 2005; Schelly *et al.* 2006; Nevado *et al.* 2009) or AFLPs and mitochondrial markers (Egger *et al.* 2007; Koblmüller *et al.* 2007a,b, 2010; Takahashi *et al.* 2007; Sturmbauer *et al.* 2010). These studies often led to new insights regarding hybridization, introgression or incomplete lineage sorting events (Nevado *et al.* 2009, 2011; Koblmüller *et al.* 2010). However, no taxon-rich phylogenetic study exists that is based on sequence data from various nuclear markers. This is in contrast to the many advantages that a (nuclear) multilocus phylogeny would provide. Most importantly, a species tree inferred from the gene trees of many independent loci should be more accurate than a species tree obtained from a few loci or a single locus only (Pamilo & Nei 1988).

Here, we present the design and general validation of primer pairs for 24 nuclear loci in East African cichlids. Our main goal was to obtain a set of nuclear markers for multilocus phylogenetic analyses. We focused on the development of markers with a length suitable for high-throughput sequencing. At the same time, we designed primers to amplify genes with known functions and from different functional categories. As a proof of principle, we tested our marker set in 16 East African cichlid species across a broad phylogenetic range and performed a phylogenetic analysis.

Materials and methods

Primer design

First, we defined the following general requirements for our primer sets:

1. the primers should work in a phylogenetically representative set of East African cichlid fishes
2. the primers should amplify between ca. 400–600 bp (based on the current read length of 454 sequencing/GS FLX Titanium)
3. they should have a maximal length of 24 bp (based on recommendations for fusion primer design)
4. all primers should have a similar melting temperature (Tm) at an optimum between 57–59 °C (according to the table of thermodynamic parameters from (SantaLucia (1998))
5. the genes to be amplified should be well characterized.

We first screened the literature for candidate primers, which were then, if necessary, modified to match the above requirements. Second, to generate new markers, we selected a set of candidate genes with known functions, for example, in coloration and pigmentation, growth factor activity, (craniofacial) bone development, protein processing, cell cycle, metabolism, or as transcription factors and ribosomal proteins. In the absence of a cichlid genome assembly (at the time the study was performed), the distribution of these candidate genes across fish genomes was determined using the available assemblies of Zebrafish (*Danio rerio*) and Medaka (*Oryzias latipes*) in Ensembl (Flicek *et al.* 2011). These two assemblies, in combination with available cichlid cDNA/EST sequences (Watanabe *et al.* 2004; Tsai *et al.* 2007; Salzburger *et al.* 2008; Tine *et al.* 2008; Kobayashi *et al.* 2009; Lee *et al.* 2010; Baldo *et al.* 2011), were also used to infer exon/intron boundaries for each locus, which was important to estimate intron lengths. The final primer design was based on additional cichlid sequences (from NCBI databases 'wgs' (whole genome shotgun) and 'nr' as well as unpublished sequences from our laboratory). To avoid the amplification of ancient paralogs, primers were designed in regions where paralogs differed. If possible, primers were designed for exon-primed intron-crossing (EPIC) markers, which anneal in conserved exons and amplify mainly the introns. Given a read length of ca. 400 bp (after trimming) by 454/GS FLX Titanium and ca. 600–800 bp by Sanger sequencing, only relatively short introns could be considered. All primers were designed with PRIMER-BLAST (Sayers *et al.* 2011), which includes the software PRIMER3 (Rozen & Skaletsky 2000) and a BLAST search (Altschul *et al.* 1990, 1997; Sayers *et al.* 2011), using the nr nucleotide database with the 'taxid' (NCBI taxonomy id) for cichlids (8113).

Taxon sampling

To assess the applicability of the newly designed primers in a broad spectrum of cichlid species, we tested them in a phylogenetically representative set of 16 cichlid species representing 12 tribes (Tylochromini, Tilapiini, Bathybatini, Eretmodini, Lamprologini, Ectodini, Cyprichromini, Perissodini, Limnochromini, Haplochromini/Tropheini, Cyphotilapiini). *Tylochromis polylepis*, a relatively recent colonizer of Lake Tanganyika and a representative of an ancestral lineage, was included as outgroup (Salzburger *et al.* 2002b; Clabaut *et al.* 2005; Koch *et al.* 2007). Note that most species are from Lake Tanganyika to account for its greater diversity in cichlid lineages; however, as we also included several haplochromines, our taxon sampling represents the entire phylogenetic spectrum of East African cichlids. Samples were collected in the years 2007 and 2008.

Molecular data

Genomic DNA was extracted from fin clips preserved in 95% ethanol, using the robotic workstation BioSprint 96 following the manufacture's protocol (Qiagen, Hombrechtikon, Switzerland). PCRs were performed in a final volume of 12.5 μL containing REDTaq® DNA Polymerase (0.04 units/μL), its PCR Buffer (1×) (Sigma-Aldrich, Buchs, Switzerland), 200 μM of each dNTP (Promega, Dübendorf, Switzerland), 0.2 μM of each sense and antisense primer (Microsynth, Balgach, Switzerland), 5–10 ng of DNA and water. The PCR conditions of all target fragments consisted of an initial denaturation for 2 min at 94 °C, followed by 32 cycles with a denaturation step at 94 °C for 30 s, an annealing step at 52–54 °C for 30 s and finalized by an extension step at 72 °C for 1 min. PCR success was evaluated using gel electrophoresis (1.5% agarose; buffered in 1× TAE). To assess the length of the PCR product, a size standard (BenchTop 100bp DNA Ladder; Promega) was added as reference to the gel. For visualization under UV-light, the gel was stained with GelRed™ (Biotium; VWR International, Dietikon, Switzerland).

Prior to DNA sequencing, the PCR products were purified from excess primers and dNTPs using ExoSAP-IT (GE Healthcare, Glattbrugg, Switzerland) following the manufacturer's protocol. Sequencing reactions were performed using 1.5 μL purified PCR products, the primers specified in Table 1 (0.5 μL/10 μM), and 1 μL BigDye® Terminator v3.1 Ready Reaction Mix (Applied Biosystems, Rotkreuz, Switzerland) in a total volume of 8 μL under standard conditions (1 min 94 °C, followed by 25 cycles with 10 s at 94 °C, 20 s at 52 °C, 4 min at 60 °C). To scavenge all unincorporated BigDye® terminators, the BigDye XTerminator® Purification Kit with its standard protocol (Applied Biosystems) was used. After this purification step, sequences were obtained with the 3130*xl* Genetic Analyzer (Applied Biosystems). Chromatograms were edited in CODONCODE ALIGNER (CodonCode, Dedham, MA, USA). Double peaks with equally high intensities in the chromatograms were assigned as heterozygous sites (SNPs). These polymorphic sites were coded as ambiguous nucleotides following the IUPAC-IUB code.

Sequence analysis

Initial alignments were performed with MAFFT (–auto) (Katoh & Toh 2008). Thereafter, a 'supermatrix' was generated by concatenating the single genes of the 16 species using MESQUITE 2.73 (Maddison & Maddison 2010). We then used MEGA5 (Tamura *et al.* 2011) to calculate genetic p-distance between the ingroup species (excluding *T. polylepis)* with complete deletion, for each single gene (and, in a second step, within exons and introns separately). The percentage of missing data, the gaps and polymorphic sites were accessed using MACCLADE 4.08 (Maddison & Maddison 2005).

Phylogenetic analysis

Prior to phylogenetic analysis, we determined the best fitting substitution model for each gene with JMODELTEST 0.1.1 (Guindon & Gascuel 2003; Posada 2008) on the basis of the Bayesian information criterion (BIC) (Schwarz 1978). We first performed a maximum likelihood analysis with GARLI 2.0 (Zwickl 2006) and our partitioned supermatrix. We run ten independent replicates, which were terminated automatically after 5000 generations with no significant ($P < 0.01$) improvements in topology scoring. To access confidence in the tree topology, 1000 bootstrap replicates were executed and a majority-rule consensus tree was constructed with PAUP* 4.0a114 (Swofford 2002). Bayesian phylogenetic inference for the partitioned data set was conducted with MRBAYES v3.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). The starting trees were set to be random. Prior probability distributions for all parameters were assumed to be flat. Two simultaneous MCMC were conducted for 21 000 000 generations, each of which had three heated and one cold chain, and the trees were sampled every 1000 generations. The first 25% of the sampled trees were discarded as burnin. To diagnose convergence in the two runs, we used AWTY (Nylander *et al.* 2008) and TRACER v1.5 (Rambaut & Drummond 2007). The majority-rule consensus tree derived from GARLI and PAUP* as well as the tree from MRBAYES were finally processed in FIGTREE v1.3.1 (Rambaut 2009).

Results

We designed 24 new primer pairs that amplify nuclear markers in East African cichlid fishes. Two of these primers are variations of already existing primers. The S7 reverse primer is adopted from Chow & Hazama (1998), but with an extra degenerated nucleotide. The other primer, the *bmp4* reverse primer, is a variation of an existing primer from Albertson *et al.* (2003), slightly elongated and with more specific nucleotides. The length of the resulting PCR products ranges between 357–707 base pairs, with an average length of 497 bp and a median of 483.5 bp. Table 1 lists all loci with their specific forward and reverse primer sequences, their location in Medaka chromosomes and the number of base pairs belonging to intron or exon.

The amplification of these loci was successful in most of the 16 tested cichlid species; on average, 15 species showed a band on the agarose gels. Sequencing success with the Sanger method was less successful, which we

Table 1 List of the 24 primer pairs with their forward and reverse sequences using IUPAC code, the length of the respective PCR product with primers, the ENSEMBL-ID (or NCBI Accession no) of the Medaka ortholog, known chromosome position (Chr. pos.) in Medaka (M), and the amount of base pairs in exon and intron

Locus	Primer forward/reverse (5′–3′)	PCR product (bp)	ID of Medaka ortholog	Chr pos M	Exon (bp)	Intron (bp)
bmp4	GAGGACCCATGCCCATTCGTTT/GCCACTATCCAGTCATTCCAGCC	577	ENSORLG00000013304	22	482	0
bmp2	AGGCCCTGGCCAGCCTAAAA/TCCTGCGTCTGTGGGCATCCTT	414	ENSORLG00000009772	24	315	0
fgf6	CGCAAAGGTGCCACTACAG/TCGCACTGCACGGATGCAAA	512	ENSORLG00000015820	23	286	158
furina	GCTGCATGGGGACAGACAGTCA/ATAGTCACTGGCACCCGCCACA	357	ENSORLG00000009133	3	154	94
runx2	CGGGGTTGGTGTTTGAGGGCAA/GCTGACATGGTGTCACTGTGCTGA	411	ENSORLG00000010169	24	95	218
shh	TGGCACCAAGGAAGCCGTCA/CACTGCTTGGAGGCTGGGA	512	ENSORLG00000010463	20	421	0
pax9	TCCCACGGCTGTGTCAGYAA/ACAGAGTGCGAGGAAGGCCA	434	AB187122.1	?	338	0
sox10b	TSCRGGGTCTGGGAAACCTCAT/TGGTGGTCGGCGTATTCTGCAA	486	ENSORLG00000014587	8	310	0
ednrb1	CGTTGGCCTGCACTGCCATT/AGGCAGCCAGCACAGAGCAAA	479	ENSORLG00000011054	17	54	320
mc1r	GACCACGGCCTCCTGGATGT/GTTGCAGAAGGGGCTGGTGG	510	ENSORLG00000009400	3	401	0
c-ski	CGACCAGCTGGAGATCCT/TCCTCTTGTACTTGTTGGCG	491	ENSORLG00000016855	7	408	0
kita	CAGAGTACTGCTGTTTCGGMGAT/GGCTAAGAACTCCATGCCT TTGGC	611	ENSORLG00000000569	4	237	270
mitfa	CCTGGCATGAAGCARGTACTGGAC/TTGCYAGAGCACGAA CTTCRGC	456	ENSORLG00000003123	5	25	373
tyr	TGGGTGGACGCAACTCCCTT/TGGCAAATCGGTCCATGGGT	659	ENSORLG00000010905	13	155	413
csfr1	AAGCACAGATGGGACACGCC/TGTACTGGCCCTGCTCCTGT	459	ENSORLG00000004849	10	25	324
pax3	AAGAGCCCGGTGGAGGAAGCAA/TGACGGCGTTGGTGTGTCCT	471	ENSORLG00000015932	17	254	130
hag	AAACTGGTACARYGGGVTCTGC/AGCGRCAGACGTCACCCTTGT	470	ENSORLG00000000906	15	115	309
rag	TCGGCGCTTTCGGTACGATGTG/TGCCCCTGAAGTGGAASSGA	461	ENSORLG00000011969	6	373	0
b2m	GCCACGTGAGTRATTTCCACCCC/ACGCTAYACRGYGGACYCTGA	508	ENSORLG00000012506	23	235	183
gapdhs	CCCTGGCCAAAGTCATCCACGATA/CACCACTGACACATCGGCCACT	499	ENSORLG00000006033	16	171	258
ccng1	CTGCTTGCCCTGGCTCTCCT/AGCTGACTCAGGTATGGTCGGA	707	ENSORLG00000005817	10	210	444
ptr-like	GCGGGTAGTGAATGTGAGTGCG/ACCCAAGACACCCAGCTCCA	436	ENSORLG00000015652	24	368	0
enc1	CRGTTCGCCTTGCGCTRTTGC/TGGGTGCCGCCTTTGACCAT	417	ENSORLG00000003288	12	329	0
s7	CGTGCCATTTTACTCTGGACTKGC/AACTCGTCYGGCTTCTCGCC	569	ENSORLG00000018123	24	0	414

attribute in part to the existence of alleles with different lengths resulting in double peaks (note: this is not an issue when using next-generation sequencing techniques). The percentage of missing data and gaps per species are listed in Table 2.

In total, we obtained a concatenated data set of 24 partial gene sequences containing 9669 bp. A total of 583 sites were variable (6.03%), of which 130 are parsimony–informative sites (1.3%) (calculated without the outgroup taxon *Tylochromis polylepis* and without indels or polymorphic sites coded with ambiguous IUPAC code). The combined sequence matrix consists of 5761 bp (59.58%) from exons and 3908 bp (40.42%) from introns (Table 1). In 18 of the 24 loci, we detected heterozygous SNPs (46 SNPs in total; referred to as polymorphic sites in the tables, Table 2 and Table S1, Supporting information). Detailed information about the number of variable sites, the number and location of polymorphisms, the calculated BIC value and the gene ontology (GO) terms for every gene are listed in Table S1 (Supporting information). The PCR conditions for each primer pair (including the enzymes used and the annealing temperatures), and PCR and sequencing success (and possible reason for its failure) are shown in Table S2 (Supporting information).

Maximum likelihood (not shown) and Bayesian inference (Fig. 1) of the concatenated data yielded congruent trees. The only differences between the two trees concern weakly supported nodes (the relative positions of the Cyprichromini, Cyphotilapiini, Limnochromini, Eretmodini and Perissodini to the Lamprologini/Ectodini complex) and the placement of *Sargochromis* within the Haplochromini.

Discussion

In this study, we present a new primer set for phylogenetic inferences in East African cichlid fishes. We further show that our primers amplify successfully in most of the tested representatives from Lake Tanganyika, making our primer set applicable for a great portion of the ~250 cichlid species in this lake. As the primers amplify very well in our riverine and Tanganyikan representatives of haplochromines *sensu lato* (Salzburger *et al.* 2005), it is likely that they also work for the members of

Table 2 DNA sequencing success in our test taxon set. The total sequence length (in bp) for each taxon, the percentage of missing data relative to the supermatrix, the percentage of gaps and the number of polymorphic sites (SNPs) within each taxon are given

Taxon	Sequence length (bp)	Missing (%)	Gaps (%)	SNPs
Bathybates graueri	9528	37.2	1.5	0
Cyprichromis leptosoma	9616	54.7	0.5	1
Ophthalmotilapia ventralis	9517	56.9	1.6	7
Oreochromis tanganicae	9552	31.9	1.2	1
Cyphotilapia frontosa	9532	52.2	1.4	1
Eretmodus cyanostictus	9437	14.4	2.4	5
Tropheus moori	9482	7.3	1.9	4
Ctenochromis horei	9480	1.9	1.9	2
Astatotilapia burtoni	9504	13.1	1.7	2
Lamprologus callipterus	9593	18.6	0.8	5
Perissodus microlepis	9489	14.7	1.9	0
Neolamprologus pulcher	9530	31.8	1.4	1
Tylochromis polylepis	9633	37.9	0.4	2
Limnochromis abeelei	9494	13.1	1.8	5
Sargochromis spec.	9523	11.9	1.5	7
Pseudocrenilabrus philander	9531	23.7	1.4	3

the species flocks of Lake Malawi and the Lake Victoria region, which exclusively consist of haplochromines (Meyer *et al.* 1990; Verheyen *et al.* 2003; Salzburger *et al.* 2005).

We designed the primer pairs specifically for the use in both single-read Sanger sequencing and next-generation sequencing with the 454 technology, by restricting the amplicon product length to the read length of these methods (see Table 1). Another important attribute of our primers constitutes the similar melting temperature and consequently a comparable annealing temperature, which enables multiplexed reactions and cost-effective parallel high-throughput sequencing. Furthermore, the markers are placed into annotated genes with known functions, leading to a good coverage of the genome (see Table 1) and the avoidance (or at least relatively easy identification) of paralogs or pseudogenes. Finally, the primers were designed to cover coding (exons) and non-coding (introns) regions of these genes. Intron sequences have been used successfully for both phylogenetic inference (Hedin & Maddison 2001; Fujita *et al.* 2004; Dalebout *et al.* 2008; Jacobsen & Omland 2011; Yu *et al.* 2011) and population genetics (Palumbi & Baker 1994; Tay *et al.* 2008; Carvajal-Vallejos *et al.* 2010), particularly because

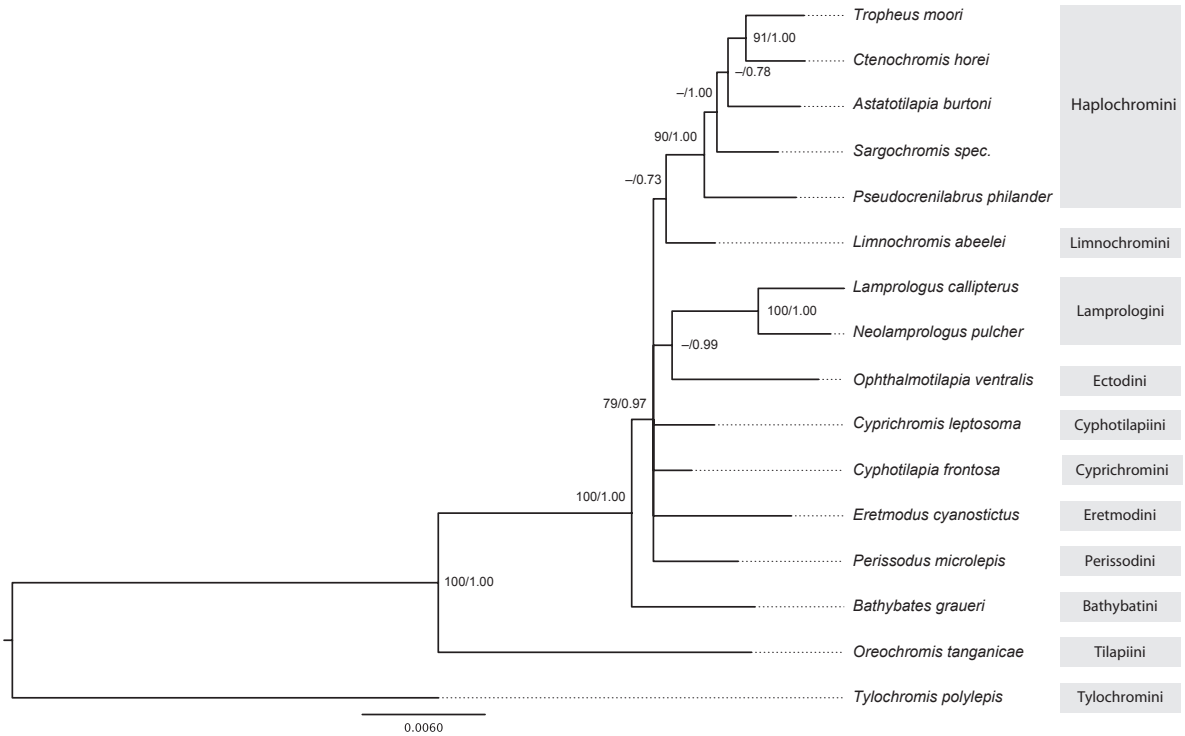


Fig. 1 Bayesian inference topology inferred with MrBayes for the 16 species in our test data set and based on 24 genes (concatenated, 9669 bp). Bootstrap support for ML > 50 and posterior probability of the MrBayes analysis >0.50 are shown. Branch lengths are proportional to the number of mutations per site.

introns typically contain a higher percentage of polymorphic sites than exons (reviewed in Zhang & Hewitt 2003). About 40% of the nuclear DNA sequences, obtained by using our newly developed primer set for cichlids, belong to introns, which indeed show greater diversity than the exons (see Table S1, Supporting information).

The primary goal of this novel primer set is to use it for phylogenetic purposes in order to refine and extend existing phylogenetic hypotheses (Salzburger *et al*. 2002b, 2005; Clabaut *et al*. 2005; Day *et al*. 2008; Koblmüller *et al*. 2008) and to address the gene tree/species tree issue in Tanganyikan cichlids (see Brito & Edwards 2009; Heled & Drummond 2010; Liu 2008 for methods and discussions).

The various drawbacks of the sole use of mitochondrial markers for phylogenetic, phylogeographic and population genetic inference have been frequently discussed (Ballard & Whitlock 2004; Ballard & Rand 2005; Rubinoff & Holland 2005; Brito & Edwards 2009; Galtier *et al*. 2009). However, also nuclear markers have some drawbacks, such as a relatively low mutation rate (Moritz *et al*. 1987) and a four times larger effective population size compared with the haploid and uniparentally inherited mitochondrial markers leading to longer coalescence times and slower fixation rates (Moore 1995). In the case of the new marker set provided here, this is counterbalanced by the relatively large amount of sequence data that can be obtained.

Taken together, we present the development and the proof of functionality of the so far largest set of independent sequence-based nuclear markers for phylogenetic purposes for East African cichlid fishes. The markers can be used in both Sanger sequencing and next-generation sequencing using the 454 approach. We thus provide an important tool that will be used for multimarker phylogenetic analyses of East African cichlids in the future.

Acknowledgements

We are grateful to M. Barluenga, F. Muenzel, A. Indermaur and M. Muschick for help during fieldwork, B. Aeschbach and N. Boileau for technical assistance and A. Indermaur and M. Roesti for valuable comments on the manuscript. We thank three anonymous reviewers for helpful feedback on the manuscript. Sampling was performed under a memorandum of understanding (MOU) with the University of Zambia in Lusaka and the Department of Fisheries, Lake Tanganyika branch, Mpulungu, Zambia. This work was supported by the European Research Council (ERC; Starting Grant 'INTERGENADAPT'), the Swiss National Science Foundation (grant 122458) and the University of Basel.

References

Albertson RC, Markert JA, Danley PD, Kocher TD (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 5107–5110.

Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5252–5257.

Allender CJ, Seehausen O, Knight ME, Turner GF, Maclean N (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 14074–14079.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Altschul SF, Madden TL, Schäffer AA *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Baldo L, Santos ME, Salzburger W (2011) Comparative transcriptomics of eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome Biology and Evolution*, **3**, 443–455.

Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 621–642.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Brandstätter A, Salzburger W, Sturmbauer C (2005) Mitochondrial phylogeny of the Cyprichromini, a lineage of open-water cichlid fishes endemic to Lake Tanganyika, East Africa. *Molecular Phylogenetics and Evolution*, **34**, 382–391.

Brito P, Edwards S (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439–455.

Carvajal-Vallejos FM, Duponchelle F, Ballivian JPT *et al*. (2010) Population genetic structure of *Cichla pleiozona* (Perciformes: Cichlidae) in the Upper Madera basin (Bolivian Amazon): sex-biased dispersal? *Molecular Phylogenetics and Evolution*, **57**, 1334–1340.

Chow S, Hazama K (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, **7**, 1255–1256.

Clabaut C, Salzburger W, Meyer A (2005) Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. *Journal of Molecular Evolution*, **61**, 666–681.

Cohen AS, Soreghan MJ, Scholz CA (1993) Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. *Geology*, **21**, 511–514.

Dalebout ML, Steel D, Baker CS (2008) Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: implications for the evolution of male tusks. *Systematic Biology*, **57**, 857–875.

Day JJ, Santini S, Garcia-Moreno J (2007) Phylogenetic relationships of the Lake Tanganyika cichlid tribe Lamprologini: the story from mitochondrial DNA. *Molecular Phylogenetics and Evolution*, **45**, 629–642.

Day JJ, Cotton JA, Barraclough TG (2008) Tempo and mode of diversification of lake Tanganyika cichlid fishes. *PLoS ONE*, **3**, e1730.

Duftner N, Koblmüller S, Sturmbauer C (2005) Evolutionary relationships of the Limnochromini, a tribe of benthic deepwater cichlid fish endemic to Lake Tanganyika, East Africa. *Journal of Molecular Evolution*, **60**, 277–289.

Egger B, Koblmüller S, Sturmbauer C, Sefc K (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. *BMC Evolutionary Biology*, **7**, 137.

Flicek P, Amode MR, Barrell D *et al*. (2011) Ensembl 2011. *Nucleic Acids Research*, **39**, D800–D806.

Fujita MK, Engstrom TN, Starkey DE, Shaffer HB (2004) Turtle phylogeny: insights from a novel nuclear intron. *Molecular Phylogenetics and Evolution*, **31**, 1031–1040.

Galtier N, Nabholz B, GlÉMin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Genner MJ, Seehausen O, Lunt DH *et al*. (2007) Age of Cichlids: new dates for ancient lake fish radiations. *Molecular Biology and Evolution*, **24**, 1269–1282.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hedin MC, Maddison WP (2001) Phylogenetic utility and evidence for multiple copies of elongation factor-1α in the spider genus *Habronattus* (Araneae: Salticidae). *Molecular Biology and Evolution*, **18**, 1512–1521.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Jacobsen F, Omland KE (2011) Species tree inference in a recent radiation of orioles (genus *Icterus*): multiple markers and methods reveal cytonuclear discordance in the northern oriole group. *Molecular Phylogenetics and Evolution*, **61**, 460–469.

Joyce DA, Lunt DH, Genner MJ *et al*. (2011) Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology*, **21**, R108–R109.

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.

Kobayashi N, Watanabe M, Horiike T, Kohara Y, Okada N (2009) Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. *Gene*, **441**, 187–191.

Koblmüller S, Salzburger W, Sturmbauer C (2004) Evolutionary relationships in the sand-dwelling cichlid lineage of Lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. *Journal of Molecular Evolution*, **58**, 79–96.

Koblmüller S, Duftner N, Katongo C, Phiri H, Sturmbauer C (2005) Ancient divergence in bathypelagic Lake Tanganyika deepwater cichlids: mitochondrial phylogeny of the tribe Bathybatini. *Journal of Molecular Evolution*, **60**, 297–314.

Koblmüller S, Duftner N, Sefc K *et al*. (2007a) Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika – the result of repeated introgressive hybridization. *BMC Evolutionary Biology*, **7**, 7.

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (2007b) Evolutionary history of Lake Tanganyika's scale-eating cichlid fishes. *Molecular Phylogenetics and Evolution*, **44**, 1295–1305.

Koblmüller S, Schliewen UK, Duftner N *et al*. (2008) Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics and Evolution*, **49**, 153–169.

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (2010) Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Molecular Phylogenetics and Evolution*, **55**, 318–334.

Koch M, Koblmüller S, Sefc K (2007) Evolutionary history of the endemic Lake Tanganyika cichlid fish *Tylochromis polylepis*: a recent intruder to a mature adaptive radiation. *Journal of Zoological Systematics and Evolutionary Research*, **45**, 64–71.

Kocher TD (2003) Evolutionary biology: fractious phylogenies. *Nature*, **423**, 489–491.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **5**, 288–298.

Lee B, Howe A, Conte M *et al*. (2010) An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. *BMC Genomics*, **11**, 278.

Liu L (2008) BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*, **24**, 2542–2543.

Maddison DR, Maddison WP (2005) *MacClade*. Sinauer Associates, Inc., Sunderland, MA.

Maddison WP, Maddison DR (2010) Mesquite: a modular system for evolutionary analysis. Available at: http://www.mesquite.org.

Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, **347**, 550–553.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, **18**, 269–292.

Nevado B, Koblmüller S, Sturmbauer C *et al*. (2009) Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. *Molecular Ecology*, **18**, 4240–4255.

Nevado B, Fazalova V, Backeljau T, Hanssens M, Verheyen E (2011) Repeated unidirectional introgression of nuclear and mitochondrial DNA between four congeneric Tanganyikan cichlids. *Molecular Biology and Evolution*, **28**, 2253–2267.

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, **24**, 581–583.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.

Parker A, Kornfield I (1997) Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of lake Malawi, East Africa. *Journal of Molecular Evolution*, **45**, 70–83.

Poll M (1986) *Classification des Cichlidae du lac Tanganika: tribus, genres et espèces*, Mémoires de la Classe des Sciences edn. Académie Royale de Belgique, Brussels.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Rambaut A (2009) FigTree v1.3.1. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummond A (2007) Tracer v1.5. Available at: http://beast.bio.ed.ac.uk/Tracer.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, NJ)*, **132**, 365–386.

Rubinoff D, Holland BS (2005) Between two extremes: mitochondrial DNA is neither the Panacea nor the Nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, **54**, 952–961.

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**, 169–185.

Salzburger W, Baric S, Sturmbauer C (2002a) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**, 619–625.

Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C (2002b) Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Systematic Biology*, **51**, 113–135.

Salzburger W, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology*, **5**, 17.

Salzburger W, Braasch I, Meyer A (2007) Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. *BMC Biology*, **5**, 51.

Salzburger W, Renn S, Steinke D *et al*. (2008) Annotation of expressed sequence tags for the East African cichlid fish *Astatotilapia burtoni* and evolutionary analyses of cichlid ORFs. *BMC Genomics*, **9**, 96.

SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 1460–1465.

Sayers EW, Barrett T, Benson DA *et al*. (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **39**, D38–D51.

Schelly R, Salzburger W, Koblmüller S, Duftner N, Sturmbauer C (2006) Phylogenetic relationships of the lamprologine cichlid genus *Lepidiolamprologus* (Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. *Molecular Phylogenetics and Evolution*, **38**, 426–438.

Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Schwarzer J, Misof B, Tautz D, Schliewen U (2009) The root of the East African cichlid radiations. *BMC Evolutionary Biology*, **9**, 186.

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1987–1998.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Shaw PW, Turner GF, Rizman Idid M, Robinson RL, Carvalho GR (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **267**, 2273–2280.

Snoeks J (2000) How well known is the ichthyodiversity of the large East African lakes? In: *Advances in Ecological Research* (eds Rossiter A, Kawanabe H), pp. 17–38. Academic Press, Waltham, MA, USA.

Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. *Nature*, **358**, 578–581.

Sturmbauer C, Salzburger W, Duftner N, Schelly R, Koblmüller S (2010) Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. *Molecular Phylogenetics and Evolution*, **57**, 266–284.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates Sunderland, Massachusetts.

Takahashi T (2003) Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). *Ichthyological Research*, **50**, 367–382.

Takahashi R, Watanabe K, Nishida M, Hori M (2007) Evolution of feeding specialization in Tanganyikan scale-eating cichlids: a molecular phylogenetic approach. *BMC Evolutionary Biology*, **7**, 195.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tay WT, Behere GT, Heckel DG, Lee SF, Batterham P (2008) Exon-primed intron-crossing (EPIC) PCR markers of *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Bulletin of Entomological Research*, **98**, 509–518.

Tine M, de Lorgeril J, D'Cotta H *et al.* (2008) Transcriptional responses of the black-chinned tilapia *Sarotherodon melanotheron* to salinity extremes. *Marine Genomics*, **1**, 37–46.

Tsai CL, Wang LH, Shiue YL, Chao TY (2007) Influence of temperature on the ontogenetic expression of neural development-related genes from developing tilapia brain expressed sequence tags. *Marine Biotechnology (New York, NY)*, **9**, 243–261.

Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL (2001) How many species of cichlid fishes are there in African lakes? *Molecular Ecology*, **10**, 793–806.

Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, **300**, 325–329.

Watanabe M, Kobayashi N, Shin-i T *et al.* (2004) Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock: identity of EST sequences between *Haplochromis chilotes* and *Haplochromis* sp. ''Redtailsheller''. *Gene*, **343**, 263–269.

Young KA, Snoeks J, Seehausen O (2009) Morphological diversity and the roles of contingency, chance and determinism in African cichlid radiations. *PLoS ONE*, **4**, e4740.

Yu L, Luan P-T, Jin W *et al.* (2011) Phylogenetic utility of nuclear introns in interfamilial relationships of Caniformia (Order Carnivora). *Systematic Biology*, **60**, 175–187.

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Zwickl DJ (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD Dissertation, The University of Texas at Austin, Austin.

B.S.M., W.S. conceived and designed the project. B.S.M. performed the experiments. B.S.M., W.S. analyzed the data and wrote the paper.

Data Accessibility

All DNA sequences from this study are available under GenBank Accession: JX135129–JX135389 (for more details see: Table S3, Supporting information).

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 List of the 24 genes used for primer design. Information is provided with regard to the alignment length used for primer design; the percentage of variable and parsimony-informative sites without the outgroup; the number of polymorphic sites (SNPs), the location in the different species and its percentage including the outgroup; average p-distance calculated with complete deletion and without the outgroup; its range and SE calculated with 500 bootstrap replicates; and average p-distance including only base pairs from exon or intron without the outgroup; the recommended model from JMODELTEST 0.1.1 on the basis of the BIC with its computed likelihood scores; the GO terms from annotated fish sequences from UniProtKB.

Table S2 List of the 24 genes used for primer design. Information about the number of successful PCR and sequencing reactions; further used sequences and their Accession number of GenBank or other source; the percentage of missing data (due to sequencing errors); used annealing temperature in the PCR and used *Taq* polymerase; +/+ designates successful PCR and sequencing reaction, −/− both unsuccessful, +/− PCR successful and sequencing unsuccessful; indicating possible reason for failed sequencing reaction.

Table S3 List of used species, 24 markers and GenBank Accession numbers (asterisks indicate the usage of other unpublished primer pairs) http://www.ncbi.nlm.nih.gov/genbank/.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary Material

A novel primer set for multilocus phylogenetic inference in East African cichlid fishes

Britta S. Meyer & Walter Salzburger

Molecular Ecology Resources (2012) 12, 1097–11,
doi: 10.1111/j.1755-0998.2012.03169.x

Supplementary Table 1

List of the 24 genes used for primer design. Information is provided with regard to the alignment length used for primer design; the percentage of variable and parsimony-informative sites without the outgroup; the number of polymorphic sites (SNPs), the location in the different species and its percentage including the outgroup; average p-distance calculated with complete deletion and without the outgroup; its range and SE calculated with 500 bootstrap replicates; and average p-distance including only base pairs from exon or intron without the outgroup; the recommended model from jModelTest 0.1.1 on the basis of the Bayesian information criterion (BIC) with its computed likelihood scores; the GO terms from annotated fish sequences from UniProtKB (The UniProt Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research 40, D71-D75).

Supplementary Table 2

List of the 24 genes used for primer design. Information about the number of successful PCR and sequencing reactions; further used sequences and their Accession no of GenBank or other source; the percentage of missing data (due to sequencing errors); used annealing temperature in the PCR and used Taq polymerase; +/+ designates successful PCR and sequencing reaction, -/- both unsuccessful, +/- PCR successful and sequencing unsuccessful; indicating possible reason for failed sequencing reaction.

Supplementary Table 3

List of used species, 24 markers and GenBank Accession nos (asterisks indicate the usage of other unpublished primer pairs) http://www.ncbi.nlm.nih.gov/genbank/.

Locus	Aligned sites [bp]	Var sites without outgroup [%]	PI sites without outgroup [%]	No of poly-morphic sites with outgroup	Polymorphic sites: species, bases, position	Polymorphic sites with outgroup [%]	Gaps with outgroup [%]	Genetic distance without outgroup	SE	Range	Genetic distance (exon) without outgroup	Genetic distance (intron) without outgroup	BIC	-lnL	Molecular function	Process	Cellular Component
bmp4	482	6.22	1.24	4	Ophven, S, 215 Ophven, Y, 216 Ctehor, M, 221 Sarspc, R, 431	0.05	0	0.012	0.002	0-0.048	-	-	HKY+I	924.25	growth factor activity	BMP signaling pathway involved in heart jogging; BMP signaling pathway involved in mesodermal cell fate specification; determination of ventral identity; embryonic hemopoiesis; growth; heart looping; notochord cell differentiation; positive regulation of cell proliferation; pronephros development	extracellular region
bmp2	315	1.59	0	1	Ophven, Y, 232	0.02	0	0.003	0.001	0-0.006	-	-	K80	520.66	growth factor activity; cytokine activity; protein heterodimerization activity; protein homodimerization activity	cartilage development; growth; ossification; BMP signaling pathway; cell-cell signaling involved in cell fate commitment; determination of left/right symmetry; determination of ventral identity; dorsal convergence; liver development; protein complex assembly; semicircular canal formation; ventricular cardiac muscle cell	extracellular region; extracellular space
fgf6	444	3.38	0.68	0	-	0	0.01	0.007	0.002	0-0.020	0.003	0.013	K80	756.25	growth factor activity	-	-
furina	248	2.82	0	0	-	0	0.15	0.011	0.004	0.004-0.020	0.008	0.017	JC	454.82	serine-type endopeptidase activity	proteolysis; cartilage morphogenesis; fin morphogenesis	Golgi apparatus; extracellular space
runx2	313	0.64	0	2	Tylpol, Y, 92 Tylpol, Y, 122	0.04	0.11	0.002	0.001	0-0.006	-	-	JC	463.89	ATP binding; DNA binding; sequence-specific DNA binding transcription factor activity; vitamin D receptor binding	chondrocyte differentiation; positive regulation of transcription, DNA-dependent	nucleus
shh	421	2.85	1.19	0	-	0	0	0.008	0.004	0-0.018	-	-	K80	1678.28	peptidase activity	adenohypophysis development; axon guidance; blood circulation; cardioblast differentiation; caudal fin morphogenesis; cell-cell signaling; dorsal/ventral pattern formation; embryonic camera-type eye development; embryonic digestive tract morphogenesis; embryonic neurocranium morphogenesis; embryonic viscerocranium morphogenesis; endocrine pancreas development; eye photoreceptor cell development; floor plate formation; glomerulus development; glial cell differentiation; heart looping; hindbrain development; intein-mediated protein splicing; muscle cell fate specification; negative regulation of cell proliferation; neural plate morphogenesis; neuron fate commitment; pectoral fin development; photoreceptor cell maintenance; post-embryonic foregut morphogenesis; proteolysis; regulation of BMP signaling pathway involved in heart jogging; regulation of gene expression; regulation of vascular endothelial growth factor receptor signaling pathway; skeletal muscle fiber development; somitogenesis; swim bladder morphogenesis; multicellular organismal development	extracellular space; plasma membrane
pax9	338	1.18	0.3	0	-	0	0	0.004	0.002	0-0.009	-	-	F81	488.29	DNA binding	multicellular organismal development; regulation of transcription, DNA-dependent; transcription, DNA-dependent	nucleus
sox10b	310	4.52	0.97	2	Cphfro, Y, 121 Lamcal, W, 281	0.04	1.81	0.01	0.003	0-0.017	-	-	HKY	571.82	DNA binding	Schwann cell development; enteric nervous system development; negative regulation of neurogenesis; neural crest cell migration; oligodendrocyte differentiation; otic vesicle morphogenesis; otolith development; positive regulation of gene expression; semicircular canal morphogenesis	-
ednrb1	374	5.35	1.34	3	Lamcal, M, 89 Lamcal, Y, 132 Lamcal, Y, 249	0.05	0.54	0.013	0.003	0.003-0.022	0.014	0.013	HKY	746.24	endothelin receptor activity	-	integral to membrane
mc1r	401	3.74	0.75	3	Lamcal, Y, 179 Lchabe, K, 242 Sarspc, Y, 394	0.05	0	0.009	0.002	0-0.015	-	-	K80	733.99	melanocyte-stimulating hormone receptor activity	-	integral to membrane; plasma membrane
c-ski	408	4.41	0.49	0	-	0	0	0.007	0.002	0-0.020	-	-	HKY	706.27	nucleotide binding; SMAD binding; transcription repressor activity	dorsal/ventral pattern formation; negative regulation of BMP signaling pathway; negative regulation of transcription from RNA polymerase II promoter; negative regulation of transforming growth factor beta receptor signaling pathway; regulation of apoptotic process	nucleus; cytoplasm
kita	507	4.93	0.99	2	Astbur, R, 285 Psephi, K, 416	0.02	0	0.009	0.002	0-0.020	0.008	0.01	K80	989.06	ATP binding; metal ion binding; vascular endothelial growth factor-activated receptor activity; transmembrane receptor protein tyrosine kinase activity;	melanocyte differentiation; transmembrane receptor protein tyrosine kinase signaling pathway	integral to plasma membrane; plasma membrane
mitfa	398	9.8	1.76	1	Lchabe, R, 48	0.02	5.94	0.022	0.004	0-0.058	0.011	0.023	HKY	807.9	nucleic acid binding transcription factor activity	regulation of transcription, DNA-dependent; melanocyte differentiation; transcription, DNA-dependent	nucleoplasm
tyr	568	9.68	2.11	5	Sarspc, Y, 65 Sarspc, Y, 95 Lchabe, Y, 142 Lchabe, R, 167 Lchabe, R, 518	0.06	1.53	0.027	0.004	0.006-0.066	0.013	0.032	HKY+I	1313.29	metal ion binding; monophenol monooxygenase activity; oxidoreductase activity	melanin biosynthetic process; dopamine biosynthetic process from tyrosine	integral to membrane; melanosome membrane
csfr1	349	4.58	0.29	1	Psephi, K, 200	0.02	0	0.013	0.004	0.003-0.026	0	0.014	HKY+I	594.86	ATP binding; vascular endothelial growth factor-activated receptor activity; transmembrane receptor protein tyrosine kinase	macrophage chemotaxis; xanthophore differentiation; transmembrane receptor protein tyrosine kinase signaling pathway	integral to plasma membrane; membrane
pax3	384	5.47	1.04	3	Tromoo, K, 181 Astbur, Y, 234 Tromoo, R, 373	0.05	0.24	0.01	0.003	0.003-0.026	0.006	0.02	JC	689.05	sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity	multicellular organismal development; transcription, DNA-dependent; enteric nervous system development; neural crest cell migration; xanthophore differentiation	nucleus
hag	424	16.04	4.95	2	Neopul, R, 309 Ophven, R, 418	0.03	14.18	0.042	0.006	0.004-0.177	0.012	0.059	TPM3+G	1260.68	-	-	-
rag	373	5.63	0.54	5	Cyplep, S, 96 Erecya, R, 234 Erecya, S, 289 Ophven, Y, 363 Ctehor, Y, 334	0.08	0.05	0.01	0.003	0-0.022	-	-	JC+I	743.72	endonuclease activity; histone binding; protein homodimerization activity; sequence-specific DNA binding; ubiquitin-protein ligase activity; zinc ion binding; sequence-specific DNA binding	B cell differentiation; T cell differentiation in thymus; V(D)J recombination; histone monoubiquitination	intracellular; nucleus
b2m	418	7.66	4.31	1	Erecya, R, 163	0.01	0.69	0.025	0.004	0.005-0.059	0.031	0.018	HKY+I	864.14	-	antigen processing and presentation of peptide antigen via MHC class I; immune response	MHC class I protein complex
gapdhs	429	6.53	0.23	6	Sarspc, R, 36 Tromoo, K, 100 Sarspc, Y, 126 Ophven, Y, 284 Sarspc, S, 331 Tromoo, K, 367	0.09	3.05	0.012	0.002	0-0.050	0.001	0.021	JC	775	NAD binding; NADP binding; glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity; oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	glycolysis; glucose metabolic process	cytoplasm
ccng1	654	11.31	2.29	2	Psephi, R, 116 Oretan, M, 231	0.02	2.09	0.021	0.003	0.002-0.049	0.013	0.026	K80	1417.81	-	-	nucleus
ptr-like	368	4.35	0.82	0	-	0	0	0.008	0.002	0-0.030	-	-	K80	614.24	hedgehog receptor activity	-	integral to membrane; plasma membrane
enc1	329	3.04	1.22	1	Erecya, Y, 92	0.02	0	0.008	0.003	0-0.024	-	-	K80+I	531.27	-	-	-
s7	414	8.7	2.17	2	Erecya, S, 291 Ophven, R, 369	0.03	2.26	0.022	0.004	0.003-0.065	-	-	HKY	975.7	structural constituent of ribosome	embryo development; hemopoiesis; rRNA processing; ribosomal small subunit biogenesis; regulation of cell cycle; translation	90S preribosome; cytosolic small ribosomal subunit; small-subunit processome; ribosome
supermatrix	9669	6.03	1.34	46	-	0.03	1.46	-	-	-	-	-	HKY+I	19424.3	-	-	-

Supplementary Table 2

Locus	PCR worked	Sequences available	Extra sequences accession	Missing data [%]	anneal T [°C]	Taq	*Bathybates graueri*	*Cyprichromis leptosoma*	*Ophthal-motilapia ventralis*	*Oreochromis tanganicae*	*Cyphotilapia frontosa*	*Eretmodus cyanosticus*	*Tropheus moori*	*Ctenochromis horei*	*Astatotilapia burtoni*	*Lamprologus callipterus*	*Perissodus microlepis*	*Neolam-prologus pulcher*	*Tylochromis polylepis*	*Limno-chromis abeelei*	*Sargochromis spec.*	*Pseudo-crenilabrus philander*	Possible reasons for an unsuccessful sequencing reaction
bmp4	14	13 (14)	AB084658.1*	12.5	52	Redtaq	-/-	+/-	+/+	+/+	+/+	+/+	+/+	+/+	-/- /AB084658.1	+/+	+/+	+/+	+/+	+/+	+/+	+/+	unclear / too much PCR product in sequencing reaction
bmp2	14	12	-	25	52	Redtaq	+/ -	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	+/+	-/-	weak PCR / too much PCR product
fgf6	16	15	-	6.25	52	Redtaq	+/+	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	unclear
furina	14	6	-	62.53	52	Redtaq	+/+	+/ -	+/ -	+/ -	+/ -	+/ -	+/+	+/+	+/+	+/ -	+/ -	+/ -	+/+	-/-	+/+	-/-	unclear / too much PCR product
runx2	15	8	-	50	52	Redtaq	+/ -	+/ -	-/-	+/ -	+/ -	+/+	+/+	+/+	+/+	+/ -	+/+	+/ -	+/+	+/+	+/+	+/ -	too much PCR product / homopolymer / indel
shh	14	9	-	46.62	52	Redtaq	-/-	+/+	+/ -	+/ -	+/ -	+/ -	+/+	+/+	-/-	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	too much PCR product / conflict with other band (only Erecya)
pax9	15	5	-	68.77	52	Redtaq	-/-	+/ -	+/ -	+/ -	+/ -	+/ -	+/ -	+/+	+/+	+/+	+/+	+/ -	+/ -	+/ -	+/ -	+/+	too much PCR product / indel?
sox10b	15	11	-	31.99	52	Redtaq	-/-	+/ -	+/ -	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/ -	+/+	+/+	indel / too much PCR product
ednrb1	16	11	-	31.25	52	Redtaq	+/+	+/ -	+/-	+/ -	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/ -	+/+	+/+	+/+	+/+	increase annealing temperature
mc1r	15	12	-	25	54	Redtaq	+/+	+/ -	evaporated/?	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	evaporated PCRs / high salt concentration / too much PCR product
c-ski	16	14	-	12.51	52	Redtaq	+/+	+/ -	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	unclear / too much PCR product
kita	16	14	-	12.5	54	Redtaq	+/+	+/+	+/ -	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	unclear
mitfa	15	14	-	12.52	54	Redtaq	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/ -	+/+	homopolymer / indel
tyr	15	11	-	32.29	52	Redtaq	+/ -	+/ -	+/ -	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	high salt concentration / too much PCR product
esfr1	16	7(8)	EU042710.1**	50	54	Redtaq	+/+	+/ -	+/- /EU042710.1	+/ -	+/ -	+/ -	+/+	+/+	+/+	+/ -	+/ -	+/ -	+/ -	+/ -	+/+	+/+	unclear
pax3	16	11(14)	unpublished Münzel & Salzburger	12.5	52	Redtaq	+/+	+/- /unpublished sequence	+/ -	+/- /unpublished sequence	+/+	+/+	+/+	+/+	+/+	+/+	+/- /unpublished sequence	+/+	+/+	+/ -	+/+	+/+	unclear
hag	15	13(14)	other primers, unpublished Mever	12.79	52	Redtaq	evaporated/?/unpublished primer	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	evaporated / weak PCR/ too much PCR product
rag	16	8(14)	DQ012234.1*** DQ012223.1*** DQ012219.1*** DQ012236.1*** FJ706526.1**** DO012231.1***	12.5	54	Redtaq	+/+	+/- /DQ012234.1	+/+	+/- /DQ012223.1	+/- /DQ012219.1	+/DQ012236.1	+/+	+/+	+/+	+/-/FJ706526.1	+/+	+/-	+/- /DQ012231.1	+/-	+/+	+/+	unclear / too much PCR product
b2m	13	10	-	37.51	52	Redtaq	-/-	+/ -	+/-	+/ -	-/-	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/+	+/+	unclear/ conflict with other band (in Cyplep and Oretan)?
gapdhs	14	11(12)	other primers, unpublished Mever	25	54	Redtaq	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/+	-/unpublished primer	+/-	+/-	+/+	+/+	homopolymer / indel / too much PCR product
ccng1	16	13	-	18.75	53	Redtaq	+/+	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/-	+/+	+/+	too much PCR product / indel / weak PCR
ptr-like	15	12	-	25	53	Redtaq	+/+	+/+	+/ -	+/+	+/ -	+/+	-/-	+/+	+/+	+/+	+/+	+/+	+/+	+/ -	+/+	+/+	secondary structure hairpin / high salt concentration / too much PCR product
enc1	16	12(14)	GQ168338.1*** ** GQ168285.1*** ** HM050005.1*** ***	12.5	52	Redtaq	+/+	+/- /GQ168338.1	+/+	+/- /HM050005.1/ GQ168285.1	+/ -	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/ -	+/+	+/+	unclear
s7	12	9(11)	GQ168149.1*** ** DQ055082.1*** *****	32.15	72 / 52	Phusiontaq/Redtaq	-/-	+/- /GQ168149.1	+/+	+/+	+/ -	+/+	+/+	+/+	-/-	+/- /DQ055082.1	+/+	+/+	+/+	+/+	-/-	-/-	unclear
super-matrix	-	-	-	26.33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -	

References

*Terai Y, Morikawa N, Okada N
2002
The Evolution of the Pro-Domain of Bone Morphogenetic Protein 4 (Bmp4) in an Explosively Speciated Lineage of East African Cichlid Fishes.
Molecular biology and evolution 19, 1628-1632.

**Salzburger W, Braasch I, Meyer A
2007
Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes.
BMC Biology 5, 51.

***Clabaut C, Salzburger W, Meyer A
2005
Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: nuclear sequences are less homoplasious but also less informative than mitochondrial DNA.
Journal of Molecular Evolution 61, 666-681.

****Nevado B, Koblmüller S, Sturmbauer C, et al.
2009
Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish.
Molecular Ecology 18, 4240-4255.

*****Schwarzer J, Misof B, Tautz D, Schliewen U
2009
The root of the East African cichlid radiations.
BMC Evolutionary Biology 9, 1 86.

******Matschiner M, Hanel R, Salzburger, W
2011
On the origin and trigger of the notothenioid adaptive radiation
PLoS ONE 6 (4), E18911

*******Schelly R, Salzburger W, Koblmüller S, Duftner N, Sturmbauer C
2006
Phylogenetic relationships of the lamprologine cichlid genus Lepidiolamprologus (Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization.
Molecular Phylogenetics and Evolution 38, 426-438.

Supplementary Table 3

Species	bmp4	bmp2	fgf6	furina	runx2	shh	pax9	sox10b	ednrb1	mc1r	c-ski	kita	mitfa	tyr	csfr1	pax3	hag	rag	b2m	gapdhs	ccng1	ptr-like	enc1	s7
	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession	Accession
Bathybates graueri	-	-	JX135153	JX135190	-	-	-	-	JX135142	JX135178	JX135251	JX135216	JX135296	-	JX135196	JX135379	JX135345*	JX135282	-	JX135310	JX135129	JX135239	JX135322	-
Cyprichromis leptosoma	-	JX135270	JX135154	-	-	JX135359	-	-	-	-	-	JX135217	JX135297	-	-	-	-	-	-	JX135311	-	JX135240	-	-
Ophthalmotilapia ventralis	JX135203	JX135271	-	-	-	-	-	-	-	-	-	-	JX135298	-	-	-	JX135346	-	-	JX135312	JX135130	-	JX135323	JX135230
Oreochromis tanganicae	JX135204	-	JX135155	-	-	-	-	-	-	JX135179	JX135252	JX135218	JX135299	JX135368	-	-	JX135347	-	-	JX135313	JX135131	JX135241	-	JX135231
Cyphotilapia frontosa	JX135205	JX135272	JX135156	-	-	-	-	JX135334	-	-	JX135253	-	JX135300	-	-	JX135380	JX135348	-	-	JX135314	JX135132	-	-	-
Eretmodus cyanostictus	JX135206	JX135273	JX135157	-	JX135288	-	-	JX135335	JX135143	JX135180	JX135254	JX135219	JX135301	JX135369	-	JX135381	JX135349	-	JX135168	JX135315	JX135133	JX135242	JX135324	JX135232
Tropheus moori	JX135207	JX135274	JX135158	JX135191	JX135289	JX135360	-	JX135336	JX135144	JX135181	JX135255	JX135220	JX135302	JX135370	JX135197	JX135382	JX135350	-	JX135169	JX135316	JX135134	-	JX135325	JX135233
Ctenochromis horei	JX135208	JX135275	JX135159	JX135192	JX135290	JX135361	JX135265	JX135337	JX135145	JX135182	JX135256	JX135221	JX135303	JX135371	JX135198	JX135383	JX135351	JX135283	JX135170	JX135317	JX135135	JX135243	JX135326	JX135234
Astatotilapia burtoni	-	JX135276	JX135160	JX135193	JX135291	-	JX135266	JX135338	JX135146	JX135183	JX135257	JX135222	JX135304	JX135372	JX135199	JX135384	JX135352	JX135284	JX135171	-	JX135136	JX135244	JX135327	-
Lamprologus callipterus	JX135209	-	JX135161	-	-	JX135362	JX135267	JX135339	JX135147	JX135184	JX135258	JX135223	JX135305	JX135373	-	JX135385	-	-	JX135172	JX135318	JX135137	JX135245	JX135328	-
Perissodus microlepis	JX135210	JX135277	JX135162	-	JX135292	-	JX135268	JX135340	JX135148	-	JX135259	JX135224	JX135306	JX135374	-	-	JX135353	JX135285	JX135173	JX135319*	JX135138	JX135246	JX135329	JX135235
Neolamprologus pulcher	JX135211	JX135278	JX135163	-	-	JX135363	-	JX135341	-	JX135185	JX135260	JX135225	JX135307	JX135375	-	JX135386	JX135354	-	JX135174	-	-	JX135247	JX135330	JX135236
Tylochromis polylepis	JX135212	JX135279	JX135164	JX135194	JX135293	JX135364	-	JX135342	JX135149	JX135186	JX135261	JX135226	-	JX135376	-	-	JX135355	-	-	-	-	-	-	JX135237
Limnochromis abeelei	JX135213	JX135280	JX135165	-	JX135294	JX135365	-	-	JX135150	JX135187	JX135262	JX135227	JX135308	JX135377	JX135200	JX135387	JX135356	-	JX135175	JX135320	JX135139	JX135248	JX135331	JX135238
Sargochromis spec.	JX135214	JX135281	JX135166	JX135195	JX135295	JX135366	-	JX135343	JX135151	JX135188	JX135263	JX135228	-	JX135378	JX135201	JX135388	JX135357	JX135286	JX135176	JX135321	JX135140	JX135249	JX135332	-
Pseudocrenilabrus philander	JX135215	-	JX135167	-	-	JX135367	JX135269	JX135344	JX135152	JX135189	JX135264	JX135229	JX135309	-	JX135202	JX135389	JX135358	JX135287	JX135177	-	JX135141	JX135250	JX135333	-
No of sequences	13	12	15	6	8	9	5	11	11	11	14	14	14	11	7	11	14	6	10	12	13	12	12	9

Chapter 2

A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach

Britta S. Meyer, Michael Matschiner & Walter Salzburger

Molecular Phylogenetics and Evolution (2015) 83, 56-71,
http://dx.doi.org/10.1016/j.ympev.2014.10.009

BSM developed the primer pairs, conducted the laboratory work, analyzed the resulting sequence data and drafted the manuscript. MM participated/helped analyzing the data and participated in manuscript drafting. WS designed the study and helped with the data analyses and participated in manuscript drafting.

Molecular Phylogenetics and Evolution 83 (2015) 56–71

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach

Britta S. Meyer [a,b,*], Michael Matschiner [a,c], Walter Salzburger [a,c,*]

[a] Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland
[b] Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany
[c] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

ARTICLE INFO

Article history:
Received 25 November 2013
Revised 5 October 2014
Accepted 8 October 2014
Available online 26 November 2014

Keywords:
Adaptive radiation
Cichlidae
454 amplicon sequencing
Hybridization
Incomplete lineage sorting

ABSTRACT

The species-flocks of cichlid fishes in the East African Great Lakes Victoria, Malawi and Tanganyika constitute the most diverse extant adaptive radiations in vertebrates. Lake Tanganyika, the oldest of the lakes, harbors the morphologically and genetically most diverse assemblage of cichlids and contains the highest number of endemic cichlid genera of all African lakes. Based on morphological grounds, the Tanganyikan cichlid species have been grouped into 12–16 distinct lineages, so-called tribes. While the monophyly of most of the tribes is well established, the phylogenetic relationships among the tribes remain largely elusive. Here, we present a new tribal level phylogenetic hypothesis for the cichlid fishes of Lake Tanganyika that is based on the so far largest set of nuclear markers and a total alignment length of close to 18 kb. Using next-generation amplicon sequencing with the 454 pyrosequencing technology, we compiled a dataset consisting of 42 nuclear loci in 45 East African cichlid species, which we subjected to maximum likelihood and Bayesian inference phylogenetic analyses. We analyzed the entire concatenated dataset and each marker individually, and performed a Bayesian concordance analysis and gene tree discordance tests. Overall, we find strong support for a position of the Oreochromini, Boulengerochromini, Bathybatini and Trematocarini outside of a clade combining the substrate spawning Lamprologini and the mouthbrooding tribes of the 'H-lineage', which are both strongly supported to be monophyletic. The Eretmodini are firmly placed within the 'H-lineage', as sister-group to the most species-rich tribe of cichlids, the Haplochromini. The phylogenetic relationships at the base of the 'H-lineage' received less support, which is likely due to high speciation rates in the early phase of the radiation. Discordance among gene trees and marker sets further suggests the occurrence of past hybridization and/or incomplete lineage sorting in the cichlid fishes of Lake Tanganyika.

1. Introduction

The species-flocks of cichlid fishes in the East African Great Lakes Victoria, Malawi and Tanganyika (LT) represent the most species-rich adaptive radiations known in vertebrates (see e.g. Kocher, 2004; Salzburger, 2009; Seehausen, 2006). Several hundred of endemic cichlid species have evolved in each of these lakes in only the last few million to several thousand of years (see e.g. Genner et al., 2007; Kocher, 2004; Salzburger, 2009; Salzburger and Meyer, 2004; Snoeks, 2000; Turner et al., 2001; Verheyen et al., 2003). Because of their taxonomic diversity, their ecological and morphological disparity and the high proportion of endemism, East African cichlid fishes are a prime model system in evolutionary biology (reviewed in: Kocher, 2004; Salzburger, 2009; Santos and Salzburger, 2012).

With a maximum estimated age of 9–12 million years (my), LT is the oldest lake in Africa (Cohen et al., 1997; Salzburger et al., 2014) and contains the genetically, morphologically and ecologically most diverse group of cichlid fishes counting ca. 200 species in more than 50 genera (Koblmüller et al., 2008b; Salzburger et al., 2002a; Snoeks, 2000). Based on morphological grounds, Poll (1986) grouped the LT cichlid species into 12 tribes (a taxonomic rank between subfamily and genus): Bathybatini, Cyprichromini, Ectodini, Eretmodini, Haplochromini, Lamprologini, Limnochromini, Perissodini, Tilapiini, Trematocarini, Tropheini,

* Corresponding authors at: Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany (B.S. Meyer). Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland (W. Salzburger).
E-mail addresses: britta-meyer@gmx.de (B.S. Meyer), walter.salzburger@unibas.ch (W. Salzburger).

http://dx.doi.org/10.1016/j.ympev.2014.10.009
1055-7903/

and Tylochromini. Takahashi (2003) revised Poll's tribal assignment and suggested to (i) taking *Boulengerochromis microlepis* out of the Tilapiini into its own tribe, Boulengerochromini, leaving behind *Oreochromis tanganicae* as the only representative of the Tilapiini in LT; (ii) splitting the Limnochromini into Limnochromini *sensu stricto*, Benthochromini and Greenwoodochromini; (iii) establishing a separate tribe, Cyphotilapiini, for *Cyphotilapia frontosa* and *C. gibberosa*; (iv) moving '*Ctenochromis*' *benthicola* into its own tribe; and (v) putting the species of the Trematocarini into the Bathybatini. Only some of these revisions are backed up by molecular data, such as the establishment of the new tribes Benthochromini, Boulengerochromini, and Cyphotilapiini (Koblmüller et al., 2008b; Muschick et al., 2012; Salzburger et al., 2002a). *Greenwoodochromis*, on the other hand, is clearly nested within the Limnochromini in molecular phylogenies (Duftner et al., 2005; Muschick et al., 2012; Kirchberger et al., 2014), and should hence remain within the Limnochromini; the Trematocarini consistently form a separate lineage outside the Bathybatini (see e.g. Koblmüller et al., 2005; Muschick et al., 2012) and should remain in their own tribe (note that Koblmüller et al. (2008b) suggested splitting the Bathybatini into Bathybatini *sensu stricto* and Hemibatini); and '*Ctenochromis*' *benthicola* has recently been identified as member of the Cyphotilapiini (Muschick et al., 2012). Finally, the Tropheini were consistently found to be nested within the Haplochromini (Salzburger et al., 2005, 2002a; see also below) and should, hence, not be considered as separate tribe but as part of the Haplochromini.

Not all of the cichlid tribes occurring in LT are endemic to this lake, though, and four tribes show a distribution range that exceeds the LT basin by far. The Tylochromini have their center of divergence in West Africa (Stiassny, 1990), and the only LT species, *T. polylepis*, is likely to have invaded LT only recently (Koch et al., 2007). The same might be true for *O. tanganicae*, the only native representative of the widely distributed Tilapiini in LT (Klett and Meyer, 2002). Note that the Tilapiini were recently taxonomically revised and that the genus *Oreochromis* has been placed into a new tribe, namely the Oreochromini (Dunz and Schliewen, 2013). The Lamprologini, the most species-rich tribe of cichlids in LT, contain a few species that have secondarily colonized the Congo and Malagarasi River systems (Salzburger et al., 2002a; Schelly et al., 2003; Schelly and Stiassny, 2004; Sturmbauer et al., 2010). The Haplochromini (including the Tropheini) represent the most species-rich tribe of cichlids overall, and are distributed across large parts of Africa, where they have seeded various radiations including the ones of Lake Malawi and the Lake Victoria Region (Koblmüller et al., 2008a; Salzburger et al., 2005; Schwarzer et al., 2012; Verheyen et al., 2003; Wagner et al., 2012). The LT cichlid fishes thus show faunal affinities across a large geographical range to both older cichlid lineages such as the Tylochromini and Tilapiini/Oreochromini and younger ones such as the Haplochromini.

The phylogenetic relationships among East African cichlid tribes has been the subject of various studies over the past two decades, yet remain enigmatic (reviewed in: Koblmüller et al., 2008b). The first comprehensive phylogenetic study of LT's cichlid fishes using molecular information dates back to the early 1990s, when Nishida (1991) used allozyme data to examine the relationships among tribes. He established the so-called 'H-lineage' consisting of the tribes Cyprichromini, Ectodini, Eretmodini, Haplochromini/Tropheini (which he already found to be monophyletic), Limnochromini, and Perissodini as sister-group to the Lamprologini; the Bathybatini, Trematocarini plus *Boulengerochromis microlepis*, *Oreochromis tanganicae*, and *Tylochromis polylepis* were placed outside of a clade formed by the 'H-lineage' and Lamprologini. Yet, the relative position of the 'H-lineage' tribes differed depending on the algorithms used (UPGMA and neighbour-joining; NJ) (Fig. 1a). Sturmbauer and Meyer (1993) used two mitochondrial (mt) DNA markers (cytochrome *b* and control region) and suggested, based on phylogenetic analyses with NJ and maximum parsimony (MP), a sister-group relationship between the Cyprichromini and the Ectodini and between the Eretmodini and the Haplochromini (Fig. 1b). Kocher et al. (1995) established the mitochondrial NADH dehydrogenase subunit 2 (ND2) gene as marker for phylogenetic analyses in cichlid fishes and provided the most inclusive phylogenetic hypothesis for LT cichlids so far. In their MP and NJ phylogenies, the Bathybatini, the Tylochromini, *B. microlepis* and *O. tanganicae* formed a clade, and the Eretmodini were placed outside the 'H-lineage', as sister-group to the Lamprologini (Fig. 1c). The Cyprichromini were resolved as the sister-group to all remaining 'H-lineage' taxa (i.e. without the Eretmodini). Using three mitochondrial markers (control region, cytochrome *b*, ND2) and NJ, MP and maximum-likelihood (ML) phylogenetic analyses, Salzburger et al. (2002a) confirmed the position of *B. microlepis*, the Bathybatini and the Trematocarini outside all other tribes occurring in Lake Tanganyika, with the exception of the Tylochromini, and the Eretmodini were placed as sister-group to the Lamprologini and the remaining 'H-lineage' tribes (Fig. 1d). Within the 'H-lineage', the Ectodini appeared as the sister to the remaining taxa. This study was also the first to establish phylogenetic affinities between the LT cichlid fishes and the riverine genus *Orthochromis* (not shown in Fig. 1d; see also Salzburger et al., 2005). Clabaut et al. (2005) combined sequences of the mitochondrial ND2 gene and the nuclear recombinase activating gene (*rag*) and applied ML and Bayesian inference (BI). They placed the Eretmodini as sister-group to the Lamprologini and established the 'C-lineage', i.e. the 'H-lineage' of Nishida (1991) but without the Eretmodini. Within this 'C-lineage', the Limnochromini plus *C. frontosa* appeared as the sister-group to the Perissodini, the Ectodini, the Cyprichromini and the Haplochromini (Fig. 1e). Day et al. (2008) provided one of the most comprehensive datasets to date (cytochrome *b*, ND2) including 157 taxa. Their ML and BI phylogenies supported the existence of the 'C-lineage' by placing the Eretmodini as sister-group to the Lamprologini. In their analyses, a clade formed by the Ectodini and Cyprichromini was placed as the sister-group of the remaining 'C-lineage' taxa (Fig. 1f). In the ML phylogeny of Muschick et al. (2012), who used the mitochondrial ND2 gene and two nuclear markers (*ednrb1*, *phpt1*), the Eretmodini were placed as sister group to the Lamprologini and the 'C-lineage', within which the Limnochromini appeared outside of all other included taxa (Fig. 1g). The study of Friedman et al. (2013), which was based on ten nuclear makers and did not focus specifically on the species of LT but on a larger cichlid phylogeny, revealed a clade formed by the Lamprologini, the Perissodini plus the Cyprichromini, and the Cyphotilapiini plus the Limnochromini as sister group to the Ectodini, the Eretmodini and the Haplochromini (Fig. 1h).

In summary, after more than 20 years of research, the composition of individual LT tribes has been well investigated, whereas the phylogenetic relationships among these cichlid tribes remain largely elusive. All studies performed so far revealed different results (Fig. 1), and the support values for many of the deeper nodes were consistently low. While there is consensus about the position of *T. polylepis*, *O. tanganicae*, the Bathybatini, Boulengerochromini and Trematocarini outside of the other tribes, the following main areas of uncertainty persist: (i) the relative position of the Bathybatini, Boulengerochromini and Trematocarini to each other; (ii) the placement of the Eretmodini, which were suggested as either being part of the 'H-lineage' and sister to the Haplochromini (Friedman et al., 2013; Nishida, 1991; Sturmbauer and Meyer, 1993), as sister-group to the Lamprologini (Clabaut et al., 2005; Day et al., 2008; Kocher et al., 1995), or as separate lineage outside the Lamprologini-'C-lineage' clade (Muschick et al., 2012; Salzburger et al.,

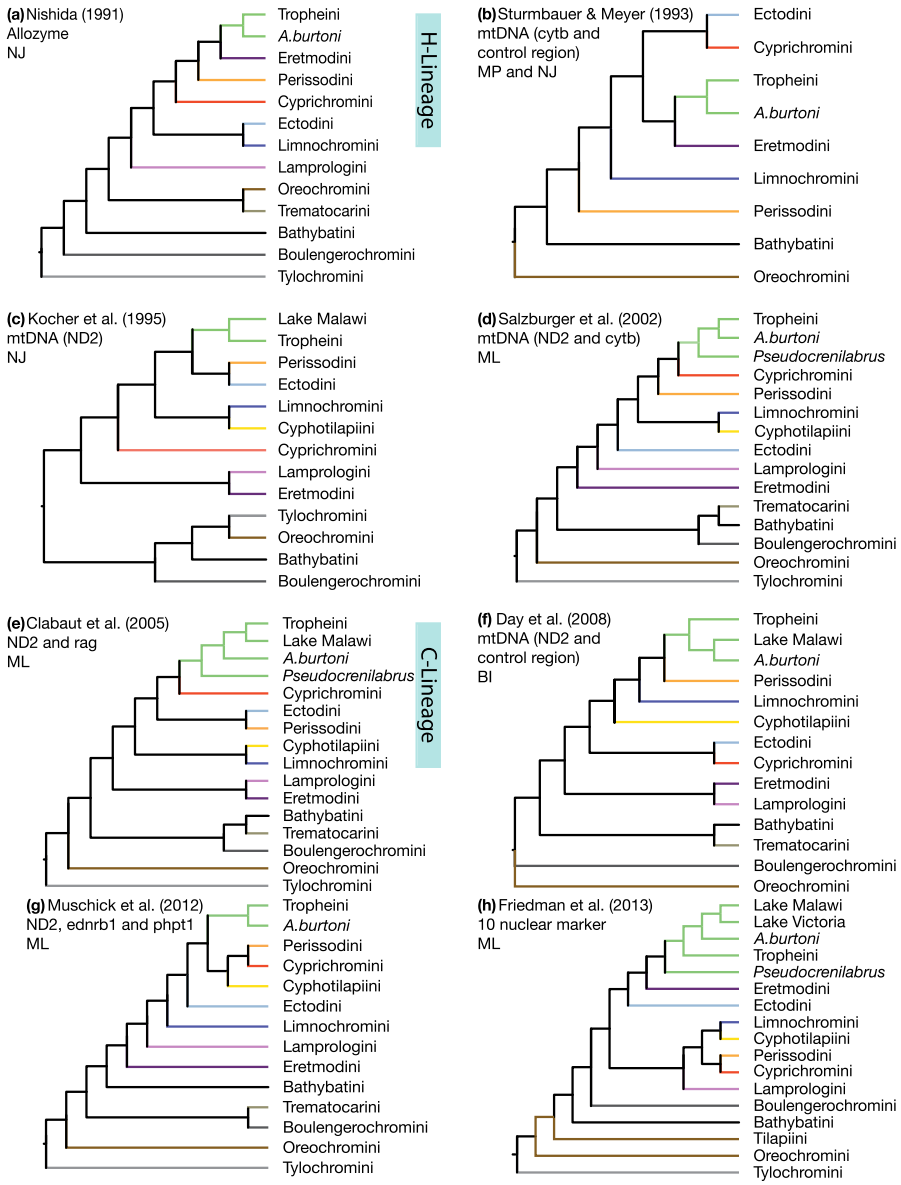


Fig. 1. Previous hypotheses for the phylogenetic relationships among cichlid tribes in Lake Tanganyika. The figure depicts simplified cladograms based on the studies of (a) Nishida (1991), (b) Sturmbauer and Meyer (1993), (c) Kocher et al. (1995), (d) Salzburger et al. (2002a), (e) Clabaut et al. (2005), (f) Day et al. (2008), (g) Muschick et al. (2012), and (h) Friedman et al. (2013). The markers used in the respective study and the phylogenetic algorithms applied are indicated; the color code for cichlid tribes follows that of Muschick et al. (2012). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2002a); and (iii) the relative position of the 'H-lineage'/'C-lineage' taxa with respect to each other.

The apparent intricacy with resolving the phylogenetic relationships of the cichlid tribes in LT might have various reasons. First, the conflict between the various phylogenetic hypotheses might in part result from the different phylogenetic algorithms used (see above), although this would not apply to the more recent studies, all of which relied on ML and BI methods. Second, we might face the problem here that the previously used markers do not provide enough power of resolution for the question at hand. Alternatively, the inability to resolve some of the phylogenetic relationships of LT's cichlid tribes might reflect biological reality in the context of an adaptive radiation, where speciation is not necessarily bifurcating and multiple lineages may evolve nearly contemporaneously from a common ancestor ('soft polytomy' versus 'hard polytomy' problem: Maddison, 1989; Slowinski, 2001; Sturmbauer et al., 2003; Walsh et al., 1999; Whitfield and Lockhart, 2007). Conflicting topologies may also be the result of reticulate evolution due to (introgressive) hybridization, which is a commonly observed phenomenon in LT's cichlid assemblage (e.g. Koblmüller et al., 2007; Salzburger et al., 2002b) and might have acted as trigger of cichlid adaptive radiations in the first place (Joyce et al., 2011; Seehausen,

2004). Finally, discordance between different sets of markers could reflect incomplete lineage sorting, which is expected to have a strong impact on phylogenetic inference in rapidly diversifying clades (Kubatko and Degnan, 2007) and has been demonstrated in LT cichlid fishes before (Takahashi et al., 2001).

With decreasing sequencing costs and increasing computational resources, single marker and mtDNA-based phylogenies are rapidly being replaced by phylogenies inferred from large-scale nuclear marker sets based on selected loci, transcriptomes, or even whole genomes (McCormack et al., 2013). This recent development enables comparisons between the phylogenetic histories of multiple sets of individual markers. Here, we analyze the phylogenetic history of cichlid fishes from LT on a tribal level, including representatives from the East African Lakes Victoria and Malawi. We sampled 45 species and 42 nuclear loci and thus assembled the largest DNA sequence dataset available for LT cichlid fishes to date. In order to account for potential hybridization and incomplete lineage sorting, we explore gene tree concordance in addition to concatenation as ways for species tree estimation. We further test the strength of our dataset using random resampling of different numbers of markers.

2. Material and methods

2.1. Sample collection and DNA extraction

Specimens for this study were collected between 2007 and 2011 at the Kafue River (Kafue National Park) and at LT in the Northern Province of the Republic of Zambia following the standard operating procedure described in Muschick et al. (2012). Additional samples were obtained from aquaria stocks at the University of Basel and at EAWAG, Kastanienbaum, Switzerland. In total, we analyzed data for 45 specimens, each representing a different East African cichlid species. Our sampling comprised 34 cichlid species from LT covering all major cichlid lineages in this lake. In addition we included 11 further species of riverine clades and from Lakes Victoria and Malawi, to place the LT cichlid taxa into a larger phylogenetic context. A detailed list of specimens, their IDs and sample locations is provided in Table S1. Genomic DNA was extracted from ethanol preserved tissue of whole specimens (see Muschick et al., 2012 for details).

2.2. Marker selection, sequencing and quality control

To infer the phylogenetic history of the cichlid fishes of LT on the basis of an informative set of nuclear (nc) DNA markers, we selected a set of 42 nuclear loci. Twenty-four primer pairs were taken from earlier studies (Meyer and Salzburger, 2012; Muschick et al., 2012; Won et al., 2005) and 18 primer pairs were newly designed following the strategy described in Meyer and Salzburger (2012). In short, we selected genes with known functions and aimed for amplification products between 400 and 600 bp in length to enable the application of next-generation amplicon sequencing. Twenty-four of the markers were developed as exon-primed intron crossing (so-called EPIC) primers (Lessa, 1992; Slade et al., 1993). The markers for *enc1*, *ptr*, *tbr* and *snx33* were taken from Li et al. (2007), but modified to meet our requirements. The same strategy was applied for *ednrb* (Lang et al., 2006), *bmp4* (Albertson et al., 2003), and the reverse primer of *s7* (Chow and Hazama, 1998). The genome of the Nile Tilapia (*Oreochromis niloticus*) (Brawand et al., 2014) was used to define exon–intron boundaries and UTRs. A detailed list of all primers, their base composition, the length of the amplification products, their source, the ENSEBML reference of the respective locus in Tilapia, the chromosomal position of the respective locus in the Medaka genome and the number of variable sites are provided in Tables 1 and 2.

The 42 nuclear markers were PCR amplified in several separate multiplex reactions in a final volume of 25 μL on a Veriti or 2720 thermal cycler (both Applied Biosystems, Rotkreuz, Switzerland). All PCR reactions contained the Multiplex PCR Kit (QIAGEN, Hombrechtikon, Switzerland) and a primer mix including eight to ten barcoded primer pairs (0.1 μM of each primer), water, and template DNA (5–20 ng/μL). We used barcoded fusion primers synthesized by Microsynth (Balgach, Switzerland). The PCR conditions were standardized for all reactions with an initial heat activation phase of 95 °C for 15 min, followed by 35 amplification cycles with denaturation steps at 94 °C for 30 s, annealing steps at 60–62 °C for 90 s and extension steps at 72 °C for 90 s; reactions were completed by a final extension phase at 72 °C for 10 min.

To remove small fragments, residual primers and primer-dimers, we applied the Agencourt AMPure XP magnetic bead system following the manufacturer's protocol (Beckman Coulter, Nyon, Switzerland) and using a bead/DNA ratio of 1:1. Purification results were inspected with a 2100 Bioanalyzer (Agilent, Basel, Switzerland) using the DNA 1000 Kit. The amplification products of five individual PCR reactions with different primer combinations were then pooled (on the basis of the concentration measurements with the Bioanalyzer) to obtain the final libraries containing all 42 markers of one individual. In a second pooling step, 16 barcoded individuals were pooled for one 1/16th run on a 454 PicoTiterPlate. The subsequent library handling and sequencing was conducted by Microsynth (Balgach, Switzerland) with the GS FLX system (454 Sequencing, Roche). Sequencing was unidirectional starting at the forward primer, which also contained the barcodes.

Individual sequences (in both fasta and fastq format) were separated and extracted with Roche's sffinfo tool (described in 454 Sequencing System Software Manual Version 2.6). Quality control was conducted with the software PRINSEQ (v0.20.3) (Schmieder and Edwards, 2011). We excluded individual reads that were shorter than 150 bp, that had an average Phred quality score below 15, or that contained more than 1% unidentified bases coded as "N". In a second step, we filtered out exact duplicates. The assembly to reference sequences from the *A. burtoni* genome (Brawand et al., 2014) was performed with the software bwa and the BWA-SW algorithm (the Burrows-Wheeler Aligner's Smith-Waterman Alignment) (Li and Durbin, 2010). The resulting SAM files were imported into Geneious (v6.1.6–7.0.3, Biomatters Ltd, Auckland, New Zealand; available from http://www.geneious.com), visually inspected, if necessary reassembled, and further trimmed (we allowed a 0.05 error probability limit and a maximum of 10 low quality bases at the 3′ end). The final consensus sequences for each individual and marker were constructed with a 50% threshold, where bases were called "N" if the Phred score was below 20. Sequence data has been deposited on GenBank under the accession numbers KP129679-KP131427 (see Table S2 for details) and KM263618-KM263752 (Santos et al., 2014).

2.3. Alignment and sequence characterization

Sequences for each locus were aligned with the software MAFFT (v7.017) (Katoh and Standley, 2013), using the "—auto" option. Resulting alignments were visually inspected and manually improved when obvious sequencing artefacts (e.g. homopolymers) were observed or homology appeared questionable.

Overall mean distance for each locus was calculated with the software MEGA (v5.2.1) (Tamura et al., 2011) as the total number of differences and the *p*-distance. This was done for all ingroup taxa (i.e. excluding *Tylochromis polylepis*), with pairwise deletion for missing and ambiguous data. For the concatenated alignment the within group mean distance was also calculated for the three most species-rich lineages, the Haplochromini, the Lamprologini and the Ectodini.

Table 1
List of the 42 markers used in this study. The marker name, the forward and reverse sequence of each primer, the Ensembl Gene-ID for the respective locus in Tilapia, the link to the Ensemble entry for Tilapia, the chromosomal position of each locus in Medaka and the reference for the primer sequences are provided.

Name (synonym)	Forward primer [5′-3′]	Reverse primer [5′-3′]	Ensembl-Gene-ID	Link to Ensembl	Chr Medaka	Reference
rag1	TCGGCGCTTTCGGTACGATGTG	TGCCCCTGAAGTGGAASSGA	ENSONIG00000014593	RAG1	6	Meyer and Salzburger (2012)
b2m	GCCACGTGAGTRATTTCCACCCC	ACGCTAYACRGYGGACYCTGA	ENSONIG00000014176	B2M	23	Meyer and Salzburger (2012)
gapdhs	CCCTGGCCAAAGTCATCCACGATA	CACCACTGACACATCGGCCACT	ENSONIG00000007262	GAPDHS	16	Meyer and Salzburger (2012)
Ptchd4	GCGGGTAGTGAATGTGAGTGCG	ACCCAAGACACCCAGCTCCA	ENSONIG00000006708	PTCHD4	24	Meyer and Salzburger (2012)
enc1	CRGTTCGCCTTGCGCTRTTGC	TGGGTGCCGCCTTTGACCAT	ENSONIG00000020511	ENC1	12	Meyer and Salzburger (2012)
phpt1	AGCAGGGTTGACCTTCTCAA	TGGCTAAAATCCCCGATGTA	ENSONIG00000002175	novel gene	4	Muschick et al. (2012)
rps7	CGTGCCATTTTACTCTGGACTKGC	AACTCGTCYGGCTTCTCGCC	ENSONIG00000018698	RPS7	24	Meyer and Salzburger (2012)
tbr1	ATCGTGCCGGGTGCGAGATA	AGGACGGCGTCTCAATCCAGCT	ENSONIG00000008933	TBR1	21	This study
aqp1a.1	ATCAACCCTGCTCGCTCCTTCG	TGCATCGTTGCCTCCGTTGACG	ENSONIG00000009446	novel gene	17	This study
hprt1	TCAGYGATGAGGAGCAGGGTTATG	CGACCGTCATTGGGATGGAGC	ENSONIG00000017584	HPRT1	10	This study
anxa4	TGGACGAGGCCCAGGCTATTCAAG	ACGTCTTCCAGGCAGCCAGACA	ENSONIG00000003465	ANXA4	12	This study
pgk1	CGGTACCTCCCTGTATGACGAGGA	GCAGCCAGATTTGGTCACCTCGA	ENSONIG00000017337	PGK1	14	This study
bmp4	GAGGACCCATGCCCATTCGTTT	GCCACTATCCAGTCATTCCAGCC	ENSONIG00000001366	BMP4	22	Meyer and Salzburger (2012)
bmp2	AGGCCCTGGCCAGCCTAAAA	TCCTGCGTCTGTGGGCATCCTT	ENSONIG00000000958	BMP2	24	Meyer and Salzburger (2012)
TMO-4C4	TTATGCTGAGGTGTTTGGCCTAC	CCACAGCACCCTCCTCATAAAT	ENSONIG00000017439	novel gene	–	This study
fgf6b	CGCAAAGGTGCCACTACAG	TCGCACTGCACGGATGCAAA	ENSONIG00000000017	FGF6 (2 of 2)	23	Meyer and Salzburger (2012)
runx2	CGGGGTTGGTGTTTGAGGGCAA	GCTGACATGGTGTCACTGTGCTGA	ENSONIG00000001025	RUNX2	24	Meyer and Salzburger (2012)
furina	GCTGCATGGGGACAGACAGTCA	ATAGTCACTGGCACCCGCCACA	ENSONIG00000005696	FURIN (1 of 2)	3	Meyer and Salzburger (2012)
wnt7b	GCGTCTCGGGATCCTGTACCACTA	TGCAGGTAAACACCTCCGTCCT	ENSONIG00000008839	WNT7B	6	This study
pax9	TCCCACGGCTGTGTCAGYAA	ACAGAGTGCGAGGAAGGCCA	ENSONIG00000000990	PAX9	–	Meyer and Salzburger (2012)
sox10b	TSCRGGGTCTGGGAAACCTCAT	TGGTGGTCGGCGTATTCTGCAA	ENSONIG00000008392	SOX10 (1 of 2)	8	Meyer and Salzburger, 2012
otx2	GCAGAACAAAGTGCGACCTGCC	GTCTGCTGTGGAGTTGAAGCCCA	ENSONIG00000020156	OTX2	22	This study
otx1	TACACCTCCTGCTGTCTCCAGCAC	ATAGATGAGGCCGTCATGGGGC	ENSONIG00000001278	OTX1 (1 of 2)	15	This study
dlx2a	ATCGCCAACTCCCGCAGACA	TCCGTTGAAGYGCAGCCAGT	ENSONIG00000008722	DLX2	21	This study
dlx4b	GCGTGGATTTCTTCCAGGCTGTC	CTGTGTGCTCTAATCTGCTGTGGG	ENSONIG00000019896	DLX4 (1 of 2)	19	This study
barx1	TCTCGCAGAGTCTCTCGGTCTG	TCGCTGCTGGGGATGGAGTT	ENSONIG00000003234	BARX1	–	This study
ednrb1a	CGTTGGCCTGCACTGCCATT	AGGCAGCCAGCACAGAGCAAA	ENSONIG00000018701	EDNRB (1 of 2)	17	Meyer and Salzburger (2012)
mc1r	GACCACGGCCTCCTGGATGT	GTTGCAGAAGGGGCTGGTGG	ENSONIG00000021393	MC1R	3	Meyer and Salzburger (2012)
skia	CGACCAGCTGGAGATCCT	TCCTCTTGTACTTGTTGGCG	ENSONIG00000017935	SKI (1 of 2)	7	Meyer and Salzburger (2012)
kita	CAGAGTACTGCTGTTTCGGMGAT	GGCTAAGAACTCCATGCCTTTGGC	ENSONIG00000002981	KIT (1 of 2)	4	Meyer and Salzburger (2012)
mitfa	CCTGGCATGAAGCARGTACTGGAC	TTGCYAGAGCACGAACTTCRGC	ENSONIG00000020270	MITF (2 of 2)	5	Meyer and Salzburger (2012)
tyr	TGGGTGGACGCAACTCCCTT	TGGCAAATCGGTCCATGGGT	ENSONIT00000006471	TYR (1 of 2)	13	Meyer and Salzburger (2012)
hagoromo (fbxw4)	AAACTGGTACARYGGGVTCTGC	AGCGRCAGACGTCACCCTTGT	ENSONIG00000013182	HAGOROMO	15	Meyer and Salzburger (2012)
slc45a2 (aim)	GAGCTATGGACTGGGGTCAC	TGGCTGTTTGACACTTGAGG	ENSONIG00000007610	SLC45A2	12	Won et al. (2005)
rh1	TCGCCTTGGCTGCAATCTGG	ACCATGCCGGGTGACTTCCCT	ENSONIG00000021142	RH1	7	This study
opn1mw (lws)	ATTGCTGCTCTTTGGTCCCTGACA	AGCCAGAGGGTGGAAGGCAT	ENSONIG00000020292	OPN1MW	5	This study
opn1sw (sws)	TGGGTCACACGCTGTGTGCT	CAGCAGCTGGGAGTAGCAGAARA	ENSONIG00000007620	OPN1SW	scaffold1021	This study
ccng1	CTGCTTGCCCTGGCTCTCCT	AGCTGACTCAGGTATGGTCGGA	ENSONIG00000012912	CCNG1	10	Meyer and Salzburger (2012)
snx33	TGGCTGTACAACCGCCTGCT	CCAAYRTGAATGCSTGGCTGA	ENSONIG00000012857	SNX33	6	This study
rpl13a	ACCTGGCTTTCCTGCGCAAGA	TTGCGAGAGGGCTTCAGACGCA	ENSONIG00000003560	RPL13A	22	This study
edar	TGAGCAGCTGTTGAGCCGCA	CRCATKGCARGYYCTGGCATACA	ENSONIG00000004260	EDAR	21	this study
csf1ra	AAGCACAGATGGGACACGCC	TGTACTGGCCCTGCTCCTGT	ENSONIG00000013065	CSF1R (1 of 2)	10	Meyer and Salzburger (2012)

Table 2
Characterization of the 42 loci used in this study. The marker name, the alignment length of each marker, the sequenced gene regions, the number of variable (*V*) and parsimony informative (PI) sites in the ingroup taxa, the mean number of differences (genetic distance) and the *p*-distance in the ingroup taxa, and the assignment to one of six subsets according to the CONCATERILLAR analysis are specified for each marker.

Name (synonym)	Alignment lengths	Gene regions	V sites ingroup	PI sites ingroup	Genetic distance	*p*-distance	Subset
rag1	418	Exon	49	21	5.10	0.012	1
b2m	478	Exon, intron, UTR	93	50	12.88	0.031	2
gapdhs	458	Exon, intron	57	15	4.35	0.01	4
Ptchd4	394	Exon	32	11	3.59	0.009	4
enc1	376	Exon	21	7	2.95	0.008	5
phpt1	459	Exon, intron	67	31	7.14	0.017	1
rps7	470	UTR	77	31	9.24	0.021	4
tbr1	466	Exon	13	6	1.58	0.003	5
aqp1a.1	440	Exon, intron	62	24	5.69	0.014	2
hprt1	402	Exon, intron	45	14	5.12	0.014	1
anxa4	642	Exon, intron	56	20	6.31	0.014	1
pgk1	377	Exon, intron	40	16	3.55	0.01	3
bmp4	456	Exon	47	16	4.37	0.011	4
bmp2	372	Exon	26	8	1.78	0.005	1
TMO-4C4	428	Intron	54	32	8.02	0.019	2
fgf6b	471	Exon, intron	29	7	2.64	0.006	2
runx2	360	Exon, intron, UTR	16	5	2.06	0.006	1
furina	311	Exon, intron	34	8	2.88	0.009	2
wnt7b	389	Exon	16	4	1.41	0.004	2
pax9	394	Exon	22	7	2.20	0.006	1
sox10b	378	Exon	40	15	4.43	0.012	2
otx2	412	Exon	19	7	1.89	0.005	1
otx1	356	Exon	15	9	1.86	0.005	5
dlx2a	497	Exon, intron	83	27	6.94	0.015	2
dlx4b	356	UTR, exon	29	7	2.43	0.007	4
barx1	220	Exon, intron	30	11	3.47	0.019	1
ednrb1a	438	Exon, intron	59	28	6.82	0.016	6
mc1r	426	Exon	30	9	2.71	0.007	1
skia	453	Exon	38	11	2.67	0.006	2
kita	431	Exon, intron	45	20	4.93	0.012	2
mitfa	434	Exon, intron	57	21	6.41	0.016	6
tyr	525	Exon, intron	72	26	8.47	0.019	3
hagoromo (fbxw4)	493	Exon, intron	110	59	16.01	0.043	2
slc45a2 (aim)	286	Exon	38	16	4.55	0.016	3
rh1	404	Exon	43	32	9.59	0.024	6
opn1mw (lws)	420	Exon, intron	53	22	6.65	0.017	1
opn1sw (sws)	450	Exon, intron	80	36	10.01	0.024	1
ccng1	460	Exon, intron	69	20	6.55	0.017	1
snx33	437	Exon	43	19	5.10	0.012	1
rpl13a	370	Exon, intron	28	9	3.00	0.013	4
edar	372	Exon, intron	41	13	3.29	0.009	2
csf1ra	366	Exon, intron	54	19	5.29	0.015	2

2.4. Gene tree discordance tests

We first tested for topological incongruence between individual gene trees, using hierarchical likelihood ratio tests as implemented in the software CONCATERPILLAR (v1.7.2) (Leigh et al., 2008), with default settings and the assumption of linked branch lengths. As part of the CONCATERPILLAR analysis, tree inference was performed using RAxML (v7.2.8) (Stamatakis, 2006), assuming a single GTR substitution model for each sequence alignment. The two largest sets of markers identified by CONCATERPILLAR to have concordant histories (containing 13 and 14 markers, respectively) were each concatenated and subjected to phylogenetic analyses as described below.

2.5. Phylogenetic analysis of concatenated datasets

In brief, sequence alignments for sets of loci were concatenated according to different strategies (see below) and phylogenetic analyses were based on both maximum likelihood with GARLI-PART (v2.0.1019) (Zwickl, 2006) and RAxML (v7.7) (Stamatakis, 2006), and on Bayesian inference with MrBayes v3.2.1 (Ronquist et al., 2012). Prior to tree inference, sequence alignments were subdivided according to gene region (exons, introns and UTRs) and codon position, and the optimal substitution models and partitioning schemes for these subdivisions were selected with the greedy algorithm of PartitionFinder (v1.1.1) (Lanfear et al., 2012) applying the Bayesian information criterion (BIC), and always taking into account substitution models available in the respective tree inference software (Schwarz, 1978). Phylogenetic analyses were run locally or at the CIPRES Science Gateway (Miller et al., 2010) and at Bioportal (Kumar et al., 2009).

We first inferred the phylogeny for each of the two largest sets of loci with concordant histories according to CONCATERPILLAR. To this end, sequence alignments of all markers included in each set were concatenated. We then used concatenation of the full set of 42 loci to infer the phylogenetic history of LT cichlid fishes. This method assumes that all markers share a common evolutionary history and that discordant signals resulting from homoplasies can be counterbalanced by extensive and genome wide marker sampling (Rokas et al., 2003). While the assumption of a common evolutionary history seems to be violated at least for the analysis of the full marker set, concatenation may still lead to correct phylogenetic estimates when the true tree lies outside of the "anomaly zone" (Kubatko and Degnan, 2007). As there is no fully unlinked branch length option in GARLI, analyses were run with linked branch lengths (subsetspecificrates = 1, linkmodels = 0) and partitioning schemes and substitution models selected by PartitionFinder with respective settings (branchlengths = linked,

models = all, resulting in 17 distinct partitions for the full-concatenated dataset). A total of 50 independent ML inferences were conducted in GARLI, with the termination condition set to at least 10,000 generations without any substantial (0.01) topological enhancement. Node support was assessed with 500 replicates of non-parametric bootstrapping with the same settings. Bootstrap values were mapped to the ML topology with SumTrees (v3.3.1), using the DendroPy Phylogenetic Computing Library (v3.12.0) (Sukumaran and Holder, 2010).

ML phylogenies with unlinked partition-specific branch lengths were estimated with RAxML, using the -M option and applying a partitioning scheme obtained by a PartitionFinder analysis (settings: branchlengths = unlinked, model = raxml, resulting in 2 partitions). For the ML inference, we used RAxML's rapid hill-climbing algorithm and the GTR + GAMMA model in 50 alternative runs and with 500 bootstrap replicates each.

Likewise, MrBayes analyses were conducted with unlinked branch lengths (unlink brlens = (all), prset ratepr = fixed) and a partitioning scheme estimated by PartitionFinder (settings: branchlengths = unlinked, model = mrbayes, resulting in 2 partitions). Using the default prior probability distribution (exponential prior with a mean of 0.1) on branch lengths, two independent MrBayes runs were conducted with four chains for 10,000,000 MCMC generations, sampling every 100th generation, and discarding the first 25% as burn-in. All other settings were left at their defaults. Convergence of MCMC was assessed by MrBayes' Potential Scale Reduction Factor (PSRF) reaching 1.0, and the average standard deviation of split frequencies falling below 0.01. We further evaluated effective sample sizes in Tracer (v1.5) (Rambaut and Drummond, 2007) and plotted posterior probabilities of splits over the MCMC run with AWTY online to test for convergence of runs (Nylander et al., 2008).

To examine the phylogenetic signal contained in length-mutational events and to evaluate the potential power of a combined analysis (alignment plus indel information), the indels from the concatenated alignment were translated into a presence/absence matrix. This was performed with the software SeqState v1.4.1 (Müller, 2005) using the simple indel coding procedure (SIC) (Simmons and Ochoterena, 2000). Phylogenetic inference for these two datasets was conducted with GARLI, applying the Mkv model of Lewis (2001), and otherwise using default settings as described above.

2.6. Gene tree summary statistics and Bayesian concordance analysis

In order to visualize potentially conflicting signal contained in the 42 loci, gene trees for each individual marker were inferred using GARLI with settings as specified in Section 2.5. The 50 best topologies from each run and from all 42 markers (a total of 2100 gene trees) were used to generate an average consensus tree in SplitsTree (v4.12.3) (Huson and Bryant, 2006). The implemented "average consensus tree" function constructs a neighbor-net using the average pairwise distances of the individual trees.

As a further approach to investigate the discordance among the sampled gene trees and to combine conflicting data in a primary concordance and a population tree, we applied a Bayesian concordance analysis (BCA) (Ane et al., 2007; Baum, 2007), as implemented in the software BUCKy v1.4.0 (Larget et al., 2010). Using samples of MrBayes' posterior tree distribution as input, this analysis accounts for both uncertainty in individual gene trees and potential discordance among trees inferred from different loci. The primary concordance tree, as estimated by BUCKy, visualizes the most dominant history from several gene trees, along with concordance factors (CF) indicating the proportion of loci supporting a given clade (Baum, 2007). In addition, a population tree with coalescent units as branch lengths is generated by BUCKy, based on quartets of concordance factors. This population tree is known to be consistent in the presence of incomplete lineage sorting (Chung and Ané, 2011; Larget et al., 2010).

In order to apply BUCKy, MrBayes was used to infer gene trees from the individual loci, with substitution models and partitioning schemes selected by PartitionFinder (assuming linked branch lengths for all subdivisions of each locus). For each locus, we conducted two replicate MrBayes runs with six chains of 15 million generations, sampling every 100th generation. As reported by Willis et al. (2013), we found that for most loci, all of the 150,000 sampled trees represented unique topologies, suggesting a lack of resolution in some parts of the tree. This could partly be due to polytomies, which would be displayed as multiple weakly supported topologies with very short branches in MrBayes, as this software only provides fully resolved trees. To reduce the large number of distinct tree topologies, we pruned our dataset to 14 taxa, keeping only one representative per tribe (as our primary interest was a tribal level phylogeny). This deletion was done with the pruning option in BUCKy. The BUCKy analysis was conducted with 4 runs, 10 chains and 500,000 generations per chain. The alpha prior, which represents the a priori expected level of discordance, was set to 1–100.

2.7. Testing the strength of the phylogenetic signal as a function of dataset size

In order to test whether our dataset contains a sufficiently large number of markers to recover the "true" phylogenetic history of LT cichlids, we randomly resampled and concatenated different numbers of markers, and produced ML phylogenies from these sets. We then measured the topological difference between the tree resulting from one set of randomly chosen markers and the tree resulting from the complete set including of all markers and between the trees resulting from two different and mutually exclusive marker sets. As our full dataset contained 42 markers, the first comparisons were done for 1–41 randomly chosen markers, whereas the latter was performed for 1–21 randomly chosen markers. For each number of markers between 1 and 41, we compiled 20 sets drawn at random from the full set of 42 markers. Then, for each of the sets containing at most 21 markers, a comparison set was produced containing the same number of markers so that the two sets did not share any marker. In order to take into account marker concordance according to the results of the CONCATERPILLAR analysis (see Section 2.4.) we repeated the same procedure for 1–13 markers, again with 20 replications each. For the latter analysis, we always compiled two sets of markers, so that markers shared a concordant history within each set, but a discordant history between the two sets (according to CONCATERPILLAR). All generated marker sets were subjected to phylogenetic analysis with GARLI (see above, Section 2.5.), using marker-specific partitions and substitution models as suggested by PartitionFinder. Topological differences between resulting ML trees were measured by means of their *K*-score (Soria-Carrasco et al., 2007), as the *K*-score accounts for variable substitution rates between marker sets.

Then, *K*-scores of 20 replicate comparisons were plotted against the number of markers used in the datasets for which the respective ML trees had been inferred (see Camargo et al., 2012; Willis et al., 2013). We expected a general decrease of mean *K*-scores (i.e., fewer topological differences) with increasing marker number due to an increase in the phylogenetic signal for larger datasets. We further expected *K*-scores between a tree based on randomly drawn markers and the tree based on the full dataset of 42 markers to approach zero for marker numbers close to 42, as the alignments used for the reconstruction of the two trees would become increasingly similar. Nevertheless, we expected the degree to which *K*-scores decrease with increasing number of markers to inform

about the minimum number of markers needed to reliably construct the relationships among cichlid tribes in LT.

As an additional measure of discordance, we tested for statistically significant topological differences between the tree based on all 42 markers, and trees based on smaller datasets, using the Shimodaira–Hasegawa (Shimodaira and Hasegawa, 1999) and Approximated Unbiased (Shimodaira, 2002) tests as implemented in PAUP* (v.4.0a129) (Swofford, 2003). For each number of markers between 1 and 41, we plotted the number of tree replicates that fitted the full dataset significantly worse than the tree produced from all 42 markers.

3. Results

3.1. Sequencing

Amplicon sequencing was successful for most of the 42 markers for the 45 taxa. In total, we obtained 98.3% of the 1890 possible sequences. Of 789,525 bp in the final alignment, 26,854 bp (3.40%) consisted of gaps; 27,211 bp (3.45%) were undetermined ("N") and 476 bp (0.06%) were ambiguous ("WRYSMK" coded).

3.2. Alignment and sequence characterization

The concatenated alignment had a total length of 17,545 bp, of which 1932 positions (11.01%) were variable and 769 positions (4.38%) were parsimony informative (not considering the outgroup *Tylochromis polylepis*). The amount of variable sites per marker varied between 13 and 110 sites (average: 46, median: 43), the number of parsimony informative sites ranged between four and 59 (average: 18.3, median: 16) (Table 2). The average sequence length for each marker was 417.7 bp (median: 423 bp), and the average total number of differences across all sequence pairs was 208.8 (uncorrected *p*-distance: 0.013). Within three of the major lineages, we found that the Ectodini showed the highest divergence (114.1 differences; uncorrected *p*-distance: 0.007), followed by the Lamprologini (110.4; 0.007) and the Haplochromini (all species included; 103.1; 0.006). Separate analyses of the within group mean distance of the haplochromines of the three lakes indicated a higher number of base differences between the four species of Lake Malawi (14.3; 0.0009) than the four species of Lake Victoria (6.8; 0.0004). The Tropheini (*Ctenochromis horei, Lobochilotes labiatus, Gnathochromis pfefferi, Tropheus moori*) included in this study showed a higher level of diversity (73.0; 0.004).

3.3. Gene tree discordance tests

We used CONCATERPILLAR to test for topological incongruence between markers and to identify concordant sets of markers. Based on hierarchical likelihood ratio tests, CONCATERPILLAR detected six sets of markers that were concordant internally, but exhibited significant levels of discordance (p-value < 0.001) between them. The three largest sets contained 14, 13, and 6 markers, respectively, whereas the remaining three sets included 3 markers each (the assignment of each marker to one of these subsets is indicated in Table 2). The six sets exhibited no obvious clustering of markers according to gene function, coding and non-coding parts, or variability. The two largest sets of markers were subjected to individual phylogenetic analysis. Subset 1 (14 markers) contained a total of 5872 concatenated bp, of which 10.30% were variable and 3.92% were parsimony informative. The average pairwise distance was 61.05 mutational steps, and the uncorrected *p*-distance was 0.012. Subset 2 (13 markers) had a length of 5507 bp, with 12.69% variable sites, and 5.25% parsimony informative sites. This marker-set showed a somewhat higher variability (average pairwise distance: 76.07; uncorrected *p*-distance: 0.015).

3.4. Phylogenetic analysis of concatenated datasets

Phylogenetic analysis of concatenated subsets revealed conflicting topologies between subset 1 and subset 2 (Fig. 2a and b). While the base of the resultant trees (i.e. the position of the Bathybatini, Boulengerochromini and Trematocarini) was highly similar, the topologies differed with respect to the relative placement of the Eretmodini, the Lamprologini, the Limnochromini and the Cyprichromini/Perissodini clade. For subset 1 the three inferred topologies from the different analyses were congruent. In these trees, the Lamprologini were nested within the mouthbrooding tribes of the 'H-lineage', of which the Cyprichromini/Perissodini clade branched off first. The Lamprologini were resolved as sister group to the Limnochromini in BI (BPP 0.81), and the same relationship was weakly supported in GARLI and RAxML inferences (BS 37 and 45). The Ectodini were placed as sister group to a clade formed by Cyphotilapiini, Eretmodini and the Haplochromini (GARLI BS 26, RAxML BS 35, BPP 0.89).

The phylogenetic analyses of subset 2 revealed a monophyletic group containing the 'H-lineage' taxa (BS 99, BPP 1.0), which were placed as sister taxon to the Lamprologini. The Eretmodini branched off first, and the Ectodini were consistently grouped together with a clade formed by Cyprichromini/Perissodini, the Cyphotilapiini and the Limnochromini (BS 32-34, BPP 0.90). In general, the interrelationships of tribes received only moderate support, which is likely a consequence of the comparatively small number of markers in this subset (see Section 3.6). Excluding *Eretmodus cyanostictus* from these two phylogenetic analyses did not change the resulting tree topologies (data not shown).

The trees obtained with the entire concatenated dataset of 42 markers were highly congruent and most nodes were very well supported (mean GARLI BS 79.2; mean RAxML BS 78.1; mean BPP 0.941). Fig. 3b depicts the ML tree inferred with GARLI; the ML tree obtained with RAxML and the 50% majority rule consensus tree of our MrBayes analysis are shown in Fig. S1. In all three trees, *Oreochromis tanganicae* appeared as the sister to *Tilapia sparrmanii* and a strongly supported clade formed by the remaining tribes (GARLI BS 100, RAxML BS, 100, BPP 1.0). The monophyly of these tribes was strongly supported (BS 100, BPP 1.0 for all tribes of which more than two representatives have been included). Within this group *T. nigrifrons* and *B. graueri* appeared as sister taxa (BS 100, BPP 1.0) in all our analyses. The three tribes Boulengerochromini (represented by their only member, *B. microlepis*), Trematocarini (represented by *T. nigrifrons*), and Bathybatini (represented by *B. graueri*) appeared outside of a strongly supported clade (BS 100, BPP 1.0), in which the substrate spawning Lamprologini, the most species-rich tribe within LT, are clearly separated from the mouthbrooding tribes (i.e. Cyphotilapiini, Cyprichromini, Ectodini, Eretmodini, Haplochromini, Limnochromini, Perissodini; BS 73-75, BPP 1.0).

The branching order within the mouthbrooding tribes of the 'H-lineage' received less support, and there was incongruence between the tree topologies resulting from the different analyses with respect to the placement of the Cyphotilapiini and the Limnochromini relative to each other, and regarding the first divergence events within the Haplochromini (indicated by dotted lines in Fig. 3b). The Cyprichromini were consistently resolved as the sister group of Perissodini (BS 100, BPP 1.0), and the clade formed by these two tribes represented the sister of all remaining tribes of the 'H-lineage' in all analyses of the full-concatenated dataset. The Limnochromini and the Cyphotilapiini formed a monophyletic group that was sister to a clade combining the Ectodini, the Eretmodini, and the Haplochromini (GARLI BS 65, RAxML BS 59, BPP

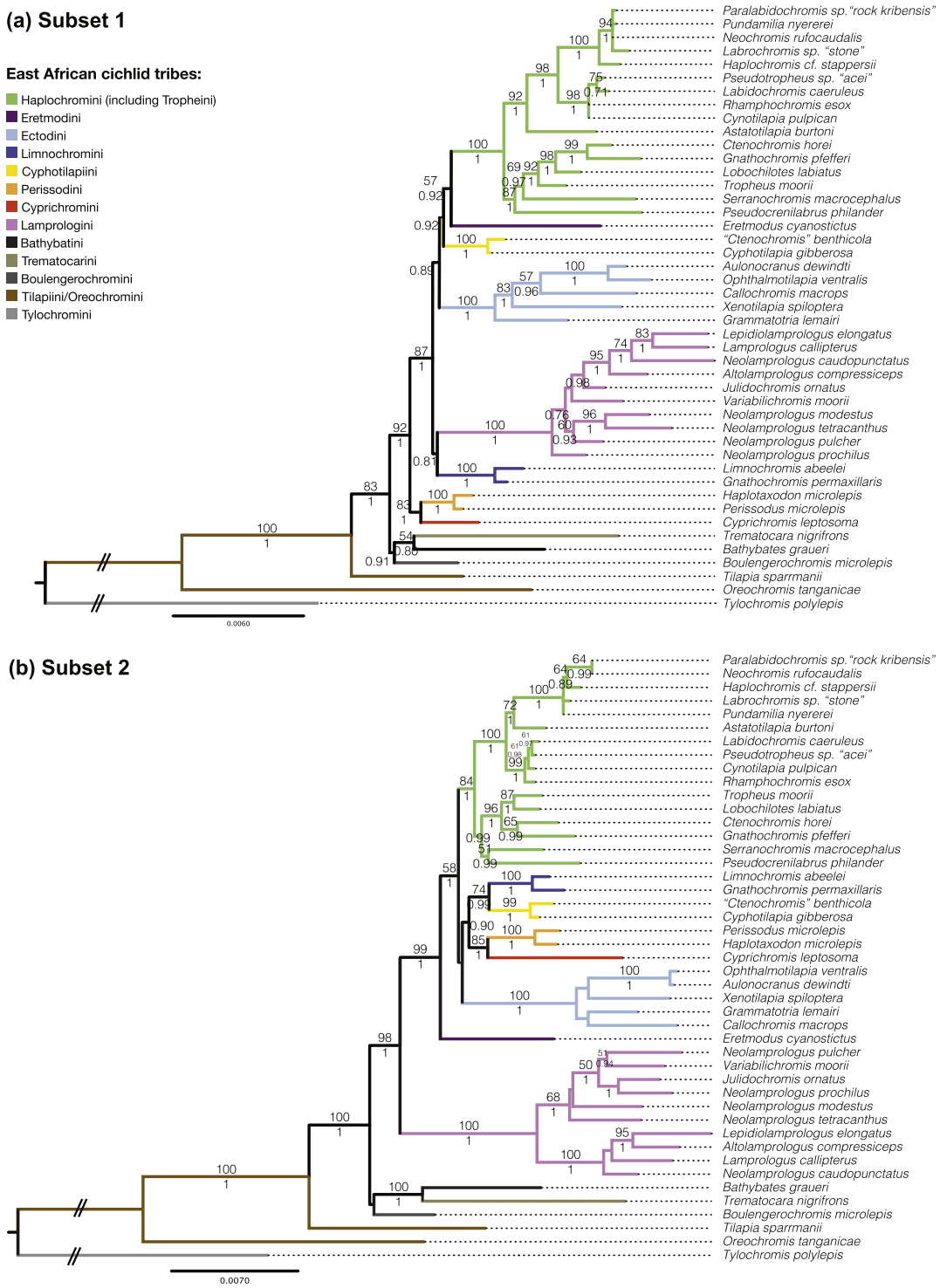


Fig. 2. Results from the phylogenetic analyses based on the two largest subsets of markers identified with CONCATERPILLAR. (a) Maximum likelihood phylogeny of subset 1 (14 markers; see Table 2) inferred with GARLI. (b) Maximum likelihood phylogeny of subset 2 (13 markers; see Table 2) inferred with GARLI. Numbers above the branches represent maximum likelihood bootstrap support values ($\geqslant$50%) as obtained with GARLI, numbers below the branches indicate Bayesian posterior probabilities ($\geqslant$0.75) as revealed with MrBayes. The branch leading to the outgroup taxon, *Tylochromis polylepis*, is shortened by one third. The colors indicate the affiliation of each taxon to one of the cichlid tribes.

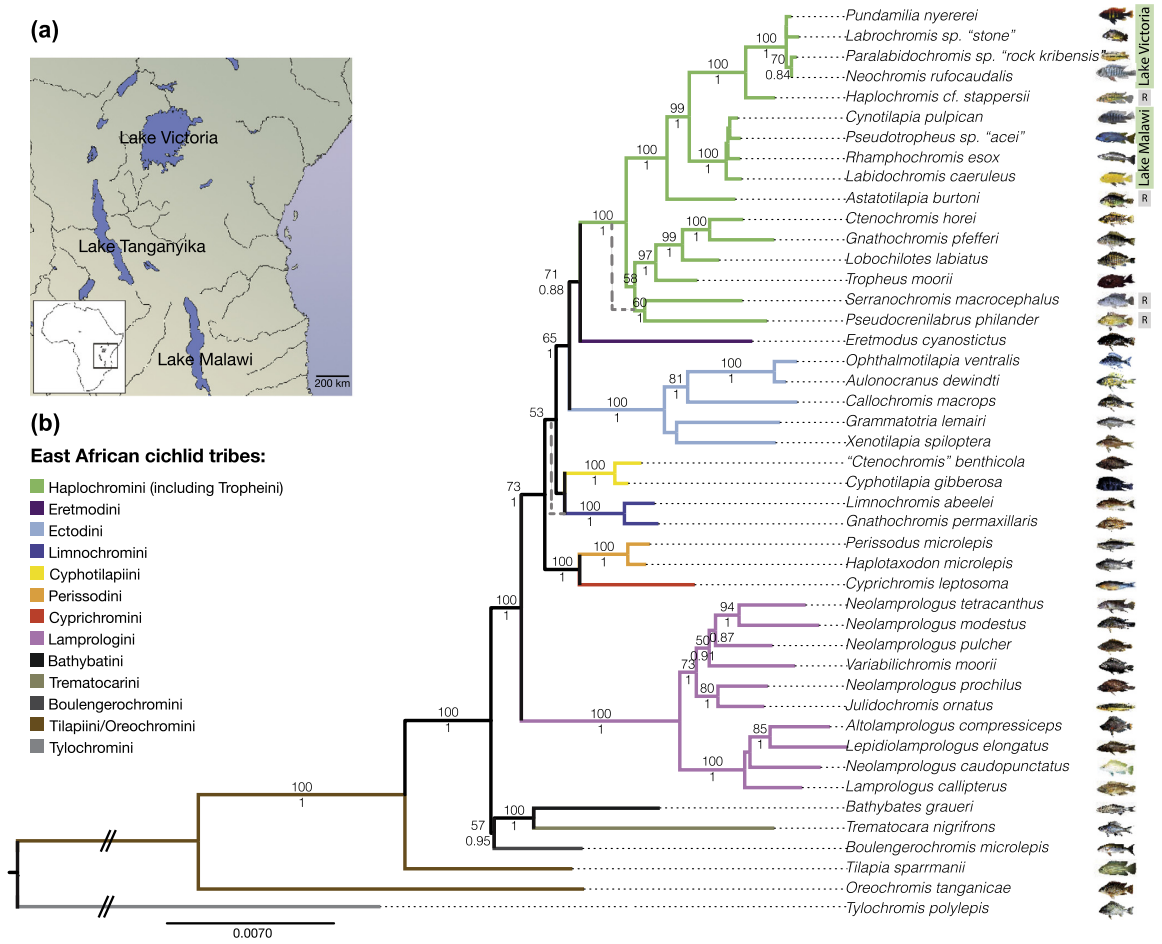


Fig. 3. Tribal level phylogeny of the Lake Tanganyika cichlid fishes. (a) Map of the area showing the three East African Great Lakes. (b) Maximum likelihood tree based on the concatenated dataset (17,545 bp) as obtained from a partitioned analysis with GARLI. Numbers above the branches indicate maximum likelihood bootstrap support values (⩾50%) produced with GARLI, numbers below the branches represent Bayesian posterior probabilities (⩾0.75) as revealed with MrBayes. Alternative branching orders between the maximum likelihood analysis with GARLI (as shown here) and the maximum likelihood analysis with RAxML (Fig. S1a) and Bayesian inference with MrBayes (Fig. S1b) are indicated with dotted lines; the branch leading to *Tylochromis polylepis* was shortened by one third; colors indicate the tribal affiliation of each taxon. Sample origin other than LT are indicated with boxes on the right; R = riverine. Fish pictures were taken in the field, except for *P. nyererei* and *R. esox* (credit: E. Schraml), *P. rockkribensis* (credit: M. Negrini) and *L. sp.* 'stone' (credit: O. Seehausen).

1.00) in the GARLI analysis, whereas the Cyphotilapiini appeared closer to this clade according to the RAxML and MrBayes analyses. Within this clade, the representative of the Eretmodini (*E. cyanostictus*) was consistently placed as sister group to the Haplochromini (GARLI BS 71, RAxML BS 50, BPP 0.88). Similarly, the species from Lake Victoria and Lake Malawi appeared reciprocally monophyletic (BS 100, BPP 1.0) within the Haplochromini. *Haplochromis cf. stappersii* from LT was resolved as sister taxon to the Lake Victoria cichlids (BS 100, BPP 1.0). The riverine species *Astatotilapia burtoni* was always placed outside of the species-flocks of the Lake Malawi and Victoria cichlids (BS 100, BPP 1.0). The haplochromines *Serranochromis macrocephalus* and *Pseudocrenilabrus philander* were either put into a separate clade (in RAxML and BI), or placed together with the LT haplochromines (Tropheini) (with GARLI).

Translating all indels of the 42 loci into a binary code resulted in a dataset comprising 167 positions, of which 70 were parsimony informative. A phylogenetic hypothesis obtained with this dataset with GARLI was, overall, concordant with the trees resulting from the concatenated dataset. However, while the monophyly of most tribes and the position of the Eretmodini as sister group to the Haplochromini was recovered, the respective support values were generally low and the position of most of the tribes relative to each other could not be recovered (see Fig. S2).

3.5. Gene tree summary statistics and Bayesian concordance analysis

Inferring single gene trees from 42 genes and 45 taxa with both GARLI and MrBayes (data not shown) resulted in 42 alternative topologies with some to numerous polytomies or low support values for certain branches, whereas other parts of the trees were well resolved. Fig. 4 shows the average consensus network of 2100 trees with 168 splits representing the conflicting affinities within the individual gene trees at the base of the tribes. The tribes themselves seem clearly defined and show only few alternative splits.

For the Bayesian concordance analysis with BUCKy, we pruned the dataset to one representative per tribe (Fig. 5). Changes in the alpha prior had no influence in the topology of both primary concordance and population tree. Its topology (with alpha default

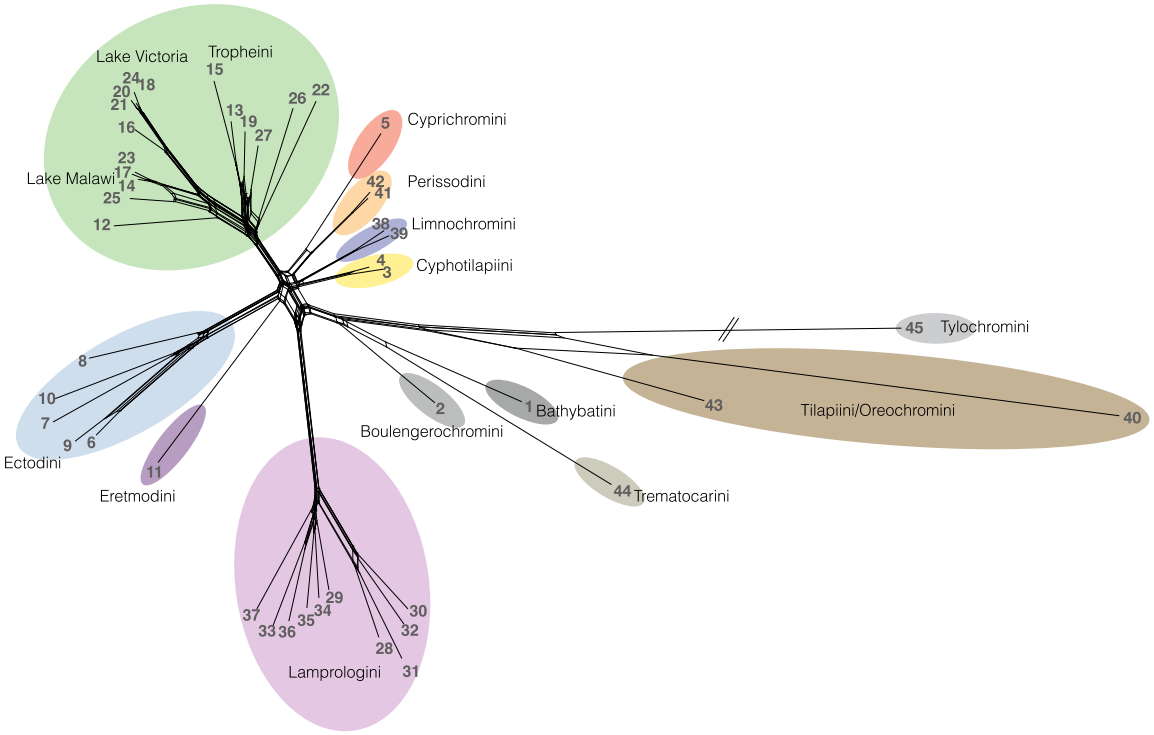


Fig. 4. Average consensus neighbor-net inferred with SplitsTree4 from average pairwise distances in the best gene trees obtained from 50 GARLI runs for each marker (2,100 trees). Note that in this consensus network each gene tree estimate contributed equally and that differences in alignment lengths, degrees of variation, and uncertainties (e.g. bootstrap values) among markers are not considered. The color code is the same is all other figures, the numbers refer to the different species (see Table S1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

prior) is mostly consistent with the species tree inferred from the full-concatenated dataset (see above; Fig. 3). However, one topological disagreement was found regarding the position of *Boulengerochromis microlepis*, which was placed as a sister group to the clade composed of the Lamprologini and the representatives of the 'H-lineage' (including the Eretmodini) in the population tree, but clustered with the Trematocarini and the Bathybatini in the primary concordance tree. Within the population tree the Eretmodini were again resolved as sister group to the Haplochromini. This close relationship is also reflected in the concordance factors of splits within the primary concordance tree (see Text S1).

3.6. Strength of the phylogenetic signal as a function of dataset size

After 20 repetitions of random resampling and concatenation of 1–41 markers, we used GARLI to infer ML phylogenies from all replicate marker sets, and compared the resulting trees between each other and with the optimal tree based on the full concatenated dataset of 42 markers, in order to test the strength of the phylogenetic signal as a function of dataset size. We expected topological differences between two trees to decrease with increasing size of the respective marker sets as shown in Camargo et al. (2012). Different types of comparisons were performed: Between one tree based on 1–41 markers and the tree resulting from the full marker set (Fig. 6a), between two trees produced from mutually exclusive sets containing 1–21 markers (Fig. 6b), and between two trees based on mutually exclusive sets of 1–13 markers found to be internally concordant but externally discordant in topology according to the CONCATERPILLAR analysis (Fig. 6c).

As expected, topological differences between two trees, as measured by their *K*-score, generally decreased with increasing marker number; the steepest decrease was observed for marker numbers between 1 and 8–10. The median *K*-score between one tree based on a randomly compiled marker set of a given size and the tree based on the full set of 42 markers was always lower than median *K*-scores between two trees based on randomly compiled marker sets of the same size (Fig. 6a versus b). Furthermore, topological comparisons involving the tree based on the full marker set generally resulted in a lower variance of *K*-scores than comparisons between two trees that were produced from randomly sampled mutually exclusive marker sets. In the latter case, the two trees represent independent phylogenetic estimates and are thus particularly useful to assess variance in discordance as a function of marker set size. For this type of comparisons, *K*-scores appear relatively constant for datasets combining between 11 and 21 markers. Nevertheless, *K*-scores between trees based on 21 markers (mean 0.0111) are significantly lower than those between trees constructed from sets of 16 markers (mean 0.0140, *t*-test *p*-value = 0.01613) or less (means $\geqslant$ 0.0128, *t*-test *p*-values $\leqslant$ 0.01704). For most marker set sizes, mean and median *K*-scores of two trees based on mutually exclusive marker sets were slightly lower when all markers with a set were concordant according to the CONCATERPILLAR analysis (Fig. 6c) compared to when sets were composed of randomly sampled markers (Fig. 6b). This reduction was significant for marker sets with eight markers or more (*t*-test *p*-values $\leqslant$ 0.0295), with the exception of sets

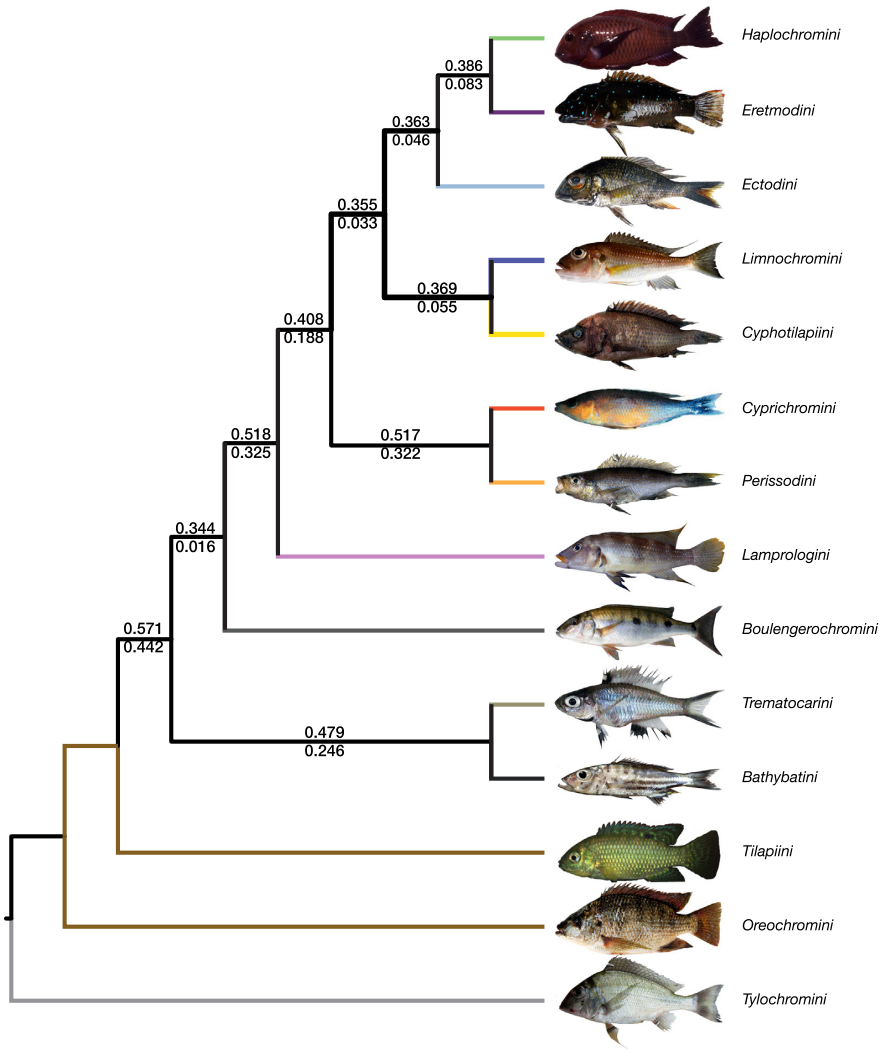


Fig. 5. Population tree topology from the Bayesian concordance analysis (conducted with BUCKy) of 14 taxa representing the different cichlid tribes in LT. Numbers above the branches represent the averaged concordance factors, numbers below are coalescence units (see Text S1 for further details). Fish pictures and color codes are the same as in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

containing eleven markers (*t*-test *p*-value = 0.0881), which suggests that the discordance between the two largest marker sets identified by CONCATERPILLAR is lower than that between randomly compiled marker sets of the same size.

Similarly, the number of marker set replicates, for which ML trees differ significantly from the ML tree based on 42 markers, shows an overall decrease with increasing size of the respective marker sets. For concatenated sets of 1–5 markers, and for sets of 8 markers, phylogenies produced from all 20 replicate sets are significantly different to the full ML tree, according to both the SH and the AU tests. On the other hand, for concatenated sets of 34 or more markers, none of the phylogenies based on these sets differ significantly from the tree obtained with the full set of markers, according to either of the two tests. Between these extremes, we observe a general decrease in the number of rejected tree replicates with increasing number of markers, based on which these trees were produced (Fig. 6a).

4. Discussion

The present study is the most extensive phylogenetic analysis of cichlid fishes in East African Lake Tanganyika with respect to the number of nuclear DNA markers and the total length of the ncDNA sequences analyzed. The main goal of our work was to establish a robust phylogenetic hypothesis for the relationships among the cichlid tribes of LT, which has so far been inferred on the basis of mtDNA or relatively few nuclear markers only (Clabaut et al., 2005; Day et al., 2008; Friedman et al., 2013; Kocher et al., 1995; Muschick et al., 2012; Salzburger et al., 2002a; Sturmbauer and Meyer, 1993).

The comparatively high information content provided by mtDNA sequences and the availability of universal primers were the main reasons for the utilization of mtDNA markers in earlier phylogenetic analyses aiming to resolve the relatively young and rapid radiation of cichlid fishes in LT. Among the many drawbacks

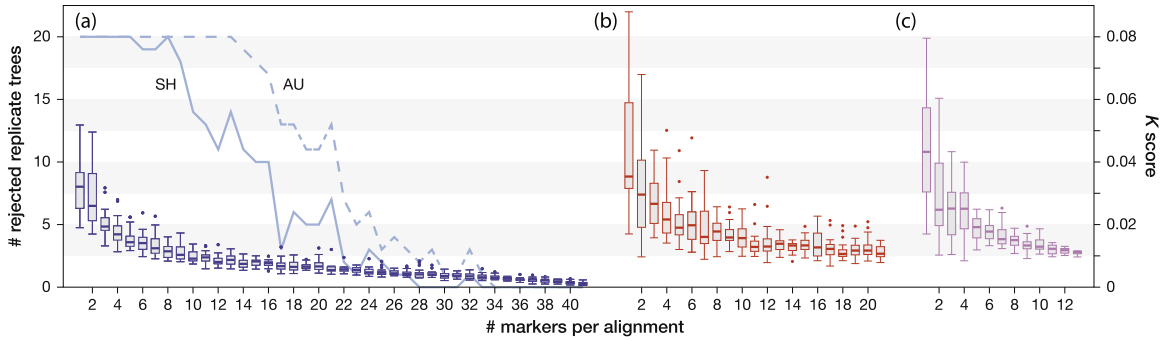


Fig. 6. Topological differences between ML trees measured by their *K*-scores as a function of the number of randomly resampled and concatenated markers. (a) *K*-scores between trees based on randomly sampled and concatenated markers and the tree based on the full dataset of 42 markers. Light blue lines indicate the number of tree replicates (out of a total of 20 replicates) significantly different to the tree based on the full dataset, according to the Shimodaira–Hasegawa (SH) test (solid line), and the Approximately Unbiased (AU) test (dashed line). (b) *K*-scores between two trees that are both based on mutually exclusive randomly sampled marker sets of the given size. (c) As (b), but strictly grouping concordant markers in each set (according to CONCATERPILLAR, see text). Boxplots are based on 20 replicates of each comparison. Whiskers indicate the lowest *K*-score still within 1.5 inter-quartile range of the lower quartile, and the highest *K*-score still within 1.5 inter-quartile range of the upper quartile. Outliers are indicated with dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of mtDNA markers are that only maternal inheritance patterns are captured and that past events of introgression and hybridization remain largely invisible (Ballard and Whitlock, 2004). In addition, a single locus (irrespective of being based on mtDNA or ncDNA) might not accurately reflect the species tree, as individual gene trees often differ from the true species tree (Pamilo and Nei, 1988). Nuclear DNA markers, on the other hand, usually contain fewer variable sites thus less phylogenetic signal. Clabaut et al. (2005) showed, for example, that in LT cichlids, ncDNA datasets would need to contain about ten times more sequence data to obtain the same quantity of phylogenetic information as provided by mtDNA markers – a task not reached by any previous study.

Here we took advantage of the 454 next-generation pyrosequencing technology and compiled a ncDNA dataset for LT cichlids containing 42 markers in well characterized genes and reaching a total alignment length of 17,545 bp. We chose a locus re-sequencing strategy with barcoded primers in order to obtain long enough sequence reads and to sample a large number of gene histories. Primers were chosen to bind in more conserved exons and to amplify (if possible) more variable intron regions (Meyer and Salzburger, 2012).

4.1. Single gene-tree discordance and evaluation of the strength of the phylogenetic signal

Not surprisingly, the individual single locus datasets did not contain enough phylogenetic information to accurately resolve the phylogenetic relationships among the cichlid tribes of LT. Most single locus trees were not very well resolved, the branch support values in these trees were generally rather low, and all 42 single locus topologies differed at least to some extent (in part because of the occurrence of polytomies; not shown). Overall, however, many of the single locus topologies follow a general trend as is illustrated in the average consensus network shown in Fig. 4. Many branches, and especially the monophyly of cichlid tribes, are well supported across the datasets. However, the consensus network indicates certain areas of uncertainties, which might result from hybridization and/or incomplete lineage sorting or simply reflect the low power of resolution in some of the individual markers (see below).

In order to estimate the strength of the phylogenetic signal as a function of dataset size and to evaluate whether our dataset contained enough phylogenetic information, we applied a strategy that compares tree topologies inferred from randomly chosen datasets with varying numbers of markers per alignment on the basis of their *K*-scores (Camargo et al., 2012). More specifically, we compiled datasets from 1 to 41 randomly chosen markers (in 20 replications each) and compared the ML trees based on these marker sets to the tree produced in the same way from the full dataset containing all 42 concatenated markers. Obviously, and as expected, the topologies resulting from the randomly drawn marker sets become increasingly similar to the best tree obtained with 42 markers the more markers are included in each concatenated dataset (Fig. 6a). Also, differences between equally large and mutually exclusive marker sets generally decrease with increases in the number of markers included in both sets (Fig. 6b). The same decrease was observed when trees were produced from two sets of markers that were identified as topologically concordant within each set, but discordant between sets (Fig. 6c). However, topological differences were generally slightly lower when marker sets were discordant to each other (Fig. 6c). This was unexpected but could in part be explained if the phylogenetic histories of marker sets 3–6 (which are included in Fig. 6b, but excluded from Fig. 6c) are even more discordant than those of marker sets 1 and 2.

Importantly, while all tree topologies resulting from datasets of 1–5 markers were significantly distinct from the best tree according to both SH and AU tests, inferred trees become successively more similar with an increasing number of markers, and statistically indifferent from the best tree when more than 34 markers are included (light blue lines in Fig. 6a). These results suggest that our full dataset is large enough to reliably resolve the phylogenetic history of the LT cichlid fishes. Whether or not an extension of our marker set to even more than 42 markers would provide additional phylogenetic signal remains to be tested.

4.2. A threefold strategy for phylogenetic analyses in LT cichlids

In order to account for potential problems with dataset concatenation (see below), we opted to apply three strategies to analyze our data. In a first step, we performed ML and BI phylogenetic analyses with a concatenated dataset containing all 42 markers of all 45 species. These analyses were based of the naïve assumptions that all gene histories equally reflect the species tree, and that the 'true' phylogenetic signal should dominate over phylogenetic noise in a large enough dataset (Rokas et al., 2003). The usage of the concatenated dataset is further backed up by our phylogenetic analyses of randomly chosen subsets of varying numbers of

markers, which demonstrate that the phylogenetic signal improves with increasing number of included markers (Fig. 6).

Although concatenation of multiple markers is often thought to improve accuracy (Bayzid and Warnow, 2013; Chen and Li, 2001; Rokas et al., 2003; but see Salichos and Rokas, 2013), this approach assumes that genes share a common evolutionary history, and it has been shown that violation of this assumption can lead to strongly supported yet incorrect phylogenies (Degnan and Rosenberg, 2009; Gadagkar et al., 2005; Kubatko and Degnan, 2007; Salichos and Rokas, 2013). One situation, in which concatenation may lead to inconsistent species tree estimates, is incomplete lineage sorting (Degnan and Rosenberg, 2009; Kubatko and Degnan, 2007; Yang and Rannala, 2012). We thus, in a second approach, applied a gene tree discordance test with CONCATERPILLAR to evaluate the incongruence between individual gene trees. This test suggested the existence of six sets of markers that were concordant within them, but discordant between each other. The two largest sets, containing 14 and 13 markers respectively, were then subjected to in-depth phylogenetic analysis.

As a third strategy, we performed a Bayesian concordance analysis with BUCKy, which accounts for uncertainty and variability in the individual locus phylogenies and has been shown to deal well with incomplete lineage sorting (Chung and Ané, 2011; Knowles and Kubatko, 2011; Yang and Warnow, 2011). In this analysis, we pruned our dataset to one species per tribe.

Overall, the three strategies applied to analyze our multi-marker dataset resulted in congruent topologies. All analyses confirm the monophyly of the LT tribes (in cases where more than one representative was included; this does, hence, not apply to the BUCKy analysis with the reduced taxon set). In all analyses, the Tylochromini, Oreochromini and Tilapiini were resolved outside of all other included species. The representatives of the Trematocarini and the Bathybatini always formed a clade, and were, together with *B. microlepis* (Boulengerochromini), consistently placed as sister-group to the remaining cichlid tribes; the Cyprichromini and Perissodini always clustered together. Furthermore, in all analyses except in those based on subset 1 of CONCATERPILLAR, the Lamprologini were resolved as sister group to the 'H-linage' consisting of Cyphotilapiini, Limnochromini, Cyprichromini, Perissodini, Ectodini, Eretmodini and Haplochromini. In all analyses, the Eretmodini appear as a member of the 'H-lineage' and, with one exception (i.e. subset 2 of CONCATERPILLAR), appear as sister-group to the Haplochromini.

Within the 'H-lineage', the relationships of the cichlid tribes differed between the three approaches. Especially the analysis of subset 1 of CONCATERPILLAR revealed a rather different topology, whereas in subset 2 the relative position of the Eretmodini and Ectodini varied in comparison to the other approaches. Note, however, that the two largest subsets of markers identified by CONCATERPILLAR contain only 14 (subset 1) and 13 markers (subset 2), respectively. Our analyses have shown that sets with as many as 34 markers can still produce significantly different trees for the same set of taxa. The phylogenetic hypotheses resulting from these small marker sets (Fig. 2a and b) should thus be taken with caution.

Taken together, we believe that, in our case, the concatenation of all markers is a justified strategy (Fig. 3), as it leads to the best-supported tree topologies, which are backed-up by similar results in both the average consensus network (Fig. 4) and the Bayesian concordance analysis (Fig. 5). The concatenation strategy is further supported by our phylogenetic signal tests, which show that the largest datasets lead to significantly more robust topologies (Fig. 6), whereas the subsets suggested by CONCATERPILLAR may not contain enough phylogenetic information. At the same time, these tests indicate the presence of a sufficient phylogenetic signal in the concatenated dataset, so that remaining uncertainties in the resultant tree topologies (GARLI, RAxML and MrBayes analyses of concatenated dataset and subsets) should not be due to lacking power of resolution ('soft polytomy' problem). Instead, it appears that the remaining uncertainties in our trees, most notably the phylogenetic relationships among 'H-lineage' tribes (see Figs. 2–4), are due to high speciation rates at the onset of radiation of the LT mouthbrooders ('hard polytomy' problem), past events of hybridization, and/or the persistence of ancestral polymorphisms. It has previously been recognized that it is notoriously difficult to resolve, with the available methodology, the phylogenetic relationships among lineages that emerged from adaptive radiation events (Glor, 2010), which is not least due to the fact that such tree topologies are expected to be 'bottom-heavy' (Gavrilets and Vose, 2005).

4.3. Conclusions

With this study, we present a novel hypothesis for the phylogenetic relationships among East African cichlid tribes, which is based on the largest set of ncDNA sequences so far, and which differs from all previous hypotheses (Fig. 1). Our analyses provide strong support for the monophyly of LT mouthbrooding cichlids (i.e. the 'H-lineage' of Nishida, 1991) as sister-group to the substrate spawning Lamprologini. We thus confirm the scenario that both lineages have radiated in parallel within LT (Salzburger and Meyer, 2004), leading to some intriguing cases of convergent evolution (Muschick et al., 2012). The clustering of the tribes within the 'H-lineage' generally reflects the life styles and habitat use of the respective tribes. The Cyprichromini and Perissodini, which are consistently put together (Figs. 2–5), are both adapted to the open-water column; the Cyphotilapiini and Limnochromini, which cluster together in most analyses (Figs. 2–5, excluding 2A), are restricted to deep-water habitats; and the Ectodini, Eretmodini and Haplochromini dominate (together with many lamprologine species) the shallow waters of LT. Our phylogenies thus reveal the general trend that the less species-rich cichlid tribes in LT (including the Bathybatini, Boulengerochromini and Trematocarini) occupy less-productive habitats such as the open-water column or deeper areas, whereas the generally more species-rich tribes of the 'H-lineage' dominate the more-productive and generally preferred shallow/rocky habitats (Muschick et al., 2012).

We further postulate a nested position of the Eretmodini within the 'H-lineage', as sister-group to the Haplochromini, which is in clear contrast to most of the studies relying on mtDNA markers (Clabaut et al., 2005; Day et al., 2008; Kocher et al., 1995; Muschick et al., 2012), yet in concordance to allozyme data (Nishida, 1991) and ncDNA phylogenies (Friedman et al., 2013). The obvious discordance between the Lamprologini-like mtDNA and Haplochromini-like ncDNA in the Eretmodini can either be explained by incomplete mtDNA lineage sorting, or, more likely, by an ancient hybridization event (Meng and Kubatko, 2009). The positions of the oldest tribes (Tylochromini, Oreochromini, Trematocarini, Bathybatini, Boulengerochromini) are largely in agreement with previous studies, as most studies suggested a sister-group relationship between the Bathybatini and Trematocarini (Clabaut et al., 2005; Day et al., 2008; Salzburger et al., 2002a) and placed the Oreochromini outside of this group (Friedman et al., 2013; Muschick et al., 2012; Salzburger et al., 2002a). The placement of the Boulengerochromini differed slightly between our analyses, but in all cases this monotypic tribe was resolved outside the clade formed by the Lamprologini and the 'H-lineage'.

5. Outlook

With this study, we provide a strong phylogenetic hypothesis for the cichlid tribes in LT based on 42 ncDNA makers. Yet, we also identified remaining areas of uncertainties, especially with respect

to the phylogenetic relationships of the mouthbrooding tribes within the 'H-lineage'. Future analyses should focus on the amount and relative proportion of shared genes among the different cichlid lineages to allow further insights into stochastic processes such as incomplete lineage sorting or hybridization. To this end, we recommend the usage of much larger datasets such as whole transcriptomes or genomes. RAD-sequencing could also provide a large random sample of ncDNA loci, although the current read lengths render the phylogenetic inference based on individual loci problematic. Another important next step to understand the evolutionary history of LT cichlids and to establish a species tree would be to perform coalescent-based analysis with BEST and *BEAST (Liu, 2008; Heled and Drummond, 2010), using phased alleles and more individuals per species. Finally, future analyses should increase taxon sampling, ultimately leading to a complete species tree for the cichlid species of LT.

Data accessibility

All sequences are accessible in Genbank KP129679-KP131427 and KM263618-KM263752.

Trees are deposited at Treebase (http://purl.org/phylo/treebase/phylows/study/TB2:S16660).

Acknowledgments

We thank Moritz Muschick, Adrian Indermaur and Frauke Münzel for help with sampling, and Ole Seehausen for providing tissues of Lake Victoria cichlids. We are grateful to Brigitte Aeschbach, Nicolas Boileau as well as Astrid Böhne and Emilía M. Santos for assistance in the laboratory, and to Christof Wunderlin and Georges Wigger from Microsynth for help with library preparation and 454 sequencing. We thank Stuart Willis and Cecile Ané for discussing the BUCKy analysis, and Sebastian Höhna for help with MrBayes. Finally, we would like to thank the editor Guillermo Orti, two anonymous reviewers, Heinz Büscher, Adrian Indermaur, and Uli Schliewen for discussion and helpful suggestions on the manuscript. This study was supported by the University of Basel, the Freiwillige Akademische Gesellschaft (FAG) Basel and the Burckhardt-Bürgin Stiftung (to BSM), the Swiss National Science Foundation (Grant PBBSP3-138680 to MM and Grants 3100A0_122458 and 3100A0_138224 to WS) and the European Research Council (ERC; Starting Grant "INTERGENADAPT" and Consolidator Grant "Cichlid∼X" to WS).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.10.009.

References

Albertson, R.C., Streelman, J.T., Kocher, T.D., 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc. Natl. Acad. Sci. USA 100, 5252–5257.
Ane, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426.
Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744.
Baum, D., 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56, 417–426.
Bayzid, M.S., Warnow, T., 2013. Naive binning improves phylogenomic analyses. Bioinformatics 29, 2277–2284.
Brawand, D. et al., 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513, 375–381.
Camargo, A., Avila, L.J., Morando, M., Sites, J.W., 2012. Accuracy and precision of species trees: effects of locus, individual, and base pair sampling on inference of species trees in lizards of the Liolaemus darwinii group (Squamata, Liolaemidae). Syst. Biol. 61, 272–288.
Chen, F.-C., Li, W.-H., 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Human Genet. 68, 444–456.
Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. Mol. Ecol. 7, 1255–1256.
Chung, Y., Ané, C., 2011. Comparing two bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Syst. Biol. 60, 261–275.
Clabaut, C., Salzburger, W., Meyer, A., 2005. Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. J. Mol. Evol. 61, 666–681.
Cohen, A.S., Lezzar, K.E., Tiercelin, J.J., Soreghan, M., 1997. New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. Basin Res. 9, 107–132.
Day, J.J., Cotton, J.A., Barraclough, T.G., 2008. Tempo and mode of diversification of Lake Tanganyika cichlid fishes. PLoS ONE 3, e1730.
Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.
Duftner, N., Koblmuller, S., Sturmbauer, C., 2005. Evolutionary relationships of the Limnochromini, a tribe of benthic deepwater cichlid fish endemic to Lake Tanganyika, East Africa. J. Mol. Evol. 60, 277–289.
Dunz, A.R., Schliewen, U.K., 2013. Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as "*Tilapia*". Mol. Phylogenet. Evol. 68, 64–80.
Friedman, M., Keck, B.P., Dornburg, A., Eytan, R.I., Martin, C.H., Hulsey, C.D., Wainwright, P.C., Near, T.J., 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc. R. Soc. Lond. B 280, 20131733.
Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. B. Mol. Dev. Evol. 304, 64–74.
Gavrilets, S., Vose, A., 2005. Dynamic patterns of adaptive radiation. Proc. Natl. Acad. Sci. USA 102, 18040–18045.
Genner, M.J., Seehausen, O., Lunt, D.H., Joyce, D.A., Shaw, P.W., Carvalho, G.R., Turner, G.F., 2007. Age of cichlids: new dates for ancient lake fish radiations. Mol. Biol. Evol. 24, 1269–1282.
Glor, R.E., 2010. Phylogenetic insights on adaptive radiation. Ann. Rev. Ecol. Evol. Syst. 41, 251–270.
Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Joyce, D.A., Lunt, D.H., Genner, M.J., Turner, G.F., Bills, R., Seehausen, O., 2011. Repeated colonization and hybridization in Lake Malawi cichlids. Curr. Biol. 21, R108–109.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kirchberger, P.C., Sefc, K.M., Sturmbauer, C., Koblmüller, S., 2014. Outgroup effects on root position and tree topology in the AFLP phylogeny of a rapidly radiating lineage of cichlid fish. Mol. Phylogenet. Evol. 70, 57–62.
Klett, V., Meyer, A., 2002. What, if anything, is a Tilapia? – mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. Mol. Biol. Evol. 19, 865–883.
Knowles, L.L., Kubatko, L.S., 2011. Estimating Speciest: Practical and Theoretical Aspects. Wiley.
Koblmüller, S., Duftner, N., Katongo, C., Phiri, H., Sturmbauer, C., 2005. Ancient divergence in bathypelagic Lake Tanganyika deepwater cichlids: mitochondrial phylogeny of the tribe Bathybatini. J. Mol. Evol. 60, 297–314.
Koblmüller, S., Duftner, N., Sefc, K.M., Aibara, M., Stipacek, M., Blanc, M., Egger, B., Sturmbauer, C., 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika – the result of repeated introgressive hybridization. BMC Evol. Biol. 7, 7.
Koblmüller, S., Schliewen, U.K., Duftner, N., Sefc, K.M., Katongo, C., Sturmbauer, C., 2008a. Age and spread of the haplochromine cichlid fishes in Africa. Mol. Phylogenet. Evol. 49, 153–169.
Koblmüller, S., Sefc, K.M., Sturmbauer, C., 2008b. The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. Hydrobiol. 615, 5–20.
Koch, M., Koblmüller, S., Sefc, K., 2007. Evolutionary history of the endemic Lake Tanganyika cichlid fish *Tylochromis polylepis*: a recent intruder to a mature adaptive radiation. J. Zool. Syst. Evol. Res. 45, 64–71.
Kocher, T.D., 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nature Rev. Genet. 5, 288–298.
Kocher, T.D., Conroy, J.A., McKaye, K.R., Stauffer, J.R., Lockwood, S.F., 1995. Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. Mol. Phylogenet. Evol. 4, 420–432.
Kubatko, L., Degnan, J., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.
Kumar, S., Skjaeveland, A., Orr, R.J., Enger, P., Ruden, T., Mevik, B.H., Burki, F., Botnen, A., Shalchian-Tabrizi, K., 2009. AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. BMC Bioinf. 10, 357.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Lang, M., Miyake, T., Braasch, I., Tinnemore, D., Siegel, N., Salzburger, W., Amemiya, C.T., Meyer, A., 2006. A BAC library of the East African haplochromine cichlid fish *Astatotilapia burtoni*. J. Exp. Zool. B. Mol. Dev. Evol. 306, 35–44.
Larget, B.R., Kotha, S.K., Dewey, C.N., Ane, C., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26, 2910–2911.
Leigh, J.W., Susko, E., Baumgartner, M., Roger, A.J., 2008. Testing congruence in phylogenomic analysis. Syst. Biol. 57, 104–115.
Lessa, E.P., 1992. Rapid surveying of DNA sequence variation in natural populations. Mol. Biol. Evol. 9, 323.
Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50, 913–925.
Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.
Li, C., Ortí, G., Zhang, G., Lu, G., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7, 44.
Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543.
Maddison, W., 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5, 365–377.
McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–538.
Meng, C., Kubatko, L.S., 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor. Popul. Biol. 75, 35–45.
Meyer, B.S., Salzburger, W., 2012. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. Mol. Ecol. Res. 12, 1097–1104.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Comput. Environ. Workshop (GCE) 2010, 1–8.
Müller, K., 2005. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. Appl. Bioinf. 4, 65–69.
Muschick, M., Indermaur, A., Salzburger, W., 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22, 2362–2368.
Nishida, M., 1991. Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes: inferences from allozyme data. Experientia 47, 974–979.
Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24, 581–583.
Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.
Poll, M., 1986. Classification des Cichlidae du lac Tanganika. Tribus, genres et especes. Academie Royale de Belgique. Classe des Sciences. Memoires 45, 1–163.
Rambaut, A., Drummond, A.J., 2007. Tracer v1.5 <http://beast.bio.ed.ac.uk/Tracer>.
Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331.
Salzburger, W., 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 18, 169–185.
Salzburger, W., Meyer, A., 2004. The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften 91, 277–290.
Salzburger, W., Meyer, A., Baric, S., Verheyen, E., Sturmbauer, C., 2002a. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. Syst. Biol. 51, 113–135.
Salzburger, W., Baric, S., Sturmbauer, C., 2002b. Speciation via introgressive hybridization in East African cichlids? Mol. Ecol. 11, 619–625.
Salzburger, W., Mack, T., Verheyen, E., Meyer, A., 2005. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 5, 17.
Salzburger, W., Van Bocxlaer, B., Cohen, A.S., 2014. Ecology and evolution of the African Great Lakes and their faunas. Annu. Rev. Ecol. Evol. Syst. 45, 519–545.
Santos, M.E., Salzburger, W., 2012. Evolution. How cichlids diversify. Science 338, 619–621.
Santos, M.E., Braasch, I., Boileau, N., Meyer, B.S., Sauteur, L., Böhne, A., Belting, H.-G., Affolter, M., Salzburger, W., 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5. http://dx.doi.org/10.1038/ncomms6149.
Schelly, R.C., Stiassny, M.L.J., 2004. Revision of the Congo River *Lamprologus* Schilthuis, 1891 (Teleostei: Cichlidae), with descriptions of two new species. Am. Mus. Nov. 3451, 1–40.
Schelly, R., Stiassny, M.L., Seegers, L., 2003. *Neolamprologus devosi* sp. n., a new riverine lamprologine cichlid (Teleostei, Cichlidae) from the lower Malagarasi River. Tanzania. Zootaxa 373, 1–11.
Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Schwarzer, J., Swartz, E.R., Vreven, E., Snoeks, J., Cotterill, F.P.D., Misof, B., Schliewen, U.K., 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proc. R. Soc. Lond. B 279, 4389–4398.
Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207.
Seehausen, O., 2006. African cichlid fish: a model system in adaptive radiation research. Proc. R. Soc. Lond. B 273, 1987–1998.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.
Slade, R.W., Moritz, C., Heideman, A., Hale, P.T., 1993. Rapid assessment of single-copy nuclear DNA variation in diverse species. Mol. Ecol. 2, 359–373.
Slowinski, J.B., 2001. Molecular polytomies. Mol. Phylogenet. Evol. 19, 114–120.
Snoeks, J., 2000. How well known is the ichthyodiversity of the large East African lakes? In: Rossiter, A., Kawanabe, H. (Eds.), Advances in Ecological Research, 0065-2504, 31. Academic Press, ISBN 9780120139316, pp. 17–38. http://dx.doi.org/10.1016/S0065-2504(00)31005-4.
Soria-Carrasco, V., Talavera, G., Igea, J., Castresana, J., 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics 23, 2954–2956.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stiassny, M., 1990. Tylochromis, relationships and the phylogenetic status of the African Cichlidae. Am. Mus. Nov. 2993, 1–14.
Sturmbauer, C., Meyer, A., 1993. Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika in eastern Africa. Mol. Biol. Evol. 10, 751–768.
Sturmbauer, C., Hainz, U., Baric, S., Verheyen, E., Salzburger, W., 2003. Evolution of the tribe Tropheini from Lake Tanganyika: synchronized explosive speciation producing multiple evolutionary parallelism. Hydrobiologia 500, 51–64.
Sturmbauer, C., Salzburger, W., Duftner, N., Schelly, R., Koblmuller, S., 2010. Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. Mol. Phylogenet. Evol. 57, 266–284.
Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
Takahashi, T., 2003. Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). Ichthyol. Res. 50, 367–382.
Takahashi, K., Terai, Y., Nishida, M., Okada, N., 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol. Biol. Evol. 18, 2057–2066.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Turner, G.F., Seehausen, O., Knight, M.E., Allender, C.J., Robinson, R.L., 2001. How many species of cichlid fishes are there in African lakes? Mol. Ecol. 10, 793–806.
Verheyen, E., Salzburger, W., Snoeks, J., Meyer, A., 2003. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science.
Wagner, C.E., Harmon, L.J., Seehausen, O., 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487, 366–369.
Walsh, H.E., Kidd, M.G., Moum, T., Friesen, V.L., 1999. Polytomies and the power of phylogenetic inference. Evolution 53, 932–937.
Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22, 258–265.
Willis, S.C., Farias, I.P., Orti, G., 2013. Multi-locus species tree for the Amazonian peacock basses (Cichlidae: *Cichla*): Emergent phylogenetic signal despite limited nuclear variation. Mol. Phylogenet. Evol. 69, 479–490.
Won, Y.-J., Sivasundar, A., Wang, Y., Hey, J., 2005. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. Proc. Natl. Acad. Sci. USA 102 (Suppl. 1), 6581–6586.
Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nature Rev. Genet., 303–314.
Yang, J., Warnow, T., 2011. Fast and accurate methods for phylogenomic analyses. BMC Bioinf. 12 (Suppl 9), S4.
Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion. PhD Thesis. The University of Texas at Austin.

Supplementary Material

A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach

Britta S. Meyer, Michael Matschiner & Walter Salzburger

Molecular Phylogenetics and Evolution (2015) 83, 56-71,
http://dx.doi.org/10.1016/j.ympev.2014.10.009

translate

1 02Calmac,	Callochromis macrops
2 06Cyplep,	Cyprichromis leptosoma
3 09Neotet,	Neolamprologus tetracanthus
4 13Boumic,	Boulengerochromis microlepis
5 13Oretan,	Oreochromis tanganicae
6 14Trenig,	Trematocara nigrifrons
7 15Batgra,	Bathybates graueri
8 16Erecya,	Eretmodus cyanostictus
9 18Cteben,	Ctenochromis benthicola
10 22Tromoo,	Tropheus moorii
11 24Lchabe,	Limnochromis abeelei
12 24Tylpol,	Tylochromis polylepis
13 27Permic,	Perissodus microlepis
14 46Tilapia;	Tilapia sparrmanii

Population Tree:
(((((((1,(8,10)),(9,11)),(2,13)),3),4),(6,7)),(5,12),14);

Primary Concordance Tree Topology:
((((((1,3),(8,10)),(9,11)),(2,13)),(4,(6,7))),(5,12),14);

Population Tree, With Branch Lengths In Estimated Coalescent Units:
(((((((1:10.000,(8:10.000,10:10.000):0.083):0.046,(9:10.000,11:10.000):0.055):0.033,(2:10.000,13:10.000):0.322):0.118,3:10.000):0.325,4:10.000):0.016,(6:10.000,7:10.000):0.246):0.442,(5:10.000,12:10.000):0.679,14:10.000);

Primary Concordance Tree with Sample Concordance Factors:
((((((1:1.000,3:1.000):0.059,(8:1.000,10:1.000):0.136):0.038,(9:1.000,11:1.000):0.116):0.062,(2:1.000,13:1.000):0.244):0.235,(4:1.000,(6:1.000,7:1.000):0.286):0.101):0.364,(5:1.000,12:1.000):0.572,14:1.000);

Four-way partitions in the Population Tree: sample-wide CF, coalescent units and Ties(if present)

{1; 2,3,4,5,6,7,9,11,12,13,14|8; 10}	0.386, 0.083,
{1,2,3,4,6,7,8,9,10,11,13; 14|5; 12}	0.662, 0.679,
{1,2,3,4,8,9,10,11,13; 6,7|5,12; 14}	0.571, 0.442,
{1,2,3,4,8,9,10,11,13; 5,12,14|6; 7}	0.479, 0.246,
{1,2,3,8,9,10,11,13; 4|5,12,14; 6,7}	0.344, 0.016,
{1,2,8,9,10,11,13; 3|4; 5,6,7,12,14}	0.518, 0.325,
{1,8,9,10,11; 3,4,5,6,7,12,14|2; 13}	0.517, 0.322,
{1,8,9,10,11; 2,13|3; 4,5,6,7,12,14}	0.408, 0.118,
{1,8,10; 2,3,4,5,6,7,12,13,14|9; 11}	0.369, 0.055,
{1; 8,10|2,3,4,5,6,7,12,13,14; 9,11}	0.363, 0.046,
{1,8,10; 9,11|2,13; 3,4,5,6,7,12,14}	0.355, 0.033,

Splits in the Primary Concordance Tree: sample-wide and genome-wide mean CF (95% credibility), SD of mean sample-wide CF across runs

{1,2,3,4,6,7,8,9,10,11,13,14|5,12}	0.572(0.476,0.667)	0.560(0.390,0.724)	0.006
{1,2,3,4,6,7,8,9,10,11,13|5,12,14}	0.364(0.286,0.429)	0.356(0.206,0.521)	0.005
{1,2,3,4,5,8,9,10,11,12,13,14|6,7}	0.286(0.190,0.381)	0.281(0.134,0.452)	0.007
{1,3,4,5,6,7,8,9,10,11,12,14|2,13}	0.244(0.167,0.333)	0.240(0.103,0.406)	0.007
{1,2,3,8,9,10,11,13|4,5,6,7,12,14}	0.235(0.167,0.286)	0.230(0.108,0.377)	0.003
{1,2,3,4,5,6,7,9,11,12,13,14|8,10}	0.136(0.048,0.214)	0.134(0.029,0.280)	0.009
{1,2,3,4,5,6,7,8,10,12,13,14|9,11}	0.116(0.048,0.190)	0.114(0.023,0.250)	0.01
{1,2,3,5,8,9,10,11,12,13,14|4,6,7}	0.101(0.048,0.190)	0.098(0.014,0.234)	0.012
{1,3,8,9,10,11|2,4,5,6,7,12,13,14}	0.062(0.048,0.119)	0.060(0.007,0.164)	0.012
{1,3|2,4,5,6,7,8,9,10,11,12,13,14}	0.059(0.000,0.119)	0.059(0.000,0.179)	0.01
{1,3,8,10|2,4,5,6,7,9,11,12,13,14}	0.038(0.024,0.071)	0.037(0.001,0.136)	0.001

Splits NOT in the Primary Concordance Tree but with estimated CF > 0.050:

{1,8|2,3,4,5,6,7,9,10,11,12,13,14}	0.119(0.048,0.190)	0.117(0.021,0.258)	0.002
{1,2,3,4,5,6,7,8,9,10,11,13|12,14}	0.094(0.048,0.167)	0.093(0.013,0.217)	0.008
{1,2,3,4,6,7,8,9,10,11,12,13|5,14}	0.090(0.048,0.143)	0.089(0.014,0.209)	0.008
{1,11|2,3,4,5,6,7,8,9,10,12,13,14}	0.086(0.024,0.167)	0.085(0.004,0.224)	0.008
{1,2,3,4,8,9,10,11,13|5,6,7,12,14}	0.084(0.024,0.190)	0.082(0.002,0.228)	0.019
{1,10|2,3,4,5,6,7,8,9,11,12,13,14}	0.076(0.024,0.143)	0.075(0.006,0.196)	0.002
{1,2,3,5,7,8,9,10,11,12,13,14|4,6}	0.073(0.000,0.167)	0.073(0.000,0.211)	0.003
{1,2,3,6,7,8,9,10,11,13|4,5,12,14}	0.073(0.000,0.143)	0.071(0.000,0.205)	0.002
{1,2,4,5,6,7,9,10,11,12,13,14|3,8}	0.071(0.024,0.143)	0.071(0.006,0.191)	0.009
{1,2,3,4,5,6,7,8,10,11,12,14|9,13}	0.067(0.024,0.143)	0.066(0.001,0.190)	0.003
{1,2,3,4,5,6,7,8,9,12,13,14|10,11}	0.064(0.024,0.119)	0.063(0.005,0.177)	0.004
{1,3,4,5,6,7,8,10,11,12,14|2,9,13}	0.062(0.000,0.119)	0.061(0.000,0.181)	0.01
{1,2,3,4,5,6,7,8,11,12,13,14|9,10}	0.061(0.024,0.119)	0.061(0.003,0.180)	0.002
{1,2,3,4,7,8,9,10,11,13|5,6,12,14}	0.059(0.000,0.119)	0.058(0.000,0.182)	0.002
{1,3,4,5,6,7,9,10,11,12,13,14|2,8}	0.059(0.024,0.119)	0.058(0.002,0.172)	0
{1,2,3,5,6,8,9,10,11,12,13,14|4,7}	0.058(0.000,0.143)	0.058(0.000,0.187)	0.002
{1,2,3,4,5,6,7,8,9,11,12,14|10,13}	0.055(0.000,0.119)	0.055(0.000,0.173)	0.006
{1,2,4,5,7,8,9,10,11,12,13,14|3,6}	0.054(0.000,0.119)	0.053(0.000,0.168)	0.001
{1,2,3,4,5,6,7,9,10,12,13,14|8,11}	0.052(0.000,0.119)	0.052(0.000,0.165)	0.002

Text S1: Extract of the concordance file from the Bayesian Concordance Analysis with BUCKy including more information about the concordance factors and coalescent units as well as the tree topologies of both the primary concordance tree and the population tree.

ID for Neigh-bornet	Species	Tribe (molecular classification)	Sample ID	Year	Location	Lake	Institution
1	***Bathybates graueri***	Bathybatini	12, 01G8	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
			3, 01F7	2007	Mpulungu Market	Lake Tanganyika; Northern Province	
2	***Boulengerochromis microlepis***	Boulengerochromini	15, 06B7	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
			16, 10D2	2008	Tobys	Lake Tanganyika; Northern Province	
3	***Ctenochromis benthicola***	Cyphotilapiini	16, DMC7	2011	Mpulungu Market	Lake Tanganyika; Northern Province	
4	***Cyphotilapia gibberosa***	Cyphotilapiini	6, 01I6	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
			16, X009	2008	Lawrence's Fishermen	Lake Tanganyika; Northern Province	
			12, 19I8	2008	Mpulungu Market	Lake Tanganyika; Northern Province	
5	***Cyprichromis leptosoma***	Cyprichromini	7, 17I7	2008	Tobys	Lake Tanganyika; Northern Province	Basel
			8, 17I8	2008	Tobys	Lake Tanganyika; Northern Province	
			5, 17I5	2008	Tobys	Lake Tanganyika; Northern Province	
6	***Aulonocranus dewindti***	Ectodini	16, X076	2008	Mbita Island NE	Lake Tanganyika; Northern Province	Basel
			7, 04F7	2007	Lukes	Lake Tanganyika; Northern Province	
			6, 04F6	2007	Lukes	Lake Tanganyika; Northern Province	
7	***Callochromis macrops***	Ectodini	11, 05G9	2007	Lukes	Lake Tanganyika; Northern Province	Basel
			7, 19C6	2008	Tobys	Lake Tanganyika; Northern Province	
			2, 19C1	2008	Tobys	Lake Tanganyika; Northern Province	
			9, 21C4	2008	Kasakalawe Lodge	Lake Tanganyika; Northern Province	
8	***Grammatotria lemairii***	Ectodini	7, 06F3	2007	Mpulungu market	Lake Tanganyika; Northern Province	Basel
9	***Ophthalmotilapia ventralis***	Ectodini	7, 03H9	2007	Mbita Island (2nd place)	Lake Tanganyika; Northern Province	Basel
			5, 03H7	2007	Mbita Island (2nd place)	Lake Tanganyika; Northern Province	
10	***Xenotilapia spiloptera***	Ectodini	14, X035	2008	Kasenga	Lake Tanganyika; Northern Province	Basel
11	***Eretmodus cyanostictus***	Eretmodini	15, 17B4	2008	Tobys	Lake Tanganyika; Northern Province	Basel
			8, 04B2	2007	Mbita Island (2nd place)	Lake Tanganyika; Northern Province	
			5, 02F6	2007	Kasakalawe (Tanganyika Lodge)	Lake Tanganyika; Northern Province	
12	***Astatotilapia burtoni***	Haplochromini	7, 05C9	2007	Kalambo	Lake Tanganyika; river; Northern Province	Basel
			6, 05C8	2007	Kalambo	Lake Tanganyika; river; Northern Province	
			1, 05C3	2007	Kalambo	Lake Tanganyika; river; Northern Province	
			5, 05C7	2007	Kalambo	Lake Tanganyika; river; Northern Province	
13	***Ctenochromis horei***	Haplochromini	12, 18C1	2008	Tobys	Lake Tanganyika; Northern Province	Basel
			16, 21C6	2008	Mbita Island NW	Lake Tanganyika; Northern Province	
14	***Cynotilapia pulpican***	Haplochromini	LM_Pstpul1	2009	Local stocks	Lake Malawi	Basel
15	***exGnathochromis pfefferi***	Haplochromini	8, 18F8	2008	Tobys	Lake Tanganyika; Northern Province	Basel
16	***Haplochromis cf. stappersii***	Haplochromini	8, Aststa_LU3	2011	Lufubu	Lake Tanganyika; river; Northern Province	Basel
			1, Aststa_CH1	2011	Kalambo	Lake Tanganyika; river; Northern Province	
17	***Labidochromis caeruleus***	Haplochromini	LM_Labcar1	2009	Local stocks	Lake Malawi	Basel
18	***Labrochromis 'stone'***	Haplochromini	2, LS02, 14119	2011	Makobe Island, Speke Gulf	Lake Victoria	EAWAG
			1,LS01,14259	2011	Makobe Island, Speke Gulf	Lake Victoria	
19	***Lobochilotes labiatus***	Haplochromini	2, 02B1	2007	Kasakalawe (Tanganyika Lodge)	Lake Tanganyika; Northern Province	Basel
			8, 07D4	2007	Mpulungu Market	Lake Tanganyika; Northern Province	
20	***Neochromis rufocaudalis***	Haplochromini	12_NR4,11314	2011	Makobe Island, Speke Gulf	Lake Victoria	EAWAG
21	***Paralabidochromis sp. 'rock kribensis'***	Haplochromini	14, PR2, 10623	2011	Makobe Island, Speke Gulf	Lake Victoria	EAWAG
			16, PR4, 11082	2011	Makobe Island, Speke Gulf	Lake Victoria	
22	***Pseudocrenilabrus philander***	Haplochromini	6, 01D5	2007	Kafue River	Kafue River, Western Zambia	Basel
			7, 01D6	2007	Kafue River	Kafue River, Western Zambia	
			16, 01F1	2007	Kafue River	Kafue River, Western Zambia	
23	***Pseudotropheus sp. 'acei'***	Haplochromini	LM_Pstace3	2009	Local stocks	Lake Malawi	Basel
24	***Pundamilia nyererei***	Haplochromini	7, PN03, 11303	2011	Makobe Island, Speke Gulf	Lake Victoria	EAWAG
			8, PN04, 11314	2011	Makobe Island, Speke Gulf	Lake Victoria	
25	***Rhamphochromis esox***	Haplochromini	LM_Rhaeso1	2009	Local stocks	Lake Malawi	Basel
26	***Serranochromis macrocephalus***	Haplochromini	16, 01C5	2007	Kafue River	Kafue River, Western Zambia	Basel
			8, 01B6	2007	Kafue River	Kafue River, Western Zambia	
			11, 01B9	2007	Kafue River	Kafue River, Western Zambia	
27	***Tropheus moorii***	Haplochromini	7, 06D2	2007	Mbita Island (2nd place)	Lake Tanganyika; Northern Province	Basel
			6, 06D1	2007	Mbita Island (2nd place)	Lake Tanganyika; Northern Province	
28	***Altolamprologus compressiceps***	Lamprologini	11, 07D2	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
			8, 04D4	2007	Mbita Island (2nd place)	Lake Tanganyika; Northern Province	
29	***Julidochromis ornatus***	Lamprologini	16, Julorn8	2009	Local Stock	Lake Tanganyika; Northern Province	Basel
30	***Lamprologus callipterus***	Lamprologini	10, 06I9	2007	Kasakalawe (Tanganyika Lodge)	Lake Tanganyika; Northern Province	Basel
			16, 20G2	2008	Woynze	Lake Tanganyika; Northern Province	
			1, 09G6	2008	Mbita Island W	Lake Tanganyika; Northern Province	
31	***Lepidiolamprologus elongatus***	Lamprologini	14, 21B2	2008	Kasakalawe Lodge	Lake Tanganyika; Northern Province	Basel
			7, 05I3	2007	Lukes	Lake Tanganyika; Northern Province	
			8, 06I7	2007	Kasakalawe (Tanganyika Lodge)	Lake Tanganyika; Northern Province	
32	***Neolamprologus caudopunctatus***	Lamprologini	11, 05B8	2007	Lukes	Lake Tanganyika; Northern Province	Basel
33	***Neolamprologus modestus***	Lamprologini	7, 08C3	2008	Mbita Island W	Lake Tanganyika; Northern Province	Basel
34	***Neolamprologus prochilus***	Lamprologini	8, CXA5	2011	Tobys	Lake Tanganyika; Northern Province	Basel
35	***Neolamprologus pulcher***	Lamprologini	7, 18H3	2008	Tobys	Lake Tanganyika; Northern Province	Basel
			1, 18G6	2008	Tobys	Lake Tanganyika; Northern Province	
			8, 18H4	2008	Tobys	Lake Tanganyika; Northern Province	
36	***Neolamprologus tetracanthus***	Lamprologini	16, 10I1	2008	Tobys	Lake Tanganyika; Northern Province	Basel
			11, 05A6	2007	Lukes	Lake Tanganyika; Northern Province	
			15, 10H9	2008	Tobys	Lake Tanganyika; Northern Province	
37	***Variabilichromis moorii***	Lamprologini	8, 03F8	2007	Mbita Island (1st place)	Lake Tanganyika; Northern Province	Basel
			1, 02G4	2007	Kasakalawe (Tanganyika Lodge)	Lake Tanganyika; Northern Province	
38	***Gnathochromis permaxillaris***	Limnochromini	16, X061	2008	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
39	***Limnochromis abeelei***	Limnochromini	7, 06E7	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
40	***Oreochromis tanganicae***	Oreochromini/Tilapiini	6, 20B5	2008	Nkupi Lodge Kitchen	Lake Tanganyika; Northern Province	Basel
			8, 20B7	2008	Nkupi Lodge Kitchen	Lake Tanganyika; Northern Province	
41	***Haplotaxodon microlepis***	Perissodini	7, 06A4	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
			5, 06A2	2007	Mpulungu Market	Lake Tanganyika; Northern Province	
42	***Perissodus microlepis***	Perissodini	16, X056	2008	Kasenga	Lake Tanganyika; Northern Province	Basel
			14, 21E5	2008	Mbita Island NW	Lake Tanganyika; Northern Province	
43	***Tilapia sparrmanii***	Tilapiini	01D4	2007	Kafue River	Kafue River, Western Zambia	Basel
44	***Trematocara nigrifrons***	Trematocarini	13, 21I1	2008	Mpulungu Market	Lake Tanganyika; Northern Province	Basel
			1, 21G7	2008	Mpulungu Market	Lake Tanganyika; Northern Province	
45	***Tylochromis polylepis***	Tylochromini	16, 07D1	2007	Mpulungu Market	Lake Tanganyika; Northern Province	Basel

Table S1: List of specimens used in this study. The species names, the classification into tribes, the sample ID's, sampling date and location, and institute where the samples are deposited are indicated. Numbers ("ID for neighbor-net") refer to Fig. 4.

	Astatotilapia burtoni		*Aulonocranus dewindti*		*Callochromis macrops*		*Grammatotria lemairii*		*Ophthalmotilapia ventralis*		*Xenotilapia spiloptera*		*Cyprichromis leptosoma*		*Neolamprologus caudopunctatus*	
	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank
rag1	05C9	KM263618	X076	KM263619	05G9	KM263620	06F3	KP131305	03H9	KM263621	X035	KP131306	17I7	KP131307	05B8	KP131308
b2m	05C9	KP129955	X076	KP129956	05G9	KP129957	06F3	KP129958	03H9	KP129959	X03shortene	KP129960	17I7	KP129961	05B8	KP129962
gapdhs	05C9	KM263633	X076	KM263634	05G9	KM263635	06F3	KP130222	03H9	KM263636	X035	KP130223	17I7	KP130224	05B8	KP130225
Ptchd4	05C9	KP131338	X076	KP131339	05G9	KP131340	06F3	KP131341	03H9	KP131342	X035	KP131343	17I7	KP131344	05B8	KP131345
enc1	05C9	KP131182	X076	KP131183	05G9	KP131184	06F3	KP131185	03H9	KP131186	X035	KP131187	17I7	KP131188	05B8	KP131189
phpt1	05C9	KP130388	X076	KP130389	05G9	KP130390	06F3	KP130391	03H9	KP130392	X035	KP130393	17I7	KP130394	05B8	KP130395
rps7	05C9	KM263648	X076	KM263649	05G9	KM263650	06F3	KP129922	03H9	KM263651	X035	KP129923	17I7	KP129924	05B8	KP129925
tbr1	05C9	KP130088	04F7	KP130089	05G9	KP130090	06F3	KP130091	03H9	KP130092	X035	KP130093	17I7	KP130094	05B8	KP130095
aqp1a.1	05C9	KP131004	X076	KP131005	05G9	KP131006	06F3	KP131007	03H9	KP131008	X035	KP131009	17I7	KP131010	05B8	KP131011
hprt1	05C9	KP130960	X076	KP130961	05G9	KP130962	06F3	KP130963	03H9	KP130964	X035	KP130965	17I7	KP130966	05B8	KP130967
anxa4	05C9	KP129998	X076	KP129999	05G9	KP130000	06F3	KP130001	03H9	KP130002	X035	KP130003	17I7	KP130004	05B8	KP130005
pgk1	05C9	KP129846	X076	KP129847	05G9	KP129848	06F3	KP129849	03H9	KP129850	X035	KP129851	17I7	KP129852	05B8	KP129853
bmp4	05C9	KM263663	X076	KM263664	05G9	KM263665	06F3	KP130888	03H9	KM263666	X035	KP130889	17I7	KP130890	05B8	KP130891
bmp2	05C9	KP129723	X076	KP129724	05G9	KP129725	06F3	KP129726	03H9	KP129727	X035	KP129728	17I7	KP129729	05B8	KP129730
TMO-4C4	05C9	KP130672	X076	KP130673	05G9	KP130674	06F3	KP130675	03H9	KP130676	X035	KP130677	17I7	KP130678	05B8	KP130679
fgf6b	05C9	KP130043	04F7	KP130044	05G9	KP130045	06F3	KP130046	03H9	KP130047	X035	KP130048	17I7	KP130049	05B8	KP130050
runx2	05C9	KP131383	X076	KP131384	05G9	KP131385	06F3	KP131386	03H9	KP131387	X035	KP131388	17I7	KP131389	05B8	KP131390
furina	05C9	KP130717	X076	KP130718	05G9	KP130719	06F3	KP130720	03H9	KP130721	X035	KP130722	17I7	KP130723	05B8	KP130724
wnt7b	05C9	KP130133	X076	KP130134	05G9	KP130135	06F3	KP130136	03H9	KP130137	X035	KP130138	17I7	KP130139	05B8	KP130140
pax9	05C8	KP130300	04F7	KP130301	05G9	KP130302	06F3	KP130303	03H9	KP130304	X035	KP130305	17I7	KP130306	05B8	KP130307
sox10b	05C9	KP129679	X076	KP129680	19C6	KP129681	06F3	KP129682	03H9	KP129683	X035	KP129684	17I7	KP129685	05B8	KP129686
otx2	05C9	KP130431	X076	KP130432	05G9	KP130433	06F3	KP130434	03H9	KP130435	X035	KP130436	17I7	KP130437	05B8	KP130438
otx1	05C9	KP131048	X076	KP131049	05G9	KP131050	06F3	KP131051	03H9	KP131052	X035	KP131053	17I7	KP131054	05B8	KP131055
dlx2a	05C9	KP130519	X076	KP130520	05G9	KP130521	06F3	KP130522	03H9	KP130523	X035	KP130524	17I7	KP130525	05B8	KP130526
dlx4b	05C9	KP131260	X076	KP131261	05G9	KP131262	06F3	KP131263	03H9	KP131264	X035	KP131265	17I7	KP131266	05B8	KP131267
barx1	05C9	KP130596	X076	KP130597	05G9	KP130598	06F3	KP130599	03H9	KP130600	X035	KP130601	17I7	KP130602	05B8	KP130603
ednrb1a	05C9	KM263678	X076	KM263679	05G9	KM263680	06F3	KP131227	03H9	KM263681	X035	KP131228	17I7	KP131229	05B8	KP131230
mc1r	05C9	KP130255	X076_low	KP130256	05G9	KP130257	06F3	KP130258	03H9	KP130259	X035	KP130260	17I7	KP130261	05B8	KP130262
skia	05C9	KP130343	X076	KP130344	05G9	KP130345	06F3	KP130346	03H9	KP130347	X035	KP130348	17I7	KP130349	05B8	KP130350
kita	05C9	KP129768	04B5	KP129769	05G9	KP129770	06F3	KP129771	03H9	KP129772	X035	KP129773	17I7	KP129774	05B8	KP129775
mitfa	05C9	KM263693	X076	KM263694	05G9	KM263695	06F3	KP129813	03H9	KM263696	X035	KP129814	17I7	KP129815	05B8	KP129816
tyr	05C9	KM263708	04F6	KM263709	05G9	KM263710	06F3	KP129891	03H9	KM263711	X035	KP129892	17I7	KP129893	05B8	KP129894
hagoromo (fbxw4)	05C9	KM263723	X076	KM263724	05G9	KM263725	06F3	KP130563	03H9	KM263726	X035	KP130564	17I7	KP130565	05B8	KP130566
slc45a2 (aim)	05C9	KP131093	X076	KP131094	19C1	KP131095	06F3	KP131096	03H9	KP131097	X035	KP131098	17I7	KP131099	05B8	KP131100
rh1	05C9	KP130800	X076	KP130801	05G9	KP130802	06F3	KP130803	03H9	KP130804	X035	KP130805	17I7	KP130806	05B8	KP130807
opn1mw (lws)	05C9	KP130476	X076_low	KP130477	05G9	KP130478	06F3	KP130479	03H9	KP130480	X035	KP130481	17I8	KP130482	05B8	KP130483
opn1sw (sws)	05C9	KP130845	04F7	KP130846	21C4	KP130847	06F3	KP130848	03H9	KP130849	X035	KP130850	17I8	KP130851	05B8	KP130852
ccng1	05D1	KP130920	0	-	05G9	KP130921	06F3	KP130922	03H9	KP130923	X035	KP130924	17I7	KP130925	05B8	KP130926
snx33	05C7	KP130762	04B5	KP130763	05G9	KP130764	06F3	KP130765	03H7	KP130766	X035	KP130767	17I5	KP130768	05B8	KP130769
rpl13a	05C8	KP130178	X076	KP130179	05G9	KP130180	06F3	KP130181	03H9	KP130182	X035	KP130183	17I7	KP130184	05B8	KP130185
edar	05C9	KP131138	X076	KP131139	05G9	KP131140	06F3	KP131141	03H9	KP131142	X035	KP131143	17I7	KP131144	05B8	KP131145
csf1ra	05C3	KM263738	X076	KM263739	05G9	KM263740	06F3	KP130639	03H9	KM263741	X035	KP130640	17I8	KP130641	05B8	KP130642

Lamprologus callipterus		*Lepidiolamprologus elongatus*		*Neolamprologus modestus*		*Neolamprologus tetracanthus*		*Neolamprologus pulcher*		*Variabilichromis moorii*		*Boulengerochromis microlepis*		*Oreochromis tanganicae*		*Trematocara nigrifrons*
Ind	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**
06I9	KP131309	21B2	KP131310	08C3	KP131311	10I1	KP131312	18H3	KP131313	03F8	KP131314	06B7	KP131315	20B5	KP131316	21I1
06I9	KP129963	21B2	KP129964	08C3	KP129965	10I1	KP129966	18G6	KP129967	03F8	KP129968	0	-	20B5	KP129969	21I1
06I9	KP130226	21B2	KP130227	08C3	KP130228	10I1	KP130229	18G6	KP130230	03F8	KP130231	06B7	KP130232	20B5	KP130233	21I1
06I9	KP131346	21B2	KP131347	08C3	KP131348	10I1	KP131349	18G6	KP131350	03F8	KP131351	06B7	KP131352	20B5	KP131353	21I1
06I9	KP131190	21B2	KP131191	08C3	KP131192	10I1	KP131193	18G6	KP131194	03F8	KP131195	06B7	KP131196	20B5	KP131197	21I1
06I9	KP130396	21B2	KP130397	08C3	KP130398	10I1	KP130399	18H3	KP130400	0	-	06B7	KP130401	20B5	KP130402	21I1
06I9	KP129926	21B2	KP129927	08C3	KP129928	10I1	KP129929	18G6	KP129930	03F8	KP129931	06B7	KP129932	20B5	KP129933	21I1
06I9	KP130096	21B2	KP130097	08C3	KP130098	10I1	KP130099	18G6	KP130100	03F8	KP130101	06B7	KP130102	20B5	KP130103	21I1
06I9	KP131012	21B2	KP131013	08C3	KP131014	10I1	KP131015	18G6	KP131016	03F8	KP131017	06B7	KP131018	20B5	KP131019	21I1
06I9	KP130968	05I3	KP130969	08C3	KP130970	10I1	KP130971	18G6	KP130972	03F8	KP130973	06B7	KP130974	20B5	KP130975	21I1
06I9	KP130006	21B2	KP130007	08C3	KP130008	10I1	KP130009	18G6	KP130010	03F8	KP130011	06B7	KP130012	20B5	KP130013	21I1
06I9	KP129854	21B2	KP129855	08C3	KP129856	10I1	KP129857	18G6	KP129858	03F8	KP129859	06B7	KP129860	20B5	KP129861	21I1
06I9	KP130892	21B2	KP130893	08C3	KP130894	10I1	KP130895	18G6	KP130896	03F8	KP130897	06B7	KP130898	20B5	KP130899	21I1
06I9	KP129731	21B2	KP129732	08C3	KP129733	10I1	KP129734	18G6	KP129735	03F8	KP129736	06B7	KP129737	20B5	KP129738	21I1
06I9	KP130680	21B2	KP130681	08C3	KP130682	10I1	KP130683	18G6	KP130684	03F8	KP130685	06B7	KP130686	20B5	KP130687	21I1
06I9	KP130051	05I3	KP130052	08C3	KP130053	05A6	KP130054	18G6	KP130055	03F8	KP130056	06B7	KP130057	20B5	KP130058	21I1
06I9	KP131391	21B2	KP131392	08C3	KP131393	10I1	KP131394	18G6	KP131395	03F8	KP131396	06B7	KP131397	20B5	KP131398	21I1
06I9	KP130725	05I3	KP130726	08C3	KP130727	10I1	KP130728	18G6	KP130729	03F8	KP130730	06B7	KP130731	20B5	KP130732	21I1
06I9	KP130141	21B2	KP130142	08C3	KP130143	10I1	KP130144	18G6	KP130145	03F8	KP130146	06B7	KP130147	20B5	KP130148	21I1
06I9	KP130308	06I7	KP130309	08C3	KP130310	07A1	KP130311	18H4	KP130312	03F8	KP130313	06B7	KP130314	20B5	KP130315	21I1
06I9	KP129687	21B2	KP129688	08C3	KP129689	10I1	KP129690	18G6	KP129691	03F8	KP129692	06B7	KP129693	20B5	KP129694	21I1
06I9	KP130439	21B2	KP130440	08C3	KP130441	10I1	KP130442	18G6	KP130443	03F8	KP130444	06B7	KP130445	20B5	KP130446	21I1
06I9	KP131056	21B2	KP131057	08C3	KP131058	10I1	KP131059	18G6	KP131060	03F8	KP131061	06B7	KP131062	20B5	KP131063	21I1
06I9	KP130527	21B2	KP130528	08C3	KP130529	10I1	KP130530	18G6	KP130531	03F8	KP130532	06B7	KP130533	20B5	KP130534	21I1
06I9	KP131268	21B2	KP131269	08C3	KP131270	10I1	KP131271	18G6	KP131272	03F8	KP131273	06B7	KP131274	20B5	KP131275	21I1
06I9	KP130604	21B2	KP130605	08C3	KP130606	10I1	KP130607	18H3	KP130608	03F8	KP130609	06B7	KP130610	20B5	KP130611	21I1
06I9	KP131231	21B2	KP131232	08C3	KP131233	10I1	KP131234	18G6	KP131235	03F8	KP131236	06B7	KP131237	20B5	KP131238	21I1
06I9	KP130263	21B2	KP130264	08C3	KP130265	10I1	KP130266	18G6	KP130267	03F8	KP130268	06B7	KP130269	20B5	KP130270	21I1
06I9	KP130351	21B2	KP130352	08C3	KP130353	10I1	KP130354	18G6	KP130355	03F8	KP130356	06B7	KP130357	20B7	KP130358	21G7
06I9	KP129776	21B2	KP129777	08C3	KP129778	10I1	KP129779	18G6	KP129780	03F8	KP129781	06B7	KP129782	20B5	KP129783	21I1
06I9	KP129817	21B2	KP129818	08C3	KP129819	10I1	KP129820	18H4	KP129821	03F8	KP129822	06B7	KP129823	20B5	KP129824	21I1
06I9	KP129895	21B2	KP129896	08C3	KP129897	10I1	KP129898	18G6	KP129899	03F8	KP129900	06B7	KP129901	20B5	KP129902	0
06I9	KP130567	05I3	KP130568	08C3	KP130569	10I1	KP130570	18G6	KP130571	03F8	KP130572	06B7	KP130573	20B5	KP130574	21I1
06I9	KP131101	21B2	KP131102	08C3	KP131103	10I1	KP131104	18G6	KP131105	03F8	KP131106	06B7	KP131107	20B5	KP131108	21I1
06I9	KP130808	21B2	KP130809	08C3	KP130810	10I1	KP130811	18G6	KP130812	03F8	KP130813	06B7	KP130814	20B5	KP130815	21I1
06I9	KP130484	21B2	KP130485	08C3	KP130486	10I1	KP130487	18G6	KP130488	03F8	KP130489	06B7	KP130490	20B5	KP130491	21I1
20G2	KP130853	X077	KP130854	08C3	KP130855	10H9	KP130856	18G6	KP130857	03F8	KP130858	06B7	KP130859	20B5	KP130860	0
06I9	KP130927	06I7	KP130928	08C3	KP130929	10I1	KP130930	Neopul7	KP130931	03F8	KP130932	10D2	KP130933	0	-	21I1
09G6	KP130770	04C6	KP130771	08C3	KP130772	0	-	18G6	KP130773	02G4	KP130774	0	-	20B5	KP130775	Mpulungu Market
06I9	KP130186	21B2	KP130187	08C3	KP130188	10I1	KP130189	18G6	KP130190	03F8	KP130191	06B7	KP130192	0	-	21I1
06I9	KP131146	21B2	KP131147	08C3	KP131148	10I1	KP131149	18G6	KP131150	03F8	KP131151	06B7	KP131152	20B5	KP131153	21I1
06I9	KP130643	21B2	KP130644	08C3	KP130645	10I1	KP130646	18G6	KP130647	03F8	KP130648	06B7	KP130649	20B5	KP130650	21I1

	Bathybates graueri		*Eretmodus cyanostictus*		*Pseudocrenilabrus philander*		*Ctenochromis benthicola*		*Neolamprologus prochilus*		*Ctenochromis horei*		*exGnathochromis pfefferi*		*Lobochilotes labiatus*	
Genbank	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**
KP131317	01G8	KP131318	17B4	KP131319	01D5	KM263622	DMC7	KP131320	CXA5	KP131321	18C1	KM263623	18F8	KP131322	02B1	KM263624
KP129970	01G8	KP129971	17B4	KP129972	01D5	KP129973	DMC7	KP129974	CXA5	KP129975	18C1	KP129976	18F8	KP129977	02B1	KP129978
KP130234	01G8	KP130235	17B4	KP130236	01D5	KM263637	DMC7	KP130237	CXA5	KP130238	18C1	KM263638	18F8	KP130239	02B1	KM263639
KP131354	01G8	KP131355	17B4	KP131356	01D5	KP131357	DMC7	KP131358	CXA5	KP131359	18C1	KP131360	18F8	KP131361	02B1	KP131362
KP131198	01G8	KP131199	17B4	KP131200	01D5	KP131201	DMC7	KP131202	CXA5	KP131203	18C1	KP131204	18F8	KP131205	02B1	KP131206
KP130403	01G8	KP130404	17B4	KP130405	01D5	KP130406	DMC7	KP130407	CXA5	KP130408	18C1	KP130409	18F8	KP130410	02B1	KP130411
KP129934	01G8	KP129935	17B4	KP129936	01D5	KM263652	DMC7	KP129937	CXA5	KP129938	18C1	KM263653	18F8	KP129939	02B1	KM263654
KP130104	01G8	KP130105	17B4	KP130106	01D5	KP130107	DMC7	KP130108	CXA5	KP130109	18C1	KP130110	18F8	KP130111	02B1	KP130112
KP131020	01G8	KP131021	17B4	KP131022	01D5	KP131023	DMC7	KP131024	CXA5	KP131025	18C1	KP131026	18F8	KP131027	02B1	KP131028
KP130976	01G8	KP130977	17B4	KP130978	01D5	KP130979	DMC7	KP130980	CXA5	KP130981	18C1	KP130982	18F8	KP130983	02B1	KP130984
KP130014	01G8	KP130015	17B4	KP130016	01F1	KP130017	DMC7	KP130018	CXA5	KP130019	18C1	KP130020	18F8	KP130021	02B1	KP130022
KP129862	01G8	KP129863	17B4	KP129864	01D5	KP129865	DMC7	KP129866	CXA5	KP129867	18C1	KP129868	18F8	KP129869	02B1	KP129870
KP130900	01G8	KP130901	17B4	KP130902	01D5	KM263674	DMC7	KP130903	CXA5	KP130904	18C1	KM263673	18F8	KP130905	02B1	KM263671
KP129739	01G8	KP129740	17B4	KP129741	01D5	KP129742	DMC7	KP129743	CXA5	KP129744	18C1	KP129745	18F8	KP129746	02B1	KP129747
KP130688	01G8	KP130689	17B4	KP130690	01D5	KP130691	DMC7	KP130692	CXA5	KP130693	18C1	KP130694	18F8	KP130695	02B1	KP130696
KP130059	01G8	KP130060	17B4	KP130061	01D5	KP130062	DMC7	KP130063	CXA5	KP130064	18C1	KP130065	18F8	KP130066	02B1	KP130067
KP131399	01G8	KP131400	17B4	KP131401	01D5	KP131402	DMC7	KP131403	CXA5	KP131404	18C1	KP131405	18F8	KP131406	02B1	KP131407
KP130733	01G8	KP130734	17B4	KP130735	01D5	KP130736	DMC7	KP130737	CXA5	KP130738	18C1	KP130739	18F8	KP130740	02B1	KP130741
KP130149	01G8	KP130150	17B4	KP130151	01D5	KP130152	DMC7	KP130153	CXA5	KP130154	18C1	KP130155	18F8	KP130156	02B1	KP130157
KP130316	01G8	KP130317	04B2	KP130318	01D5	KP130319	DMC7	KP130320	CXA5	KP130321	18C1	KP130322	18F8	KP130323	02B1	KP130324
KP129695	01G8	KP129696	17B4	KP129697	01D5	KP129698	DMC7	KP129699	CXA5	KP129700	18C1	KP129701	18F8	KP129702	02B1	KP129703
KP130447	01G8	KP130448	17B4	KP130449	01D5	KP130450	DMC7	KP130451	CXA5	KP130452	18C1	KP130453	18F8	KP130454	02B1	KP130455
KP131064	01G8	KP131065	17B4	KP131066	01D5	KP131067	DMC7	KP131068	CXA5	KP131069	18C1	KP131070	18F8	KP131071	02B1	KP131072
KP130535	01G8	KP130536	17B4	KP130537	01D5	KP130538	DMC7	KP130539	CXA5	KP130540	18C1	KP130541	18F8	KP130542	02B1	KP130543
KP131276	01G8	KP131277	17B4	KP131278	01D5	KP131279	DMC7	KP131280	CXA5	KP131281	18C1	KP131282	18F8	KP131283	07D4	KP131284
KP130612	01G8	KP130613	17B4	KP130614	01D5	KP130615	DMC7	KP130616	CXA5	KP130617	18C1	KP130618	18F8	KP130619	02B1	KP130620
KP131239	01G8	KP131240	17B4	KP131241	01D5	KM263682	DMC7	KP131242	CXA5	KP131243	18C1	KM263683	18F8	KP131244	02B1	KM263684
KP130271	01G8	KP130272	17B4	KP130273	01D5	KP130274	DMC7	KP130275	CXA5	KP130276	18C1	KP130277	18F8	KP130278	02B1	KP130279
KP130359	01G8	KP130360	17B4	KP130361	01D5	KP130362	DMC7	KP130363	CXA5	KP130364	18C1	KP130365	18F8	KP130366	02B1	KP130367
KP129784	01G8	KP129785	17B4	KP129786	01D5	KP129787	DMC7	KP129788	CXA5	KP129789	18C1	KP129790	18F8	KP129791	02B1	KP129792
KP129825	01G8	KP129826	17B4	KP129827	01D5	KM263697	DMC7	KP129828	CXA5	KP129829	18C1	KM263698	18F8	KP129830	02B1	KM263699
-	01G8	KP129903	17B4	KP129904	01D5	KM263712	DMC7	KP129905	CXA5	KP129906	18C1	KM263713	18F8	KP129907	02B1	KM263714
KP130575	01G8	KP130576	17B4	KP130577	01D5	KM263727	DMC7	KP130578	CXA5	KP130579	18C1	KM263728	18F8	KP130580	02B1	KM263729
KP131109	01G8	KP131110	17B4	KP131111	01D5	KP131112	DMC7	KP131113	CXA5	KP131114	18C1	KP131115	18F8	KP131116	02B1	KP131117
KP130816	01G8	KP130817	17B4	KP130818	01D5	KP130819	DMC7	KP130820	CXA5	KP130821	18C1	KP130822	18F8	KP130823	02B1	KP130824
KP130492	01G8	KP130493	17B4	KP130494	01D5	KP130495	DMC7	KP130496	CXA5	KP130497	18C1	KP130498	18F8	KP130499	02B1	KP130500
-	01G8	KP130861	17B4	KP130862	01D5	KP130863	DMC7	KP130864	CXA5	KP130865	18C1	KP130866	18F8	KP130867	02B1	KP130868
KP130934	01G8	KP130935	02F6	KP130936	01D6	KP130937	DMC7	KP130938	CXA5	KP130939	18C1	KP130940	18F8	KP130941	07D4	KP130942
KP130776	01F7	KP130777	02F6	KP130778	Kafue River	KP130779	DMC7	KP130780	0	-	21C6	KP130781	18F8	KP130782	02B1	KP130783
KP130193	01G8	KP130194	17B4	KP130195	01D5	KP130196	DMC7	KP130197	CXA5	KP130198	18C1	KP130199	18F8	KP130200	02B1	KP130201
KP131154	01G8	KP131155	17B4	KP131156	01D5	KP131157	DMC7	KP131158	CXA5	KP131159	18C1	KP131160	0	-	02B1	KP131161
KP130651	01G8	KP130652	17B4	KP130653	01D5	KM263742	DMC7	KP130654	CXA5	KP130655	18C1	KM263743	18F8	KP130656	02B1	KM263744

Tropheus moorii		*Cyphotilapia gibberosa*		*Limnochromis abeelei*		*Tylochromis polylepis*		*Gnathochromis permaxillaris*		*Haplotaxodon microlepis*		*Perissodus microlepis*		*Serranochromis macrocephalus*		*Altolamprolo compressice*
Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind
06D2	KP131323	X009	KP131324	6E7	KM263625	07D1	KP131325	X061	KM263626	06A4	KP131326	X056	KP131327	01C5	KP131328	07D2
06D2	KP129979	01I6	KP129980	6E7	KP129981	0	-	X061	KP129982	06A4	KP129983	X056	KP129984	01C5	KP129985	07D2
06D2	KP130240	X009	KP130241	6E7	KM263640	07D1	KP130242	X061	KM263641	06A4	KP130243	X056	KP130244	01C5	KP130245	07D2
06D2	KP131363	X009	KP131364	6E7	KP131365	07D1	KP131366	X061	KP131367	06A4	KP131368	X056	KP131369	01C5	KP131370	07D2
06D2	KP131207	X009	KP131208	6E7	KP131209	07D1	KP131210	X061	KP131211	06A4	KP131212	X056	KP131213	01C5	KP131214	07D2
06D2	KP130412	X009	KP130413	6E7	KP130414	07D1	KP130415	X061	KP130416	06A4	KP130417	X056	KP130418	01C5	KP130419	07D2
06D2	KP129940	X009	KP129941	consensus	KM263655	07D1	KP129942	X061	KM263656	06A4	KP129943	X056	KP129944	01C5	KP129945	07D2
06D2	KP130113	X009	KP130114	6E7	KP130115	07D1	KP130116	X061	KP130117	06A4	KP130118	X056	KP130119	01C5	KP130120	07D2
06D2	KP131029	X009	KP131030	6E7	KP131031	0	-	X061	KP131032	06A4	KP131033	X056	KP131034	01C5	KP131035	07D2
06D2	KP130985	X009	KP130986	6E7	KP130987	0	-	X061	KP130988	06A4	KP130989	X056	KP130990	01C5	KP130991	07D2
06D2	KP130023	X009	KP130024	6E7	KP130025	07D1	KP130026	X061	KP130027	06A4	KP130028	X056	KP130029	01C5	KP130030	07D2
06D2	KP129871	X009	KP129872	6E7	KP129873	07D1	KP129874	X061	KP129875	06A4	KP129876	X056	KP129877	01C5	KP129878	07D2
06D2	KP130906	X009	KP130907	6E7	KM263670	07D1	KP130908	X061	KM263672	06A4	KP130909	X056	KP130910	01C5	KP130911	07D2
06D2	KP129748	X009	KP129749	6E7	KP129750	07D1	KP129751	X061	KP129752	06A4	KP129753	X056	KP129754	01C5	KP129755	07D2
06D2	KP130697	X009	KP130698	6E7	KP130699	07D1	KP130700	X061	KP130701	06A4	KP130702	X056	KP130703	01C5	KP130704	07D2
06D2	KP130068	X009	KP130069	6E7	KP130070	07D1	KP130071	X061	KP130072	06A4	KP130073	X056	KP130074	01C5	KP130075	07D2
06D2	KP131408	X009	KP131409	6E7	KP131410	07D1	KP131411	X061	KP131412	06A4	KP131413	X056	KP131414	01C5	KP131415	07D2
06D2	KP130742	X009	KP130743	6E7	KP130744	07D1	KP130745	X061	KP130746	06A4	KP130747	X056	KP130748	01C5	KP130749	07D2
06D2	KP130158	X009	KP130159	6E7	KP130160	07D1	KP130161	X061	KP130162	06A4	KP130163	X056	KP130164	01C5	KP130165	07D2
09D9	KP130325	19I8	KP130326	6E7	KP130327	07D1	KP130328	X061	KP130329	06A4	KP130330	21E5	KP130331	01B6	KP130332	07D2
06D2	KP129704	0	-	6E7	KP129705	07D1	KP129706	X061	KP129707	06A4	KP129708	X056	KP129709	01B9	KP129710	07D2
06D2	KP130456	X009	KP130457	6E7	KP130458	07D1	KP130459	X061	KP130460	06A4	KP130461	X056	KP130462	01C5	KP130463	07D2
06D2	KP131073	X009	KP131074	6E7	KP131075	07D1	KP131076	X061	KP131077	06A4	KP131078	X056	KP131079	01C5	KP131080	07D2
06D2	KP130544	X009	KP130545	6E7	KP130546	0	-	X061	KP130547	06A4	KP130548	X056	KP130549	01C5	KP130550	07D2
06D2	KP131285	X009	KP131286	6E7	KP131287	07D1	KP131288	X061	KP131289	06A4	KP131290	X056	KP131291	01C5	KP131292	07D2
06D2	KP130621	X009	KP130622	6E7	KP130623	07D1	KP130624	X061	KP130625	06A4	KP130626	X056	KP130627	0	-	07D2
06D2	KP131245	X009	KP131246	6E7	KM263685	07D1	KP131247	X061	KM263686	06A4	KP131248	X056	KP131249	01C5	KP131250	07D2
06D2	KP130280	X009	KP130281	6E7	KP130282	07D1	KP130283	X061	KP130284	06A4	KP130285	X056	KP130286	01C5	KP130287	07D2
06D2	KP130368	X009	KP130369	6E7	KP130370	07D1	KP130371	X061	KP130372	06A4	KP130373	X056	KP130374	01C5	KP130375	07D2
06D2	KP129793	X009	KP129794	6E7	KP129795	07D1	KP129796	X061	KP129797	06A4	KP129798	X056	KP129799	01C5	KP129800	07D2
06D2	KP129831	X009	KP129832	6E7	KM263700	07D1	KP129833	X061	KM263701	06A2	KP129834	X056	KP129835	01C5	KP129836	04D4
06D2	KP129908	X009	KP129909	6E7	KM263715	07D1	KP129910	X061	KM263716	06A4	KP129911	X056	KP129912	01C5	KP129913	07D2
06D2	KP130581	X009	KP130582	6E7	KM263730	07D1	KP130583	X061	KM263731	06A4	KP130584	X056	KP130585	01C5	KP130586	07D2
06D2	KP131118	X009	KP131119	6E7	KP131120	07D1	KP131121	X061	KP131122	06A4	KP131123	X056	KP131124	01C5	KP131125	07D2
06D2	KP130825	X009	KP130826	6E7	KP130827	07D1	KP130828	X061	KP130829	06A4	KP130830	X056	KP130831	01C5	KP130832	07D2
06D2	KP130501	X009	KP130502	6E7	KP130503	07D1	KP130504	X061	KP130505	06A4	KP130506	X056	KP130507	01C5	KP130508	07D2
06D1	KP130869	X009	KP130870	6E7	KP130871	07D1	KP130872	X061	KP130873	06A2	KP130874	X056	KP130875	01C5	KP130876	07D2
06D2	KP130943	03E1	KP130944	06E7	KP130945	0	-	X061	KP130946	06A5	KP130947	X056	KP130948	01C5	KP130949	04D4
06D2	KP130784	20I8	KP130785	Mpulungu Market	KP130786	0	-	X061	KP130787	06A3	KP130788	X056	KP130789	01C3	KP130790	0
06D2	KP130202	X009	KP130203	6E7	KP130204	07D1	KP130205	X061	KP130206	06A4	KP130207	X056	KP130208	01C5	KP130209	07D2
06D2	KP131162	X009	KP131163	6E7	KP131164	07D1	KP131165	X061	KP131166	06A4	KP131167	X056	KP131168	01C5	KP131169	07D2
06D2	KP130657	X009	KP130658	6E7	KM263745	07D1	KP130659	X061	KM263746	06A4	KP130660	X056	KP130661	01C5	KP130662	07D2

ogus eps		*Julidochromis ornatus*		*Labrochromis 'stone'*		*Neochromis rufocaudalis*		*Pundamilia nyererei*		*Paralabidochromis sp. 'rock kribensis'*		*Haplochromis cf. stappersii*		*Labidochromis caeruleus*		*Pseudotropheus sp. 'acei'*	
Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	Ind	Genbank	
KP131329	Julorn8	KM263627	LS02	KP131330	12_NR4	KP131331	PN03	KM263628	PR2	KP131332	Aststa_LU3	KP131333	LM_Labcar1	KP131334	LM_Pstace3	KP131335	
KP129986	Julorn8	KP129987	LS02	KP129988	12_NR4	KP129989	PN04	KP129990	PR2	KP129991	Aststa_LU3	KP129992	LM_Labcar1	KP129993	LM_Pstace3	KP129994	
KP130246	Julorn8	KM263642	LS02	KP130247	12_NR4	KP130248	PN04	KM263643	PR2	KP130249	Aststa_LU3	KP130250	LM_Labcar1	KP130251	LM_Pstace3	KP130252	
KP131371	Julorn8	KP131372	LS02	KP131373	12_NR4	KP131374	PN04	KP131375	PR2	KP131376	Aststa_LU3	KP131377	LM_Labcar1	KP131378	LM_Pstace3	KP131379	
KP131215	Julorn8	KP131216	LS02	KP131217	12_NR4	KP131218	PN04	KP131219	PR2	KP131220	Aststa_LU3	KP131221	LM_Labcar1	KP131222	LM_Pstace3	KP131223	
KP130420	0	-	LS02	KP130421	12_NR4	KP130422	PN04	KP130423	PR2	KP130424	Aststa_LU3	KP130425	LM_Labcar1	KP130426	LM_Pstace3	KP130427	
KP129946	Julorn8	KM263657	LS02	KP129947	12_NR4	KP129948	PN04	KM263658	PR2	KP129949	Aststa_LU3	KP129950	LM_Labcar1	KP129951	LM_Pstace3	KP129952	
KP130121	Julorn8	KP130122	LS02	KP130123	12_NR4	KP130124	PN04	KP130125	PR2	KP130126	Aststa_LU3	KP130127	LM_Labcar1	KP130128	LM_Pstace3	KP130129	
KP131036	Julorn8	KP131037	LS02	KP131038	12_NR4	KP131039	PN04	KP131040	PR2	KP131041	Aststa_LU3	KP131042	LM_Labcar1	KP131043	LM_Pstace3	KP131044	
KP130992	Julorn8	KP130993	LS02	KP130994	12_NR4	KP130995	PN04	KP130996	PR2	KP130997	Aststa_LU3	KP130998	LM_Labcar1	KP130999	LM_Pstace3	KP131000	
KP130031	Julorn8	KP130032	LS02	KP130033	12_NR4	KP130034	PN04	KP130035	PR2	KP130036	Aststa_LU3	KP130037	LM_Labcar1	KP130038	LM_Pstace3	KP130039	
KP129879	Julorn8	KP129880	LS02	KP129881	12_NR4	KP129882	PN04	KP129883	PR2	KP129884	Aststa_LU3	KP129885	LM_Labcar1	KP129886	LM_Pstace3	KP129887	
KP130912	Julorn8	KM263669	LS02	KP130913	12_NR4	KP130914	PN04	KM263668	PR2	KP130915	Aststa_LU3	KP130916	LM_Labcar1	KP130917	LM_Pstace3	KP130918	
KP129756	Julorn8	KP129757	LS02	KP129758	12_NR4	KP129759	PN04	KP129760	PR2	KP129761	Aststa_LU3	KP129762	LM_Labcar1	KP129763	LM_Pstace3	KP129764	
KP130705	Julorn8	KP130706	LS02	KP130707	12_NR4	KP130708	PN04	KP130709	PR2	KP130710	Aststa_LU3	KP130711	LM_Labcar1	KP130712	LM_Pstace3	KP130713	
KP130076	Julorn8	KP130077	LS02	KP130078	12_NR4	KP130079	PN04	KP130080	PR2	KP130081	Aststa_LU3	KP130082	LM_Labcar1	KP130083	LM_Pstace3	KP130084	
KP131416	Julorn8	KP131417	LS02	KP131418	12_NR4	KP131419	PN04	KP131420	PR2	KP131421	Aststa_LU3	KP131421	LM_Labcar1	KP131423	LM_Pstace3	KP131424	
KP130750	Julorn8	KP130751	LS02	KP130752	12_NR4	KP130753	PN04	KP130754	PR2	KP130755	Aststa_LU3	KP130756	LM_Labcar1	KP130757	LM_Pstace3	KP130758	
KP130166	Julorn8	KP130167	LS02	KP130168	12_NR4	KP130169	PN04	KP130170	PR2	KP130171	Aststa_LU3	KP130172	LM_Labcar1	KP130173	LM_Pstace3	KP130174	
KP130333	Julorn8	KP130334	0	-	12_NR4	KP130335	PN04	KP130336	PR2	KP130337	Aststa_LU3	KP130338	LM_Labcar1	KP130339	LM_Pstace3	KP130340	
KP129711	Julorn8	KP129712	LS02	KP129713	12_NR4	KP129714	PN04	KP129715	PR2	KP129716	Aststa_LU3	KP129717	LM_Labcar1	KP129718	LM_Pstace3	KP129719	
KP130464	Julorn8	KP130465	LS02	KP130466	12_NR4	KP130467	PN04	KP130468	PR2	KP130469	Aststa_LU3	KP130470	LM_Labcar1	KP130471	LM_Pstace3	KP130472	
KP131081	Julorn8	KP131082	LS02	KP131083	12_NR4	KP131084	PN04	KP131085	PR2	KP131086	Aststa_LU3	KP131087	LM_Labcar1	KP131088	LM_Pstace3	KP131089	
KP130551	Julorn8	KP130552	LS02	KP130553	12_NR4	KP130554	PN04	KP130555	PR2	KP130556	Aststa_LU3	KP130557	LM_Labcar1	KP130558	LM_Pstace3	KP130559	
KP131293	Julorn8	KP131294	LS01	KP131295	12_NR4	KP131296	PN04	KP131297	PR2	KP131298	Aststa_LU3	KP131299	LM_Labcar1	KP131300	LM_Pstace3	KP131301	
KP130628	Julorn8	KP130629	LS02	KP130630	12_NR4	KP130631	PN04	KP130632	PR2	KP130633	Aststa_LU3	KP130634	LM_Labcar1	KP130635	LM_Pstace3	KP130636	
KP131251	Julorn8	KM263687	LS02	KP131252	12_NR4	KP131253	PN04	KM263688	PR2	KP131254	Aststa_LU3	KP131255	LM_Labcar1	KP131256	LM_Pstace3	KP131257	
KP130288	Julorn8	KP130289	LS02	KP130290	12_NR4	KP130291	PN04	KP130292	PR2	KP130293	Aststa_LU3	KP130294	LM_Labcar1	KP130295	LM_Pstace3	KP130296	
KP130376	Julorn8	KP130377	LS02	KP130378	12_NR4	KP130379	PN04	KP130380	PR2	KP130381	Aststa_LU3	KP130382	LM_Labcar1	KP130383	LM_Pstace3	KP130384	
KP129801	Julorn8	KP129802	LS02	KP129803	12_NR4	KP129804	PN04	KP129805	PR2	KP129806	Aststa_LU3	KP129807	LM_Labcar1	KP129808	LM_Pstace3	KP129809	
KP129837	Julorn8	KM263702	LS02	KP129838	12_NR4	KP129839	PN04	KM263703	PR2	KP129840	Aststa_LU3	KP129841	LM_Labcar1	KP129842	LM_Pstace3	KP129843	
KP129914	Julorn8	KM263717	LS02	KP129915	12_NR4	KP129916	PN04	KM263718	PR4	KP129917	0	-	LM_Labcar1	KP129918	LM_Pstace3	KP129919	
KP130587	Julorn8	KM263732	LS02	KP130588	12_NR4	KP130589	PN04	KM263733	PR2	KP130590	Aststa_LU3	KP130591	LM_Labcar1	KP130592	LM_Pstace3	KP130593	
KP131126	Julorn8	KP131127	LS02	KP131128	12_NR4	KP131129	PN04	KP131130	PR2	KP131131	Aststa_LU3	KP131132	LM_Labcar1	KP131133	LM_Pstace3	KP131134	
KP130833	Julorn8	KP130834	LS02	KP130835	12_NR4	KP130836	PN04	KP130837	PR2	KP130838	Aststa_LU3	KP130839	LM_Labcar1	KP130840	LM_Pstace3	KP130841	
KP130509	0	-	LS02	KP130510	12_NR4	KP130511	PN04	KP130512	PR2	KP130513	Aststa_LU3	KP130514	LM_Labcar1	KP130515	LM_Pstace3	KP130516	
KP130877	Julorn8	KP130878	LS02	KP130879	12_NR4	KP130880	PN04	KP130881	PR2	KP130882	Aststa_LU3	KP130883	LM_Labcar1	KP130884	LM_Pstace3	KP130885	
KP130950	Julorn8	KP130951	LS4	KP130952	12_NR4	KP130953	PN04	KP130954	PR2	KP130955	Aststa_LU3	KP130956	LM_Labcar1	KP130957	LM_Pstace3	KP130958	
-	Tobys	KP130791	LS02	KP130792	12_NR4	KP130793	PN04	KP130794	PR2	KP130795	Aststa_CH1	KP130796	LM_Labcar1	KP130797	LM_Pstace3	KP130798	
KP130210	Julorn8	KP130211	LS02	KP130212	12_NR4	KP130213	PN04	KP130214	PR2	KP130215	Aststa_LU3	KP130216	LM_Labcar1	KP130217	LM_Pstace3	KP130218	
KP131170	Julorn8	KP131171	LS02	KP131172	12_NR4	KP131173	PN04	KP131174	PR2	KP131175	Aststa_LU3	KP131176	LM_Labcar1	KP131177	LM_Pstace3	KP131178	
KP130663	Julorn8	KM263747	LS02	KP130664	12_NR4	KP130665	PN04	KM263748	PR2	KP130666	Aststa_LU3	KP130667	LM_Labcar1	KP130668	LM_Pstace3	KP130669	

Cynotilapia pulpican		*Rhamphochromis esox*		*Tilapia sparrmanii*	
Ind	**Genbank**	**Ind**	**Genbank**	**Ind**	**Genbank**
LM_Pstpul1	KM263629	LM_Rhaeso1	KP131336	46Tilapia	KP131337
LM_Pstpul1	KP129995	LM_Rhaeso1	KP129996	46Tilapia	KP129997
LM_Pstpul1	KM263644	LM_Rhaeso1	KP130253	46Tilapia	KP130254
LM_Pstpul1	KP131380	LM_Rhaeso1	KP131381	46Tilapia	KP131382
LM_Pstpul1	KP131224	LM_Rhaeso1	KP131225	46Tilapia	KP131226
LM_Pstpul1	KP130428	LM_Rhaeso1	KP130429	46Tilapia	KP130430
LM_Pstpul1	KM263659	LM_Rhaeso1	KP129953	46Tilapia	KP129954
LM_Pstpul1	KP130130	LM_Rhaeso1	KP130131	46Tilapia	KP130132
LM_Pstpul1	KP131045	LM_Rhaeso1	KP131046	46Tilapia	KP131047
LM_Pstpul1	KP131001	LM_Rhaeso1	KP131002	46Tilapia	KP131003
LM_Pstpul1	KP130040	LM_Rhaeso1	KP130041	46Tilapia	KP130042
LM_Pstpul1	KP129888	LM_Rhaeso1	KP129889	46Tilapia	KP129890
LM_Pstpul1	KM263667	LM_Rhaeso1	KP130919	0	-
LM_Pstpul1	KP129765	LM_Rhaeso1	KP129766	46Tilapia	KP129767
LM_Pstpul1	KP130714	LM_Rhaeso1	KP130715	46Tilapia	KP130716
LM_Pstpul1	KP130085	LM_Rhaeso1	KP130086	46Tilapia	KP130087
LM_Pstpul1	KP131425	LM_Rhaeso1	KP131426	46Tilapia	KP131427
LM_Pstpul1	KP130759	LM_Rhaeso1	KP130760	46Tilapia	KP130761
LM_Pstpul1	KP130175	LM_Rhaeso1	KP130176	46Tilapia	KP130177
LM_Pstpul1	KP130341	0	-	46Tilapia	KP130342
LM_Pstpul1	KP129720	LM_Rhaeso1	KP129721	46Tilapia	KP129722
LM_Pstpul1	KP130473	LM_Rhaeso1	KP130474	46Tilapia	KP130475
LM_Pstpul1	KP131090	LM_Rhaeso1	KP131091	46Tilapia	KP131092
LM_Pstpul1	KP130560	LM_Rhaeso1	KP130561	46Tilapia	KP130562
LM_Pstpul1	KP131302	LM_Rhaeso1	KP131303	46Tilapia	KP131304
LM_Pstpul1	KP130637	0	-	46Tilapia	KP130638
LM_Pstpul1	KM263689	LM_Rhaeso1	KP131258	46Tilapia	KP131259
LM_Pstpul1	KP130297	LM_Rhaeso1	KP130298	46Tilapia	KP130299
LM_Pstpul1	KP130385	LM_Rhaeso1	KP130386	46Tilapia	KP130387
LM_Pstpul1	KP129810	LM_Rhaeso1	KP129811	46Tilapia	KP129812
LM_Pstpul1	KM263704	LM_Rhaeso1	KP129844	46Tilapia	KP129845
LM_Pstpul1	KM263719	LM_Rhaeso1	KP129920	46Tilapia	KP129921
LM_Pstpul1	KM263734	LM_Rhaeso1	KP130594	46Tilapia	KP130595
LM_Pstpul1	KP131135	LM_Rhaeso1	KP131136	46Tilapia	KP131137
LM_Pstpul1	KP130842	LM_Rhaeso1	KP130843	46Tilapia	KP130844
LM_Pstpul1	KP130517	0	-	46Tilapia	KP130518
LM_Pstpul1	KP130886	0	-	46Tilapia	KP130887
LM_Pstpul1	KP130959	0	-	0	-
LM_Pstpul1	KP130799	0	-	0	-
LM_Pstpul1	KP130219	LM_Rhaeso1	KP130220	46Tilapia	KP130221
LM_Pstpul1	KP131179	LM_Rhaeso1	KP131180	46Tilapia	KP131181
LM_Pstpul1	KM263749	LM_Rhaeso1	KP130670	46Tilapia	KP130671

Table S2:Sample ID's and GenBank accession numbers for all sequences used in this study.

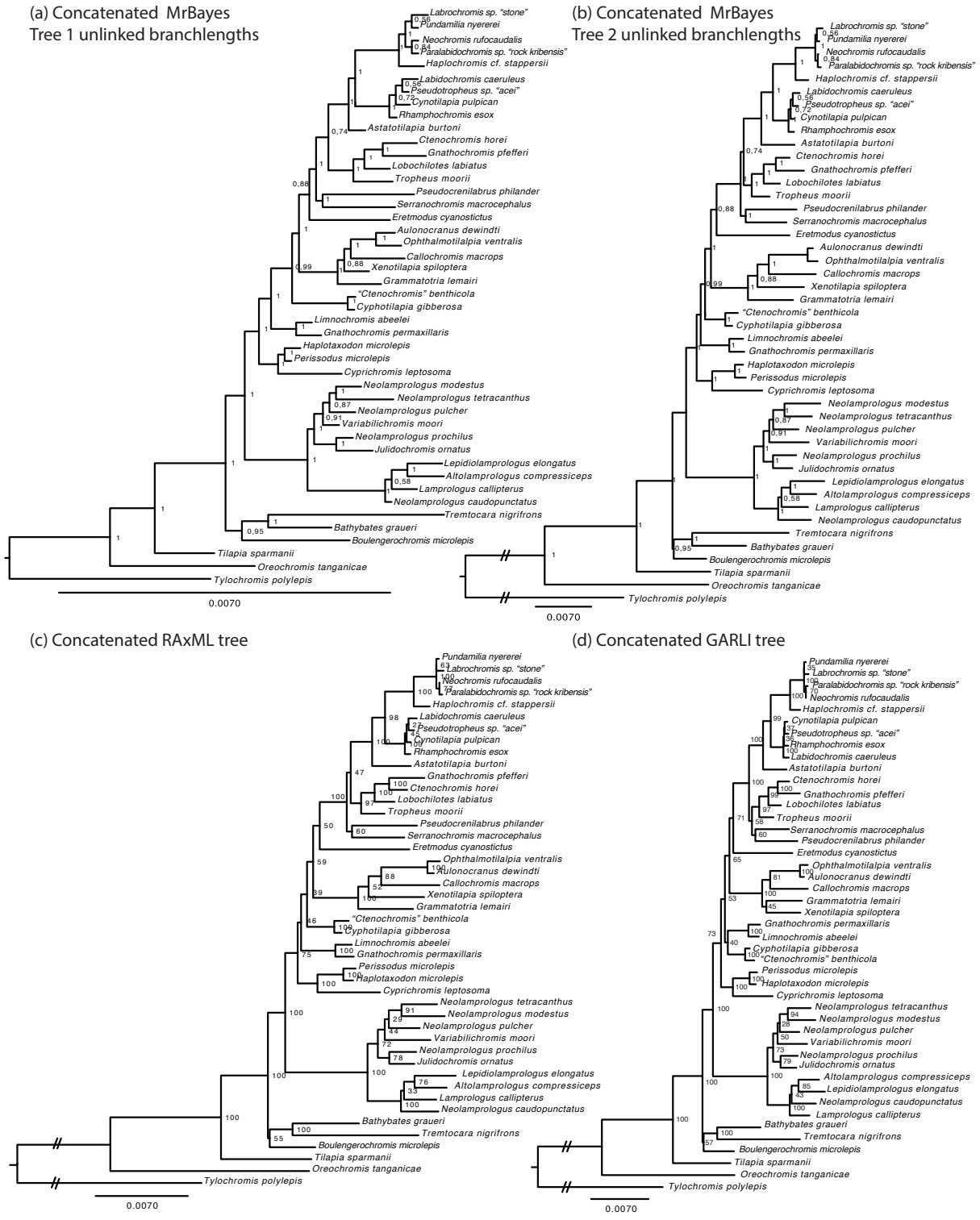


Fig. S1: Tree topologies of the concatenated supermatrix inferred with MrBayes (a and b), RAxML (c), and GARLI (d). Note that A and B result from different partitions (as suggested by PartitionFinder) and unlinking branch-lengths. All support values are plotted.

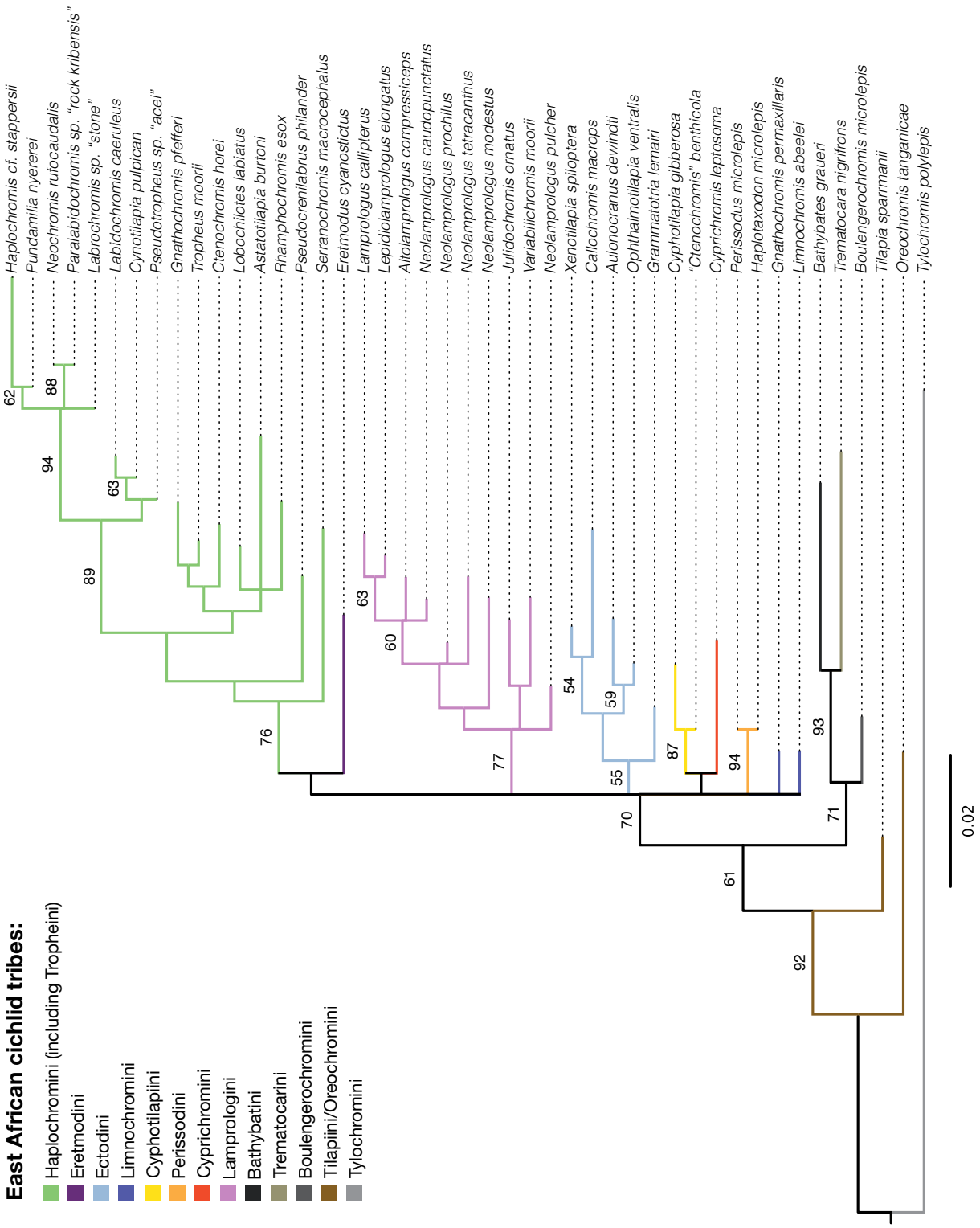


Fig. S2: Phylogram inferred using only the information provided by indels, which were transformed into a presence/absence matrix using the SIC coding procedure (Simmons and Ochoterena, 2000) and further processed with GARLI using the Mkv model. Bootstrap values (≤50%) are provided above the branches.

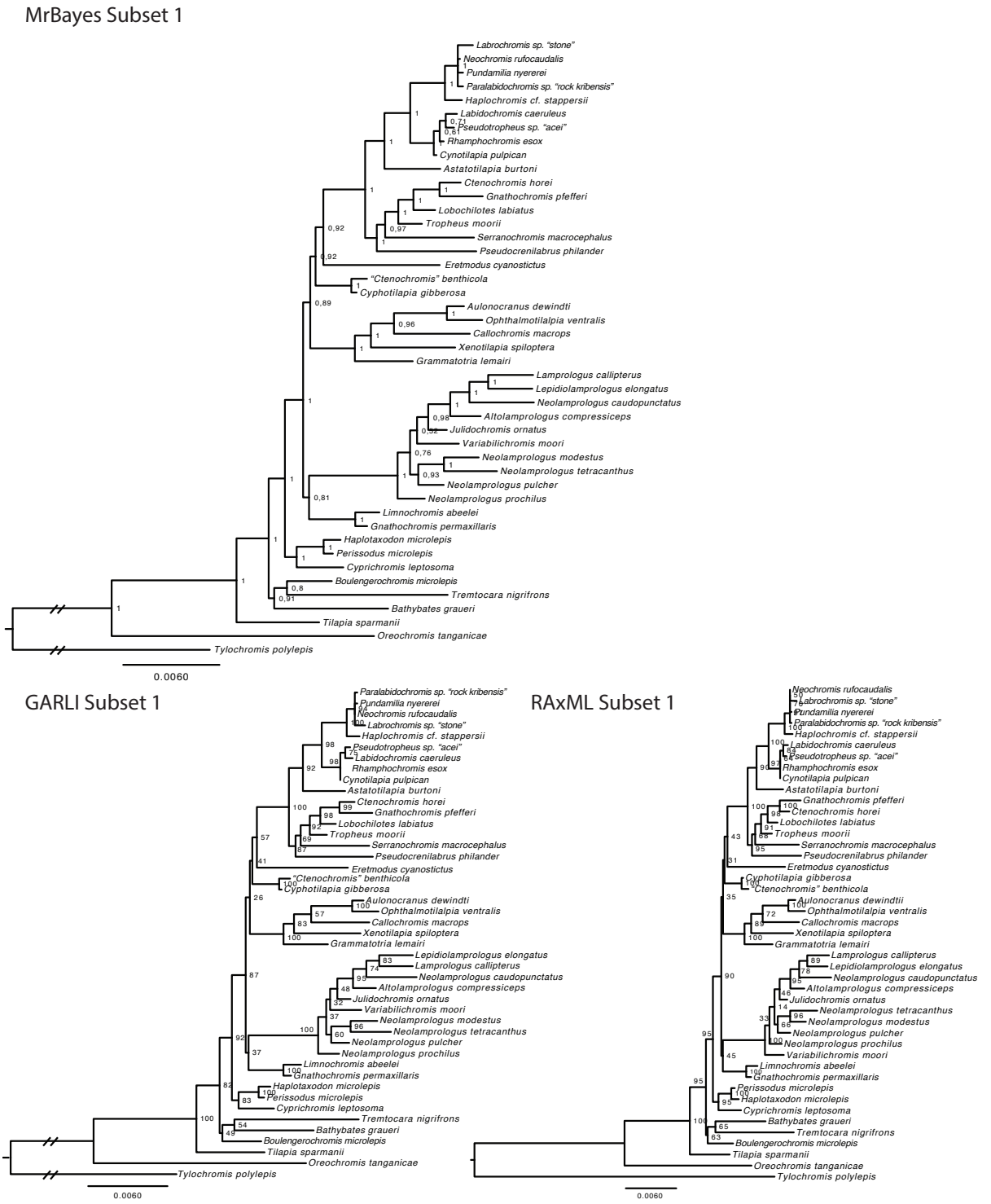


Fig. S3: Tree topologies of subset 1 from the CONCATERPILLAR analysis, inferred with GARLI, MrBayes and RAXML. All support values are plotted. The branch leading to Tylochromis polylepis was shortened by one third.

Fig. S4: Tree topologies of subset 2 from the CONCATERPILLAR analysis, inferred with GARLI, MrBayes and RAXML. All support values are plotted. The branch leading to Tylochromis polylepis was shortened by one third.

Chapter 3

Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes

Britta S. Meyer* , Adrian Indermaur *, Xenia Ehrensperger , Bernd Egger , Gaspard Banyankimbona , Jos Snoeks , Walter Salzburger

Royal Society Open Science (2015) 2,140498
http://dx.doi.org/10.1098/rsos.140498

AI, BE, GB WS were sampling specimens; GB and JS helped with taxonomic analyses; XE conducted labwork; BSM and AI equally analyzed the resulting sequence data and drafted the manuscript (*) ; BE, XE, GB, JS and WS participated in manuscript drafting; WS designed the study.

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Brief report

Cite this article: Meyer BS, Indermaur A, Ehrensperger X, Egger B, Banyankimbona G, Snoeks J, Salzburger W. 2015 Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes. *R. Soc. open sci.* **2**: 140498.
http://dx.doi.org/10.1098/rsos.140498

Received: 3 December 2014
Accepted: 9 February 2015

Subject Category:
Biology (whole organism)

Subject Areas:
evolution/taxonomy and systematics

Keywords:
Haplochromis sp. 'Chipwa', adaptive radiation, superflock, Lake Victoria

Author for correspondence:
Walter Salzburger
e-mail: walter.salzburger@unibas.ch

†Present address: Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research, Kiel 24105, Germany.
‡These authors contributed equally to this study.

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsos.140498 or via http://rsos.royalsocietypublishing.org.

Royal Society Publishing
Informing the science of the future

Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes

Britta S. Meyer[1,†,‡], Adrian Indermaur[1,‡], Xenia Ehrensperger[1], Bernd Egger[1], Gaspard Banyankimbona[2,3], Jos Snoeks[3,4] and Walter Salzburger[1]

[1]Zoological Institute, University of Basel, Basel 4051, Switzerland
[2]Department of Biology, University of Burundi, Bujumbura, Burundi
[3]Royal Museum for Central Africa, Tervuren 3080, Belgium
[4]Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Leuven 3000, Belgium

BSM, 0000-0002-2549-1825; WS 0000-0002-9988-1674

1. Summary

The species flocks of cichlid fishes in the East African Great Lakes are the largest vertebrate adaptive radiations in the world and illustrious textbook examples of convergent evolution between independent species assemblages. Although recent studies suggest some degrees of genetic exchange between riverine taxa and the lake faunas, not a single cichlid species is known from Lakes Tanganyika, Malawi and Victoria that is derived from the radiation associated with another of these lakes. Here, we report the discovery of a haplochromine cichlid species in Lake Tanganyika, which belongs genetically to the species flock of haplochromines of the Lake Victoria region. The new species colonized Lake Tanganyika only recently, suggesting that faunal exchange across watersheds and, hence, between isolated ichthyofaunas, is more common than previously thought.

2. Introduction

Adaptive radiation, the rapid evolution of novel species as a consequence of adaptation to distinct ecological niches, is thought to have played an important role in the origin of phenotypic diversity [1]. The species flocks of cichlid fishes in the African Great Lakes; Tanganyika, Malawi and Victoria are the most

species-rich vertebrate adaptive radiations, consisting of hundreds of endemic species each [2–4]. Lake Tanganyika, the oldest lake, harbours the genetically and phenotypically most diverse cichlid assemblage comprising 12–16 'tribes' [5]. The radiations in Lakes Malawi and Victoria involve only one of these tribes, the Haplochromini, making this the most species-rich cichlid lineage [4].

The haplochromines probably originated in the area of Lake Tanganyika, from where they colonized water bodies in large parts of Africa, including Lakes Malawi and Victoria [6–8]. This 'out of Tanganyika' scenario [6] implies that the seeding events of the haplochromine radiations in Lakes Malawi and Victoria date back to 1–5 and less than 0.25 Ma, respectively [6–9]. The latter radiation is not confined to only the basin of Lake Victoria, but includes the cichlid faunas of other lakes and rivers in the area, including Lakes Edward, George, Kivu and the Lake Rukwa drainage; it is hence referred to as the 'Lake Victoria region superflock' (LVRS) [6,7,10].

While Lake Tanganyika's cichlid assemblage has long been regarded as polyphyletic [11], the haplochromines from Lake Malawi and the LVRS were considered reciprocally monophyletic [7,12,13]. This view has recently been challenged with the analysis of large sets of nuclear DNA markers, which uncovered a polyphyletic origin of Lake Malawi's haplochromines [14,15], and high levels of shared genetic polymorphisms between the cichlid faunas of all three lakes [15,16]. These findings, together with the identification of similar or even identical genotypes across large geographical scales [17,18], suggest that the hydrologic systems in East Africa are more permeable for cichlids than previously thought. It has even been proposed that riverine species have 'transported' polymorphisms between lakes [15].

Interestingly, however, not a single case of a recent colonization of a Great Lake through a riverine lineage has been documented, and none of these lakes is known to contain a species belonging to a lineage associated with another Great Lake's radiation. Here we report the discovery of a haplochromine cichlid species in Lake Tanganyika, which belongs genetically to the LVRS.

3. Material and methods

In 2011 and 2012, we collected 12 specimens of a new haplochromine species (named *Haplochromis* sp. 'Chipwa' hereafter) in a shoreline habitat within Lake Tanganyika at Chipwa Village, between 500 and 1000 m south from the Kalambo River mouth. Five additional specimens were sampled in 2011 in the Lufubu River delta on Lake Tanganyika's western shoreline (open water distance between these locations: more than 55 km; figure 1*a*, *b*). In both localities, the new species co-occurs with the widespread haplochromine *Astatotilapia burtoni* found within Lake Tanganyika and in affluent rivers [20]. The new taxon was identified as undescribed species in the field by A.I.

For comparative reasons, we sampled additional haplochromines, including a morphologically similar species (*Haplochromis stappersii*) from rivers Malagarasi ($n = 4$) and Rusizi ($n = 1$) (electronic supplementary material, tables S1–S3). Sampling was performed using our standard operating procedure [21]; vouchers were deposited at the University of Basel or the Royal Museum of Central Africa, Tervuren.

In order to place the new taxon into a phylogenetic context, we amplified and sequenced two nuclear (*ednrb1*: 524 bp; *phpt1*: 434 bp) and two mitochondrial (mtDNA) loci (d-loop: 373 bp; ND2: 1047 bp), following the protocols described elsewhere [21,22]. These markers were chosen on the basis of the existence of large quantities of reference data on GenBank. The newly obtained sequences were inspected by eye in CodonCodeAligner, combined with available data from GenBank, aligned with MAFFT [23], and the appropriate models of molecular evolution were determined with jModelTest [24]. All specimens of the new species were identical in all four loci.

To identify the placement of the new species in the haplochromine phylogeny, we performed a stepwise approach using three different datasets: first, we wanted to confirm our *ad hoc* assumption that the new taxon does not belong to any of the Tanganyikan cichlid lineages (and genera) known to date. To this end, we combined the nuclear and ND2 sequences of the new species with a representative set including all East African cichlid lineages [21], resulting in a total of 83 taxa. The concatenated data (2001 bp) was analysed using Bayesian inference with MrBayes [25] (10 000 000 generations, four chains, two runs, 25% burn-in, three partitions: GTR + I + Γ; GTR + I + Γ; GTR + Γ) and maximum likelihood (ML) with GARLI (http://garli.nescent.org) (50 runs, 500 bootstrap replicates; three partitions: TIM3 + I + Γ; TVM + I + Γ; TPM2uf + Γ). In a second step, we focused on ND2 only, as many more reference data are available for this common marker in cichlids [6,8]. We again combined our data with available sequences from GenBank (216 taxa in total) and used MrBayes (3 000 000 generations, four chains, two runs, 25% burn-in; GTR + I + Γ) and GARLI (50 runs, 500 bootstraps; TIM2 + I + Γ). On the basis of this tree, we selected 86 taxa for an in-depth analysis focusing on the species belonging to the LVRS and its

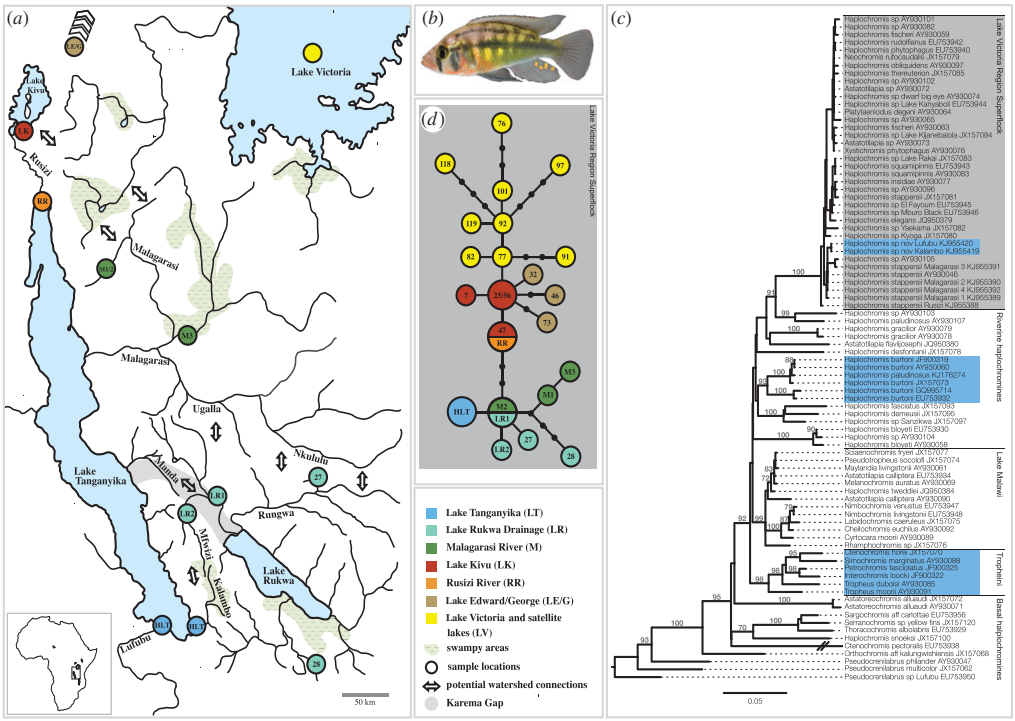


Figure 1. (*a*) Map of the study area indicating sample locations and potential watershed connections. (*b*) *Haplochromis* sp. 'Chipwa' (male) from LT. (*c*) ML phylogeny of haplochromine cichlids based on the mitochondrial ND2. *Haplochromis* sp. 'Chipwa' is firmly placed within the LVRS (grey box); the specimens from LT are depicted in blue. (*d*) Mitochondrial haplotype genealogy of representative haplotypes of the LVRS and the new species (see also the electronic supplementary material, figure S3) based on a 365 bp segment of the control region. The identification of a shared haplotype between the Malagarasi and the LR basin (M2/LR1) corroborates a recent connection between these watersheds, e.g. via 'Ugalla–Rungwa' or 'Nkululu–Rungwa' connections [19]. Colour-codes correspond to (*a*) and (*c*), haplotype numbers refer to [7].

closest sister taxa (MRBAYES: 10 000 000 generations, four chains, two runs, 25% burn-in, GTR + I + Γ; GARLI: 50 runs, 500 bootstraps, TrN + I + Γ). Finally, we integrated the mitochondrial control region sequences of *H*. sp. 'Chipwa' in the largest available dataset of members of the LVRS [7]. We performed an analysis using 178 unique mitochondrial haplotypes [7], representing about 900 specimens of the LVRS plus outgroup taxa, using GARLI (50 runs; 500 bootstraps; K81uf + I + Γ). On the basis of the resultant tree, we chose a representative subset of 27 sequences to construct a haplotype genealogy following the method described in [19] and using the first segment of the mitochondrial control region (373 bp).

4. Results

The analysis of the concatenated nuclear and mtDNA dataset resulted in highly congruent trees (electronic supplementary material, figure S1), in which *H*. sp. 'Chipwa' formed a strongly supported clade with four taxa representing the LVRS (ML bootstrap = 100, posterior probability = 1), thus confirming previous results based on a large set of nuclear DNA markers [26].

In the more inclusive ND2 phylogeny, the new species was firmly placed within the LVRS *sensu* [7] (electronic supplementary material, figure S2; ML bootstrap = 100, posterior probability = 1). Within this clade, the single ND2 haplotype of the new species from Lake Tanganyika clustered with *H. stappersii* from the Malagarasi River plus another undescribed species from Tanzania (figure 1*c*). Interestingly, two *H. stappersii* were not part of this clade: the sample from Rusizi River in Burundi and the one with unknown sampling location used by Schwartzer *et al*. [18], suggesting that specimens previously identified as *H. stappersii* are not reciprocally monophyletic and belong to at least two distinct mitochondrial lineages.

In the mtDNA haplotype genealogy, the new species was grouped into a clade of riverine taxa derived from the central haplotype of the LVRS (haplotype 25 in [7]; see the electronic supplementary material, figure S3). The reduced dataset (figure 1*d*) highlights that the single haplotype found in *H*. sp. 'Chipwa'

from Lake Tanganyika is derived from the central haplotype of this riverine clade (M2/LR1) by one mutation (nucleotide divergence: 0.29%). We refrained from performing a molecular clock analysis here, which is problematic with just one mutational difference. However, a single difference in the cichlids' mitochondrial control region is typically interpreted as recent and in the range of a maximum of tens of thousands of years [7,9].

5. Discussion

In this study, we report the discovery of a haplochromine species in Lake Tanganyika, which belongs to a clade of riverine haplochromines that is part of the LVRS (figure 1; electronic supplementary material, figures S1–S3). The phylogenetic position of the new species and the existence of identical mtDNA haplotypes on both sides of Lake Tanganyika suggest that this taxon colonized this lake recently and spread across its southern basin. Accidental translocation, e.g. with aquacultured tilapia, seems unlikely given the absence of farmed tilapia at the sampling localities. Instead, it appears likely that the new species entered Lake Tanganyika naturally.

East Africa is a geologically active area and it has been assumed that river captures mediated by tectonic movements, erosion and fluctuations in precipitation allowed for past connections between watersheds [27–30]. Since the mtDNA haplotype of the new species (HLT in figure 1) is derived from the central haplotype (M2/LR1) found in the Malagarasi and in the Lake Rukwa drainage, two alternative dispersal scenarios emerge: either via the Malagarasi River followed by southward coastal migration, or from the Lake Rukwa drainage. Given the large geographical distance between the Malagarasi River and the collection sites and that we never caught any specimen in the coastline north of the Kalambo estuary, the latter scenario appears more plausible—especially, since geological evidence suggests that Lake Rukwa was connected to Lake Tanganyika in the Early Holocene via the Karema Gap [29]. The existence of such a connection has further been corroborated with fossil molluscs and ostracods in Lake Rukwa, which resemble extant taxa from Lake Tanganyika [28]. Another recent Lake Rukwa–Lake Tanganyika connection has been hypothesized in the Kalambo-Mwimbi fault, where rivers Kalambo and Mfiwizi run, in close proximity and in opposite direction, through a swampy depression [27]. Any fish migrating downstream the Kalambo River would, however, face the challenge of a 221 m high waterfall.

With the finding of a member of the LVRS in Lake Tanganyika, we provide, to our knowledge, the first record of a cichlid species in an East African Great Lake that features genetic affinities to the fauna of another Great Lake. More precisely, we show that a haplochromine species belonging to the most recent large-scale cichlid adaptive radiation, the LVRS dated at less than 0.25 Ma [6–9], managed to migrate into the much older Lake Tanganyika, and to establish itself alongside the existing lake endemics. *Haplochromis* sp. 'Chipwa' thus represents yet another cichlid lineage that independently colonized Lake Tanganyika. Our discovery thus lends empirical support to the hypothesis that occasional migration of riverine taxa into lakes might have 'transported' genetic polymorphism between the cichlid species flocks in the East African Great Lakes [15]. Note, however, that we only demonstrated the first step required by the 'transporter hypothesis', i.e. the arrival of a distantly related haplochromine species into an established cichlid radiation. Whether this resulted in the second step, i.e. gene-flow from a divergent lineage into an established lacustrine species, remains unanswered and should be examined in the future.

Taken together, we demonstrate that recent faunal exchange occurred between the otherwise non-overlapping cichlid assemblages of the LVRS and Lake Tanganyika, thereby extending the area covered by LVRS taxa to now also include the southern part of Lake Tanganyika and affluent rivers. Our findings are in line with recent reports of shared mtDNA haplotypes across large geographical scales in haplochromines [17,18] and, particularly, with the view that faunal exchange between cichlid faunas of rivers and lakes is more common than previously thought [15]. We thus suggest that more attention should be directed towards the survey of riverine cichlid communities, which are understudied compared to the endemic faunas of Lakes Tanganyika, Malawi and Victoria.

Ethics statement. This study was performed under research permits issued by the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia and the cantonal veterinary office Basel (permit no. 2317).

Data accessibility. Sequence data has been deposited at GenBank under the accession numbers KJ955381-KJ955446.

Acknowledgements. We thank the Fisheries Department, Republic of Zambia and the Faculty of Sciences, University of Burundi, for permits; M. Colombo, F. Ronco, A. Rugirabirori and A. Theis for fieldwork assistance; T. C. Johnson for discussion and three anonymous referees for valuable suggestions.

Funding statement. This study was supported by the FAG-Basel (to B.S.M., A.I.), the Burchkardt-Bürgin-Stiftung (B.S.M.) and the Swiss National Science Foundation and European Research Council/ERC (W.S.).

Competing interests. We declare we have no competing interests.

References

1. Schluter D. 2000 *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.
2. Kocher TD. 2004 Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5**, 288–298. (doi:10.1038/nrg1316)
3. Seehausen O. 2006 African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B* **273**, 1987–1998. (doi:10.1098/rspb.2006.3539)
4. Salzburger W. 2009 The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol. Ecol.* **18**, 169–185. (doi:10.1111/j.1365-294X.2008.03981.x)
5. Koblmüller S, Sefc KM, Sturmbauer C. 2008 The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. *Hydrobiologia* **615**, 5–20. (doi:10.1007/s10750-008-9552-4)
6. Salzburger W, Mack T, Verheyen E, Meyer A. 2005 Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evol. Biol.* **5**, 17. (doi:10.1186/1471-2148-5-17)
7. Verheyen E, Salzburger W, Snoeks J, Meyer A. 2003 Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science* **300**, 325–329. (doi:10.1126/science.1080699)
8. Koblmüller S, Schliewen UK, Duftner N, Sefc KM, Katongo C, Sturmbauer C. 2008 Age and spread of the haplochromine cichlid fishes in Africa. *Mol. Phylogenet. Evol.* **49**, 153–169. (doi:10.1016/j.ympev.2008.05.045)
9. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF. 2007 Age of cichlids: new dates for ancient lake fish radiations. *Mol. Biol. Evol.* **24**, 1269–1282. (doi:10.1093/molbev/msm050)
10. Nagl S, Tichy H, Mayer WE, Takezaki N, Takahata N, Klein J. 2000 The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proc. R. Soc. Lond. B* **267**, 1049–1061. (doi:10.1098/rspb.2000.1109)
11. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. 2002 Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Syst. Biol.* **51**, 113–135. (doi:10.1080/106351502753475907)
12. Meyer A, Kocher TD, Basasibwaki P, Wilson AC. 1990 Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**, 550–553. (doi:10.1038/347550a0)
13. Moran P, Kornfield I, Reinthal PN. 1994 Molecular systematics and radiation of the haplochromine cichlids (Teleostei: Perciformes) of Lake Malawi. *Copeia* **2**, 274–288. (doi:10.2307/1446977)
14. Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O. 2011 Repeated colonization and hybridization in Lake Malawi cichlids. *Curr. Biol.* **21**, R108–R109. (doi:10.1016/j.cub.2010.11.029)
15. Loh YH *et al.* 2013 Origins of shared genetic variation in African cichlids. *Mol. Biol. Evol.* **30**, 906–917. (doi:10.1093/molbev/mss326)
16. Brawand D *et al.* 2014 The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381. (doi:10.1038/nature13726)
17. Hermann CM, Sefc KM, Koblmüller S. 2011 Ancient origin and recent divergence of a haplochromine cichlid lineage from isolated water bodies in the East African Rift System. *J. Fish Biol.* **79**, 1356–1369. (doi:10.1111/j.1095-8649.2011.03101.x)
18. Schwartzer J, Swartz ER, Vreven E, Snoeks J, Cotterill FP, Misof B, Schliewen UK. 2012 Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. *Proc. R. Soc. B* **279**, 4389–4398. (doi:10.1098/rspb.2012.1667)
19. Salzburger W, Ewing GB, Von Haeseler A. 2011 The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Mol. Ecol.* **20**, 1952–1963. (doi:10.1111/j.1365-294X.2011.05066.x)
20. Theis A, Ronco R, Indermaur A, Salzburger W, Egger B. 2014 Adaptive divergence between lake and stream populations of an East African cichlid fish. *Mol. Ecol.* **23**, 5304–5322. (doi:10.1111/mec.12939)
21. Muschick M, Indermaur A, Salzburger W. 2012 Convergent evolution within an adaptive radiation of cichlid fishes. *Curr. Biol.* **22**, 2362–2368. (doi:10.1016/j.cub.2012.10.048)
22. Meyer BS, Salzburger W. 2012 A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. *Mol. Ecol. Res.* **12**, 1097–1104. (doi:10.1111/j.1755-0998.2012.03169.x)
23. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
24. Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
25. Ronquist F *et al.* 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
26. Meyer BS, Matschiner M, Salzburger W. 2015 A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. *Mol. Phyl. Evol*. **83C**, 56–71. (doi:10.1016/j.ympev2014.10.009)
27. Seegers L. 1996 The fishes of the Lake Rukwa drainage. *Ann. Mus. R. Afr. Centr. Sci. Zool.* **287**, 1–407.
28. Cohen AS, Van Bocxlaer B, Todd JA, McGlue M, Michel E, Nkotagu HH, Grove AT, Delvaux D. 2013 Quanternary ostracods and molluscs from the Rukwa Basin (Tanzania) and their evolutionary and palaeobiogeographic implications. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **392**, 97–97. (doi:10.1016/j.palaeo.2013.09.007)
29. Delvaux D, Kervyn F, Vittori E, Kajara RSA, Kilembe E. 1998 Late Quaternary tectonic activity and lake level change in the Rukwa Rift Basin. *J. Afr. Earth Sci.* **26**, 397–421. (doi:10.1016/S0899-5362(98)00023-2)
30. Salzburger W, Van Bocxlaer B, Cohen AS. 2014 Ecology and evolution of the African Great Lakes and their faunas. *Annu. Rev. Ecol. Evol. Syst.* **45**, 519–545. (doi:10.1146/annurev-ecolsys-120213-091804)

Supplementary Material

Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes

Britta S. Meyer*, Adrian Indermaur* , Xenia Ehrensperger , Bernd Egger , Gaspard Banyankimbona , Jos Snoeks , Walter Salzburger

Royal Society Open Science (2015) 2,140498

Supplementary table 1: List of 83 cichlid specimens, their mitochondrial ND2 and their nuclear gene (ednrb, phpt1) accession numbers and their sample locations.

	GenBank accession numbers			Sampling information	
Species	**nd2**	**ednrb**	**phpt1**	**Locality**	**Coordinates**
Altolamprologus calvus	EF462256	JF900248	JF900177	Lake Tanganyika	-
Altolamprologus compressiceps	EF462257	JF900249	JF900178	Lake Tanganyika	-
Asprotilapia leptura	KJ955424	JF900251	JF900180	Lake Tanganyika	-
Haplochromis sp. nov. "Kalambo"	KJ955419	KJ955401	KJ955436	Kalambo River, Zambia	S08°36'06.34"; E031°11'12.73"
Haplochromis sp. nov. "Lufubu"	KJ955420	KJ955402	KJ955437	Lufubu River, Zambia	S08°33'41.25"; E030°43' 26.54"
Astatoreochromis alluaudi	KJ955410	KJ955393	KJ955429	Aquaria Stock, Lake Victoria	-
Astatotilapia burtoni	KJ955411	KJ955394	KJ955430	Aquaria Stock, Lake Tanganyika	-
Astatotilapia burtoni	JF900319	JF900252	JF900181	Lake Tanganyika	-
Astatotilapia calliptera	KJ955412	KJ955398	KJ955431	Aquaria Stock, Lake Malawi	-
Aulonocranus dewindti	AY337782	JF900253	JF900182	Lake Tanganyika	-
Baileychromis centropomoides	KJ955423	KJ955406	KJ955432	Mpulungu Market, Zambia	S8° 45' 56.737" E31° 6' 49.715"
Bathybates graueri	AY663726	JF900254	JF900183	Lake Tanganyika	-
Bathybates vittatus	AY663728	JF900255	JF900184	Lake Tanganyika	-
Benthochromis tricoti	AF317264	JF900256	JF900185	Lake Tanganyika	-
Boulengerochromis microlepis	AF317229	JF900257	JF900186	Lake Tanganyika	-
Callochromis macrops	AY337795	JF900258	JF900187	Lake Tanganyika	-
Chalinochromis brichardi	EF679241	JF900259	JF900188	Lake Tanganyika	-
Cyphotilapia gibberosa	EF679242	JF900260	JF900189	Lake Tanganyika	-
Ctenochromis horei	EU753935	JF900262	JF900191	Lake Tanganyika	-
Cyathopharynx furcifer	AY337781	JF900263	JF900192	Lake Tanganyika	-
Cyprichromis leptosoma	AY740337	JF900264	JF900193	Lake Tanganyika	-
Ectodus descampsii	AY337790	JF900265	JF900195	Lake Tanganyika	-
Enantiopus melanogenys	AY682517	JF900266	JF900194	Lake Tanganyika	-
Eretmodus cyanostictus	AF398220	JF900267	JF900196	Lake Tanganyika	-
Gnathochromis permaxillaris	JF900321	JF900268	JF900197	Lake Tanganyika	-
Gnathochromis pfefferi	U07248	JF900269	JF900198	Lake Tanganyika	-
Grammatotria lemairii	AY337787	JF900270	JF900199	Lake Tanganyika	-
Greenwoodochromis christyi	AY682528	JF900272	JF900201	Lake Tanganyika	-
Haplotaxodon microlepis	EF437497	JF900273	JF900202	Lake Tanganyika	-
Haplochromis obliquidens	KJ955416	KJ955403	KJ955433	Aquaria Stock, Lake Victoria	-
Haplochromis rockkribensis	KJ955418	KJ955404	KJ955434	Aquaria Stock, Lake Victoria	-
Haplotaxodon trifasciatus	EF437492	JF900274	JF900203	Lake Tanganyika	-
Interochromis loocki	JF900322	JF900303	JF900232	Lake Tanganyika	-
Julidochromis ornatus	EF462229	JF900275	JF900204	Lake Tanganyika	-
Lamprologus callipterus	AF398226	JF900276	JF900205	Lake Tanganyika	-
Lamprologus lemairii	EF462271	JF900277	JF900206	Lake Tanganyika	-
Lamprologus ornatipinnis	EF462260	JF900278	JF900207	Lake Tanganyika	-
Limnochromis abeelei	AY682533	JF900279	JF900208	Lake Tanganyika	-
Lepidiolamprologus attenuatus	EF462274	JF900282	JF900211	Lake Tanganyika	-
Lepidiolamprologus elongatus	EF462268	JF900283	JF900212	Lake Tanganyika	-
Lepidiolamprologus cf. profundicola	EF462276	JF900284	JF900213	Lake Tanganyika	-
Limnotilapia dardennii	GQ995724	JF900285	JF900214	Lake Tanganyika	-
Lobochilotes labiatus	U07254	JX402345	JF900215	Lake Tanganyika	-
Microdontochromis tenuidentatus	AY337784	JF900287	JF900216	Lake Tanganyika	-
Neolamprologus furcifer	EF679252	JF900288	JF900217	Lake Tanganyika	-
Neolamprologus modestus	DQ055012	JF900289	JF900218	Lake Tanganyika	-

Neolamprologus prochilus	EF462248	JF900290	JF900219	Lake Tanganyika	-
Neolamprologus pulcher	EF462244	JF900291	JF900220	Lake Tanganyika	-
Neolamprologus savoryi	HM623796	JF900292	JF900221	Lake Tanganyika	-
Neolamprologus sexfasciatus	HM623828	JF900293	JF900222	Lake Tanganyika	-
Neolamprologus tetracanthus	EF462220	JF900294	JF900223	Lake Tanganyika	-
Ophthalmotilapia ventralis	AY337774	JF900295	JF900224	Lake Tanganyika	-
Oreochromis tanganicae	AF317240	JF900296	JF900225	Lake Tanganyika	-
Paracyprichromis brieni	AY740378	JF900297	JF900226	Lake Tanganyika	-
Perissodus microlepis	AF398222	JF900298	JF900227	Lake Tanganyika	-
Plecodus paradoxus	EF437500	JF900299	JF900228	Lake Tanganyika	-
Petrochromis famula	JF900324	JF900301	JF900230	Lake Tanganyika	-
Petrochromis fasciolatus	JF900325	JF900302	JF900231	Lake Tanganyika	-
Petrochromis macrognathus	AY930068	JF900304	JF900233	Lake Tanganyika	-
Petrochromis polyodon	JF900326	JF900305	JF900234	Lake Tanganyika	-
Pharyngochromis acuticeps	KJ955421	KJ955396	KJ955438	Kafue, Zambia	-
Plecodus straeleni	EF437481	JF900306	JF900235	Lake Tanganyika	-
Pseudosimochromis curvifrons	GQ995777	JF900307	JF900236	Lake Tanganyika	-
Pseudotropheus sp. „acei“	KJ955413	KJ955399	KJ955439	Aquaria Stock, Lake Malawi	-
Pseudocrenilabrus multicolor	KJ955425	KJ955395	KJ955440	Aquaria Stock, Lake Malawi	-
Cynotilapia pulpican	KJ955414	KJ955400	KJ955442	Aquaria Stock, Lake Malawi	-
Pundamilia nyererei	KJ955417	KJ955405	KJ955441	Aquaria Stock, Lake Malawi	-
Reganochromis calliurus	AY682544	JF900308	JF900237	Lake Tanganyika	-
Rhamphochromis sp.	KJ955415	KJ955407	KJ955443	Aquaria Stock, Lake Malawi	-
Sarotherodon sp. "Barombi Mbo"	KJ955426	KJ955407	KJ955435	Barombi Mbo, Cameroon	-
Serranochromis macrocephalus	KJ955422	KJ955397	KJ955444	Kafue, Zambia	S14° 58' 25.315" E25° 55' 14.642"
Simochromis diagramma	AY930087	JF900310	JF900239	Lake Tanganyika	-
Telmatochromis dhonti/temporalis	EF679266	JF900311	JF900240	Lake Tanganyika	-
Oreochromis sp.	KJ955427	KJ955408	KJ955445	Kafue, Zambia	S14° 58' 25.315" E25° 55' 14.642"
Tilapia zillii	KJ955428	KJ955409	KJ955446	Daylan, Turkey	N36° 49' 56.349" E28° 38' 13.746"
Trematocara marginatum	JF900327	JF900312	JF900241	Lake Tanganyika	-
Trematochromis benthicola	JF900320	JF900261	JF900190	Lake Tanganyika	-
Trematocara nigrifrons	JF900328	JF900313	JF900242	Lake Tanganyika	-
Tropheus moorii	AY930093	JF900314	JF900243	Lake Tanganyika	-
Tylochromis polylepis	U07268	JF900315	JF900244	Lake Tanganyika	-
Variabilichromis moorii	DQ055016	JF900316	JF900245	Lake Tanganyika	-
Xenotilapia flavipinnis	AY337794	JF900317	JF900246	Lake Tanganyika	-
Xenotilapia spiloptera	AY337788	JF900318	JF900247	Lake Tanganyika	-

Supplementary table 2: List of 218 cichlid specimens and their mitochondrial ND2 sequence accession numbers. Specified are the original publications, their sample information and in which analysis there were used.

Species	Published in	Accession number	Locality	Coordinates	Collected by	Fig1c	FigS2
Haplochromis sp. "Chipwa"	Present study	KJ955419	Kalambo River Delta, Zambia	08°36'6.34"S; 031°11'12.73"E	*ZIUB NS_CH4*	+	+
Haplochromis sp. "Chipwa"	Present study	KJ955420	Lufubu River Delta, Zambia	8°33'41.25"S; 030°43' 26.54"E	*ZIUB NS_LU2*	+	+
Haplochromis stappersii "Malagarasi 1"	Present study	KJ955389	Malagarasi River	03°50'56.9"S; 030°18'01.3"E	*Gaspard Banyankimbona, MRAC1840*	+	+
Haplochromis stappersii "Malagarasi 2"	Present study	KJ955390	Malagarasi River	03°51'25.2"S; 030°17'53.5"E	*Gaspard Banyankimbona, MRAC1847*	+	+
Haplochromis stappersii "Malagarasi 3"	Present study	KJ955391	Muvumu-Nkobokobo	03°53'10.8"S; 030°15'16.1"E	*Gaspard Banyankimbona, MRAC12034*	+	+
Haplochromis stappersii "Malagarasi 4"	Present study	KJ955392	SOSUMO-Amont	03°59'33.8"S; 030°12'52.9"E	*Gaspard Banyankimbona, MRAC12087*	+	+
Haplochromis stappersii "Rusizi"	Present study	KJ955388	Gatumba marsh, Rusizi River	03°20'21,6"S; 029°13'56,9"E	*Gaspard Banyankimbona, MRAC6334*	+	+
Boulengerochromis microlepis	Klett & Meyer 2002	AF317229	n/a	n/a	*n/a*		+
Haplochromis burtoni	Kobelmüller et al. 2010	GQ995714	Kalambo, above falls	n/a	Kobelmüller et al. 2010, 7055	+	
Astatoreochromis alluaudi	Koblmüller et al. 2008	EU753923	Lake Kanyaboli, Kenya	n/a	n/a		+
Chetia brevicauda	Koblmüller et al. 2008	EU753924	Buzi River	n/a	n/a		+
Chetia brevis	Koblmüller et al. 2008	EU753925	Incomati River	n/a	n/a		+
Chetia flaviventris	Koblmüller et al. 2008	EU753926	Limpopo river	n/a	n/a		+
Chetia flaviventris	Koblmüller et al. 2008	EU753927	Limpopo river	n/a	n/a		+
Haplochromini sp. 'Lufubu'	Koblmüller et al. 2008	EU753928	Lufubu river, Zambia	n/a	n/a		+
Thoracochromis albolabris	Koblmüller et al. 2008	EU753929	Cunene River	n/a	n/a	+	+
Haplochromis bloyeti	Koblmüller et al. 2008	EU753930	Nyumba ya Mungu, Tanzania	n/a	n/a	+	+
Thoracochromis brauschi	Koblmüller et al. 2008	EU753931	Lake Fwa, DRC	n/a	n/a		+
Haplochromis burtoni	Koblmüller et al. 2008	EU753932	Kalambo River	n/a	n/a	+	+
Thoracochromis buysi	Koblmüller et al. 2008	EU753933	Cunene River	n/a	n/a		+
Astatotilapia calliptera	Koblmüller et al. 2008	EU753934	Lake Kisiba, Tanzania	n/a	n/a	+	+
Ctenochromis horei	Koblmüller et al. 2008	EU753935	Lake Tanganyika	n/a	n/a		+
Orthochromis machadoi	Koblmüller et al. 2008	EU753936	Cunene River	n/a	n/a		+
Haplochromis oligacanthus	Koblmüller et al. 2008	EU753937	Ngoko River, Congo	n/a	n/a		+
Ctenochromis pectoralis	Koblmüller et al. 2008	EU753938	Nyumba ya Mungu, Tanzania	n/a	n/a	+	+
Ctenochromis pectoralis	Koblmüller et al. 2008	EU753939	Nyumba ya Mungu, Tanzania	n/a	n/a		+
Haplochromis phytophagus	Koblmüller et al. 2008	EU753940	Lake Kenyaboli, Kenia	n/a	n/a	+	+
Haplochromis polli	Koblmüller et al. 2008	EU753941	Lower Congo River	n/a	n/a		+
Haplochromis rudolfianus	Koblmüller et al. 2008	EU753942	Lake Turkana	n/a	n/a	+	+
Haplochromis squamipinnis	Koblmüller et al. 2008	EU753943	Lake Edward Uganda	n/a	n/a	+	+
Haplochromis sp. 'Lake Kanyaboli'	Koblmüller et al. 2008	EU753944	Lake Kenyaboli, Kenia	n/a	n/a	+	+
Haplochromis sp. El Fayoum	Koblmüller et al. 2008	EU753945	El Fayoum Oasis, Egypt	n/a	n/a	+	+
Haplochromis sp. 'Mburo Black'	Koblmüller et al. 2008	EU753946	Lake Mburo, Uganda	n/a	n/a	+	+
Nimbochromis venustus	Koblmüller et al. 2008	EU753947	Lake Malawi	n/a	n/a	+	+
Nimbochromis livingstonii	Koblmüller et al. 2008	EU753948	Lake Malawi	n/a	n/a	+	+
Pharyngochromis acuticeps	Koblmüller et al. 2008	EU753949	Rundu, Namibia	n/a	n/a		+
Pseudocrenilabrus sp. Lufubu	Koblmüller et al. 2008	EU753950	Lufubu river, Zambia	n/a	n/a	+	+
Pseudocrenilabrus sp. Lunzua blue	Koblmüller et al. 2008	EU753951	Lunzua River, Zambia	n/a	n/a		+
Pseudocrenilabrus sp. Mweru orange	Koblmüller et al. 2008	EU753952	Lake Mweru	n/a	n/a		+
Pseudocrenilabrus sp. Olushandja	Koblmüller et al. 2008	EU753953	Cunene River, Olushandjia, Namibia	n/a	n/a		+
Sargochromis coulteri	Koblmüller et al. 2008	EU753954	Cunene River, Olushandjia, Namibia	n/a	n/a		+
Sargochromis coulteri	Koblmüller et al. 2008	EU753955	Olushandja, Namibia	n/a	n/a		+
Sargochromis aff. carlottae SK-2008	Koblmüller et al. 2008	EU753956	Kafue Flats, Zambia	n/a	n/a	+	+
Schwetzochromis neodon	Koblmüller et al. 2008	EU753957	Lake Fwa, Congo	n/a	n/a		+
Serranochromis angusticeps	Koblmüller et al. 2008	EU753958	Cunene River	n/a	n/a		+
Serranochromis angusticeps	Koblmüller et al. 2008	EU753959	Cunene River	n/a	n/a		+
Serranochromis stappersi	Koblmüller et al. 2008	EU753960	Lake Bangwuelu, Zambia	n/a	n/a		+
Serranochromis thumbergi	Koblmüller et al. 2008	EU753961	Lake Bangwuelu, Zambia	n/a	n/a		+
Benthochromis horii	Koblmüller et al. 2008	EU753962	Lake Tanganyika	n/a	n/a		+
Tylochromis polylepis	Kocher et al. 1995	U07268	Fish market, Uvira, Kongo	n/a	n/a		+
Haplochromis burtoni	Muschick et al. 2012	JF900319	Kalambo River, Zambia	n/a	ZIUB	+	+
Trematochromis benthicola	Muschick et al. 2012	JF900320	Lake Tanganyika	n/a	ZIUB		+
Gnathochromis permaxillaris	Muschick et al. 2012	JF900321	Lake Tanganyika	n/a	ZIUB		+
Interochromis loocki	Muschick et al. 2012	JF900322	Lake Tanganyika	n/a	ZIUB	+	+
Petrochromis ephippium	Muschick et al. 2012	JF900323	Lake Tanganyika	n/a	ZIUB		+
Petrochromis famula	Muschick et al. 2012	JF900324	Lake Tanganyika	n/a	ZIUB		+
Petrochromis fasciolatus	Muschick et al. 2012	JF900325	Lake Tanganyika	n/a	ZIUB	+	+
Petrochromis polyodon	Muschick et al. 2012	JF900326	Lake Tanganyika	n/a	ZIUB		+
Trematocara marginatum	Muschick et al. 2012	JF900327	Lake Tanganyika	n/a	ZIUB		+
Trematocara nigrifrons	Muschick et al. 2012	JF900328	Lake Tanganyika	n/a	ZIUB		+
Serranochromis macrocephalus "Cutato"	Musilová et al. 2013	KC146709	Cutato River, Angola	n/a	Musilová et al. C71		+
Serranochromis macrocephalus	Musilová et al. 2013	KC146710	Angola	n/a	Musilová et al. Z80_2		+
Serranochromis macrocephalus	Musilová et al. 2013	KC146711	Angola	n/a	Musilová et al. Z80_1		+
Serranochromis macrocephalus "Cuchi"	Musilová et al. 2013	KC146712	Cuchi River, Angola	n/a	Musilová et al. K03		+
Serranochromis macrocephalus "Cuchi"	Musilová et al. 2013	KC146713	Cuchi River, Angola	n/a	Musilová et al. K05		+
Serranochromis macrocephalus "Cuchi"	Musilová et al. 2013	KC146714	Cuchi River, Angola	n/a	Musilová et al. K07		+
Serranochromis macrocephalus "Cuito"	Musilová et al. 2013	KC146715	Cuito River, Angola	n/a	Musilová et al. K16		+
Serranochromis macrocephalus "Cuito"	Musilová et al. 2013	KC146716	Cuito River, Angola	n/a	Musilová et al. B51n		+
Serranochromis macrocephalus "Cuito"	Musilová et al. 2013	KC146717	Cuito River, Angola	n/a	Musilová et al. Z05		+
Serranochromis macrocephalus "Cuito"	Musilová et al. 2013	KC146718	Cuito River, Angola	n/a	Musilová et al. Z09		+
Serranochromis macrocephalus "Lomba"	Musilová et al. 2013	KC146719	Lomba, Angola	n/a	Musilová et al. C05n		+
Serranochromis macrocephalus "Lomba"	Musilová et al. 2013	KC146720	Lomba, Angola	n/a	Musilová et al. C11n		+
Serranochromis macrocephalus "Lomba"	Musilová et al. 2013	KC146721	Lomba, Angola	n/a	Musilová et al. C16n		+
Serranochromis macrocephalus "Lomba"	Musilová et al. 2013	KC146722	Lomba, Angola	n/a	Musilová et al. C17n		+
Serranochromis macrocephalus "Lomba"	Musilová et al. 2013	KC146723	Lomba, Angola	n/a	Musilová et al. C27n		+
Serranochromis macrocephalus "Cuemba"	Musilová et al. 2013	KC146724	Cuemba River, Angola	n/a	Musilová et al. V33		+
Serranochromis macrocephalus "Cuemba"	Musilová et al. 2013	KC146725	Cuemba River, Angola	n/a	Musilová et al. V35		+
Serranochromis macrocephalus "Cutato"	Musilová et al. 2013	KC146726	Cutato River, Angola	n/a	Musilová et al. C70		+
Haplochromis sp. Luando	Musilová et al. 2013	KC146727	Luando River, Angola	n/a	Musilová et al. Z38		+

Haplochromis sp. Luando	Musilovà et al. 2013	KC146728	Luando River, Angola	n/a	Musilovà et al. Z35		+
Haplochromis sp. Lomba	Musilovà et al. 2013	KC146729	Lomba, Angola	n/a	Musilovà et al. C52		+
Haplochromis sp. Lomba	Musilovà et al. 2013	KC146730	Lomba, Angola	n/a	Musilovà et al. C50n		+
Serranochromis sp.	Musilovà et al. 2013	KC146731	Angola	n/a	Musilovà et al. Z81		+
Thoracochromis sp. Huando	Musilovà et al. 2013	KC146732	Huando River, Angola	n/a	Musilovà et al. Z21		+
Tilapia sp.	Musilovà et al. 2013	unpublished			Z85		+
Haplochromis. stappersii	Salzburger et al. 2005	AY930046	Malagarasi River, Tanzania	n/a	L. De Vos (5-6/25/92)	+	+
Pseudocrenilabrus philander	Salzburger et al. 2005	AY930047	Zambezi River, Zambia	n/a	aquarium trade	+	+
Orthochromis uvinzae	Salzburger et al. 2005	AY930048	Malagarasi River, Tanzania	n/a	L. Seegers (TZ94-112b)		+
Orthochromis kasuluensis	Salzburger et al. 2005	AY930049	Tanzania	n/a	L. De Vos (T2-July 94)		+
Orthochromis rugufuensis	Salzburger et al. 2005	AY930050	Rugufu River, Tanzania	n/a	L. Seegers (TZ94-121)		+
Orthochromis rubrolabialis	Salzburger et al. 2005	AY930051	Tanzania	n/a	L. Seegers (TZ94-108)		+
Orthochromis luichensis	Salzburger et al. 2005	AY930052	Mkuti River, Luiche Basin, Tanzania	n/a	L. De Vos (T94/3)		+
Orthochromis mazimeroensis	Salzburger et al. 2005	AY930053	Nanganga, Burundi	n/a	L. De Vos (T1-5/27/93)		+
Orthochromis malagaraziensis	Salzburger et al. 2005	AY930054	Nyarungunga River, Burundi	n/a	L. De Vos (T5-5/28/93)		+
Orthochromis mosoensis	Salzburger et al. 2005	AY930055	Ruisseau Gytinya, Burundi	n/a	L. De Vos (T7-5/28/93)		+
Orthochromis malagaraziensis	Salzburger et al. 2005	AY930056	Nyarungunga River, Burundi	n/a	L. De Vos (7-2/19/93)		+
Orthochromis stormsi	Salzburger et al. 2005	AY930057	Kisangani (Lualaba River), DR Congo	n/a	L. De Vos (5/5/95)		+
Haplochromis bloyeti	Salzburger et al. 2005	AY930058	Lukaware River, Kenia	n/a	L. De Vos (F2A-12/93)	+	+
submitted as Ptyochromis sauvagei Haplochromis fischeri	Salzburger et al. 2005	AY930059	Lake Victoria (Kisumu, Kenya)	n/a	L. De Vos (F2B-12/93)	+	+
Haplochromis burtoni	Salzburger et al. 2005	AY930060	Lake Tanganyika	n/a	L. De Vos (31-02/6/92). T34	+	+
Maylandia livingstonii	Salzburger et al. 2005	AY930061	Lake Malawi	n/a	I. Kornfield	+	+
Haplochromis sp. 'Kisangani'	Salzburger et al. 2005	AY930062	Kisangani, (Lualaba River), DR Congo	n/a	L. De Vos (6/13/95)		+
submitted as Ptyochromis sauvagei Haplochromis fischeri	Salzburger et al. 2005	AY930063	Lake Victoria	n/a	A. Meyer, T44	+	+
Platytaeniodus degeni	Salzburger et al. 2005	AY930064	Lake Victoria	n/a	A. Meyer (Pd1)	+	+
Haplochromis sp. V7	Salzburger et al. 2005	AY930065	Lake Victoria	n/a	A. Meyer (V7-Feb 93)	+	+
Tropheus moorii	Salzburger et al. 2005	AY930066	Lake Tanganyika	n/a	E. Verheyen; T66		+
Tropheus moorii	Salzburger et al. 2005	AY930067	Lake Tanganyika	n/a	E. Verheyen; T67		+
Petrochromis macrognathus	Salzburger et al. 2005	AY930068	Lake Tanganyika	n/a	J. Snoeks, MRAC		+
Melanochromis auratus	Salzburger et al. 2005	AY930069	Lake Malawi	n/a	aquarium	+	+
Pseudocrenilabrus multicolor victoriae	Salzburger et al. 2005	AY930070	Lake Kanyaboli, Kenya	n/a	R. Abila (R082-2002)		+
Astatoreochromis alluaudi	Salzburger et al. 2005	AY930071	Lake Kanyaboli, Kenya	n/a	R. Abila (R101-2002)	+	+
Astatotilapia sp. R184	Salzburger et al. 2005	AY930072	Lake Kanyaboli, Kenya	n/a	R. Abila (R184-2002)	+	+
Astatotilapia sp. R185	Salzburger et al. 2005	AY930073	Lake Kanyaboli, Kenya	n/a	R. Abila (R185-2002)	+	+
Haplochromis sp. 'dwarf big eye'	Salzburger et al. 2005	AY930074	Lake Kanyaboli, Kenya	n/a	R. Abila (R280-2002)	+	+
Astatoreochromis alluaudi	Salzburger et al. 2005	AY930075	Lake Kanyaboli, Kenya	n/a	R. Abila (R281-2002)		+
Xystichromis phytophagus	Salzburger et al. 2005	AY930076	Lake Kanyaboli, Kenya	n/a	R. Abila (R670-2002)	+	+
Haplochromis insidiae	Salzburger et al. 2005	AY930077	Lake Kivu	n/a	E. Verheyen	+	+
Haplochromis gracilior	Salzburger et al. 2005	AY930078	Lake Kivu	n/a	E. Verheyen; K8	+	+
Thoracochromis brauschi	Salzburger et al. 2005	AY930080	Lake Fwa	n/a	R. Paul/E. Schraml (9792)		+
Serranochromis sp. 9793	Salzburger et al. 2005	AY930081	Lake Mweru-Wantipa, Zambia	n/a	T. Reuter / E. Schraml (9793)		+
Haplochromis sp. 9796	Salzburger et al. 2005	AY930082	Lake Mburo, Uganda	n/a	E. Schraml (9796)	+	+
Haplochromis squamipinnis	Salzburger et al. 2005	AY930083	Lake Edward	n/a	E. Schraml (9813)	+	+
Tropheus polli	Salzburger et al. 2005	AY930084	Lake Tanganyika	n/a	E. Verheyen		+
Tropheus duboisi	Salzburger et al. 2005	AY930085	Lake Tanganyika	n/a	E. Verheyen; M7	+	+
Tropheus brichardi	Salzburger et al. 2005	AY930086	Lake Tanganyika	n/a	E. Verheyen; M85		+
Simochromis diagramma	Salzburger et al. 2005	AY930087	Lake Tanganyika	n/a	E. Verheyen		+
Simochromis marginatus	Salzburger et al. 2005	AY930088	Lake Tanganyika	n/a	E. Verheyen	+	+
Cyrtocara moorii	Salzburger et al. 2005	AY930089	Lake Malawi	n/a	I. Kornfield	+	+
Astatotilapia calliptera	Salzburger et al. 2005	AY930090	Lake Malawi	n/a	I. Kornfield (A22)	+	+
Tropheus moorii	Salzburger et al. 2005	AY930091	Lake Tanganyika	n/a	E. Verheyen; 97	+	+
Cheilochromis euchilus	Salzburger et al. 2005	AY930092	Lake Malawi	n/a	I. Kornfield	+	+
Tropheus moorii	Salzburger et al. 2005	AY930093	Lake Tanganyika	n/a	E. Verheyen, 116		+
Pharyngochromis acuticeps	Salzburger et al. 2005	AY930094	Zambezi River, Zambia	n/a	C. Katongo / C. Sturmbauer		+
Thoracochromis brauschi	Salzburger et al. 2005	AY930095	Lake Fwa , Congo	n/a	Aquarium trade		+
Haplochromis sp. T13	Salzburger et al. 2005	AY930096	Upper Rusizi, Burundi	n/a	L. De Vos (T13-Aug 93)	+	+
Haplochromis obliquidens	Salzburger et al. 2005	AY930097	Lake Victoria	n/a	Aquarium trade	+	+
Sargochromis giardi	Salzburger et al. 2005	AY930098	Zambezi River, Zambia	n/a	C. Katongo / C. Sturmbauer		+
Cyclopharynx fwae	Salzburger et al. 2005	AY930099	Lake Fwa, Congo	n/a	U. Schliewen		+
Ctenochromis horei	Salzburger et al. 2005	AY930100	Lake Tanganyika	n/a	C. Sturmbauer/W. Salzburger		+
Haplochromis sp. 62	Salzburger et al. 2005	AY930101	Tanzania	n/a	L. De Vos (H62)	+	+
Haplochromis sp. 63	Salzburger et al. 2005	AY930102	Tanzania	n/a	L. De Vos (H63)	+	+
Haplochromis sp. 93/3	Salzburger et al. 2005	AY930103	Tanzania	n/a	L. Seegers (93/3)	+	+
Haplochromis sp. 93/40	Salzburger et al. 2005	AY930104	Tanzania	n/a	L. Seegers (93/40)	+	+
Haplochromis sp. 93/8	Salzburger et al. 2005	AY930105	Tanzania	n/a	L. Seegers (93/8)	+	+
Pseudocrenilabrus multicolor	Salzburger et al. 2005	AY930106	Tanzania	n/a	L. Seegers (91/137)		+
Haplochromis paludinosus	Salzburger et al. 2005	AY930107	Nanganga, Burundi	n/a	L. De Vos (T2-5/27/93)	+	+
Haplochromis gracilior	Salzburger et al. 2006	AY930079	Lake Kivu	n/a	E. Verheyen; K9	+	+
Congolapia bilineata	Schwarzer et al. 2011	JX157060	Itimbiri, DRC	n/a	ZSM		+
Lamprologus tigripictilis	Schwarzer et al. 2011	JX157061	Lower Congo, DRC	n/a	ZSM		+
Pseudocrenilabrus multicolor	Schwarzer et al. 2011	JX157062	Nile Delta, Egypt	n/a	ZSM	+	+
Orthochromis stormsi	Schwarzer et al. 2011	JX157063	Pool Malebo, DRC	n/a	ZSM		+
Orthochromis stormsi	Schwarzer et al. 2011	JX157064	Pool Malebo, DRC	n/a	ZSM		+
Orthochromis cf. stormsi 'Kisangani'	Schwarzer et al. 2011	JX157065	around Kisangani, DRC	n/a	ZSM		+
Orthochromis cf. stormsi 'Kisangani'	Schwarzer et al. 2011	JX157066	around Kisangani, DRC	n/a	ZSM		+
Orthochromis polyacanthus	Schwarzer et al. 2011	JX157067	LakeMweru, Zambia	n/a	EAWAG		+
Orthochromis aff. kalungwishiensis	Schwarzer et al. 2011	JX157068	LakeMweru, Zambia	n/a	EAWAG	+	+
Ctenochromis horei	Schwarzer et al. 2011	JX157069	Lake Tanganyika	n/a	CU		+
Ctenochromis horei	Schwarzer et al. 2011	JX157070	Lake Tanganyika	n/a	CU	+	+
Tropheus moorii	Schwarzer et al. 2011	JX157071	Lake Tanganyika	n/a	CU		+
Astatoreochromis alluaudi	Schwarzer et al. 2011	JX157072	Nile / Lake Victoria	n/a	EAWAG	+	+
Haplochromis burtoni	Schwarzer et al. 2011	JX157073	Lake Tanganyika	n/a	ZSM	+	+
Pseudotropheus socolofi	Schwarzer et al. 2011	JX157074	Lake Malawi	n/a	EAWAG	+	+

Labidochromis caeruleus	Schwarzer et al. 2011	JX157075	Lake Malawi	n/a	ZSM	+	+
Rhamphochromis sp.	Schwarzer et al. 2011	JX157076	Lake Malawi	n/a	ZSM	+	+
Sciaenochromis fryeri	Schwarzer et al. 2011	JX157077	Lake Malawi	n/a	ZSM	+	+
Astatotilapia desfontanii	Schwarzer et al. 2011	JX157078	Sahara, Tunesia	n/a	ZSM	+	+
Neochromis rufocaudalis	Schwarzer et al. 2011	JX157079	Nile / Lake Victoria	n/a	ZSM	+	+
Haplochromis sp. 'Kyoga'	Schwarzer et al. 2011	JX157080	Lake Kyoga, Uganda	n/a	ZSM	+	+
Haplochromis stappersii	Schwarzer et al. 2011	JX157081	Lake Tanganyika drainage, Burundi	n/a	ZSM	+	+
Haplochromis sp. 'Yaekama'	Schwarzer et al. 2011	JX157082	around Kisangani, DRC	n/a	ZSM	+	+
Haplochromis sp. 'Lake Rakai'	Schwarzer et al. 2011	JX157083	Nile / L. Rakai, Uganda	n/a	ZSM	+	+
Haplochromis sp. 'Lake Kijanebalola'	Schwarzer et al. 2011	JX157084	Nile / Lake Kijanebalola, Uganda	n/a	ZSM	+	+
Haplochromis thereuterion	Schwarzer et al. 2011	JX157085	Lake Victoria	n/a	ZSM	+	+
Haplochromis cf. polli 'Lefini'	Schwarzer et al. 2011	JX157086	Lefini River, ROC	n/a	MRAC		+
Haplochromis cf. polli 'Lefini'	Schwarzer et al. 2011	JX157087	Lefini River, ROC	n/a	MRAC		+
Haplochromis polli	Schwarzer et al. 2011	JX157088	Lower Congo River	n/a	ZSM		+
Haplochromis polli	Schwarzer et al. 2011	JX157089	Lower Congo River	n/a	ZSM		+
Haplochromis oligacanthus	Schwarzer et al. 2011	JX157090	Ubangi River, CAR	n/a	ZSM		+
Haplochromis oligacanthus	Schwarzer et al. 2011	JX157091	Ubangi River, CAR	n/a	ZSM		+
Haplochromis fasciatus	Schwarzer et al. 2011	JX157092	Lower Congo River	n/a	ZSM		+
Haplochromis fasciatus	Schwarzer et al. 2011	JX157093	Lower Congo River	n/a	ZSM	+	+
Haplochromis demeusii	Schwarzer et al. 2011	JX157094	Lower Congo River	n/a	ZSM		+
Haplochromis demeusii	Schwarzer et al. 2011	JX157095	Lower Congo River	n/a	ZSM	+	+
Haplochromis sp. 'Sanzikwa'	Schwarzer et al. 2011	JX157096	Sanzikwa River, DRC	n/a	ZSM		+
Haplochromis sp. 'Sanzikwa'	Schwarzer et al. 2011	JX157097	Sanzikwa River, DRC	n/a	ZSM	+	+
Haplochromis cf. bakongo	Schwarzer et al. 2011	JX157098	Kwilu River, DRC	n/a	ZSM		+
Haplochromis cf. bakongo	Schwarzer et al. 2011	JX157099	Kwilu River, DRC	n/a	ZSM		+
Haplochromis snoeksi	Schwarzer et al. 2011	JX157100	Inkisi River, DRC	n/a	MRAC	+	+
Thoracochromis callichromus	Schwarzer et al. 2011	JX157101	Lake Fwa, DRC	n/a	AMNH		+
Thoracochromis callichromus	Schwarzer et al. 2011	JX157102	Lake Fwa, DRC	n/a	AMNH		+
Cyclopharynx schwetzi	Schwarzer et al. 2011	JX157103	Lake Fwa, DRC	n/a	AMNH		+
Thoracochromis brauschi	Schwarzer et al. 2011	JX157104	Lake Fwa, DRC	n/a	AMNH		+
Schwetzochromis neodon	Schwarzer et al. 2011	JX157105	Lake Fwa, DRC	n/a	AMNH		+
Haplochromis stigmatogenys	Schwarzer et al. 2011	JX157106	Kasai River, DRC	n/a	AMNH		+
Haplochromis stigmatogenys	Schwarzer et al. 2011	JX157107	Kasai River, DRC	n/a	AMNH		+
Haplochromis sp. 'Kwango'	Schwarzer et al. 2011	JX157108	Kwango River, DRC	n/a	ZSM		+
Haplochromis sp. 'Kwango'	Schwarzer et al. 2011	JX157109	Kwango River, DRC	n/a	ZSM		+
Orthochromis torrenticola	Schwarzer et al. 2011	JX157110	Lufira, DRC	n/a	ZSM		+
Orthochromis torrenticola	Schwarzer et al. 2011	JX157111	Lufira, DRC	n/a	ZSM		+
Pharyngochromis sp. 'yellow lip'	Schwarzer et al. 2011	JX157112	Kwanza / Middel Kwanza (Angola)	n/a	SAIAB		+
Pharyngochromis sp. 'yellow lip'	Schwarzer et al. 2011	JX157113	Kwanza / Middel Kwanza, Angola	n/a	SAIAB		+
Pharyngochromis sp. 'yellow lip'	Schwarzer et al. 2011	JX157114	Kwanza / Middel Kwanza, Angola	n/a	SAIAB		+
Pharyngochromis sp. 'white tip'	Schwarzer et al. 2011	JX157115	Kwanza / Upper Lucalla, Angola	n/a	SAIAB		+
Pharyngochromis sp. 'white tip'	Schwarzer et al. 2011	JX157116	Kwanza / Upper Lucalla (Angola)	n/a	SAIAB		+
Serranochromis sp. 'red scales'	Schwarzer et al. 2011	JX157117	Kwanza / Upper Lucalla, Angola	n/a	SAIAB		+
Serranochromis sp. 'red scales'	Schwarzer et al. 2011	JX157118	Kwanza / Upper Lucalla, Angola	n/a	SAIAB		+
Pharyngochromis sp. 'yellow fins'	Schwarzer et al. 2011	JX157119	Kwanza / Upper Kwanza, Angola	n/a	SAIAB		+
Serranochromis sp. 'yellow fins'	Schwarzer et al. 2011	JX157120	Kwanza / Upper Kwanza, Angola	n/a	SAIAB	+	+
Serranochromis sp. 'black and white'	Schwarzer et al. 2011	JX157121	Kwanza / Upper Kwanza, Angola	n/a	SAIAB		+
Pharyngochromis acuticeps	Schwarzer et al. 2011	JX157122	Zambezi, Namibia	n/a	ZSM		+
Serranochromis robustus	Schwarzer et al. 2011	JX157123	Zambezi, Namibia	n/a	ZSM		+
Serranochromis macrocephalus	Schwarzer et al. 2011	JX157124	Zambezi, Namibia	n/a	ZSM		+
Serranochromis angusticeps	Schwarzer et al. 2011	JX157125	Zambezi, Namibia	n/a	ZSM		+
Serranochromis altus	Schwarzer et al. 2011	JX157126	Zambezi, Namibia	n/a	ZSM		+
Haplochromis elegans	Wagner et al. 2012	JQ950379	n/a	n/a	EAWAG, KAT_10	+	
Astatotilapia flavijosephi	Wagner et al. 2012	JQ950380	n/a	n/a	EAWAG, voucher 14	+	
Haplochromis tweddlei	Wagner et al. 2012	JQ950384	n/a	n/a	EAWAG, voucher 2_B6	+	
Haplochromis paludinosus	Weiss et al. unpublished	KJ176274	n/a	n/a	ZSM, P-AA-0595	+	

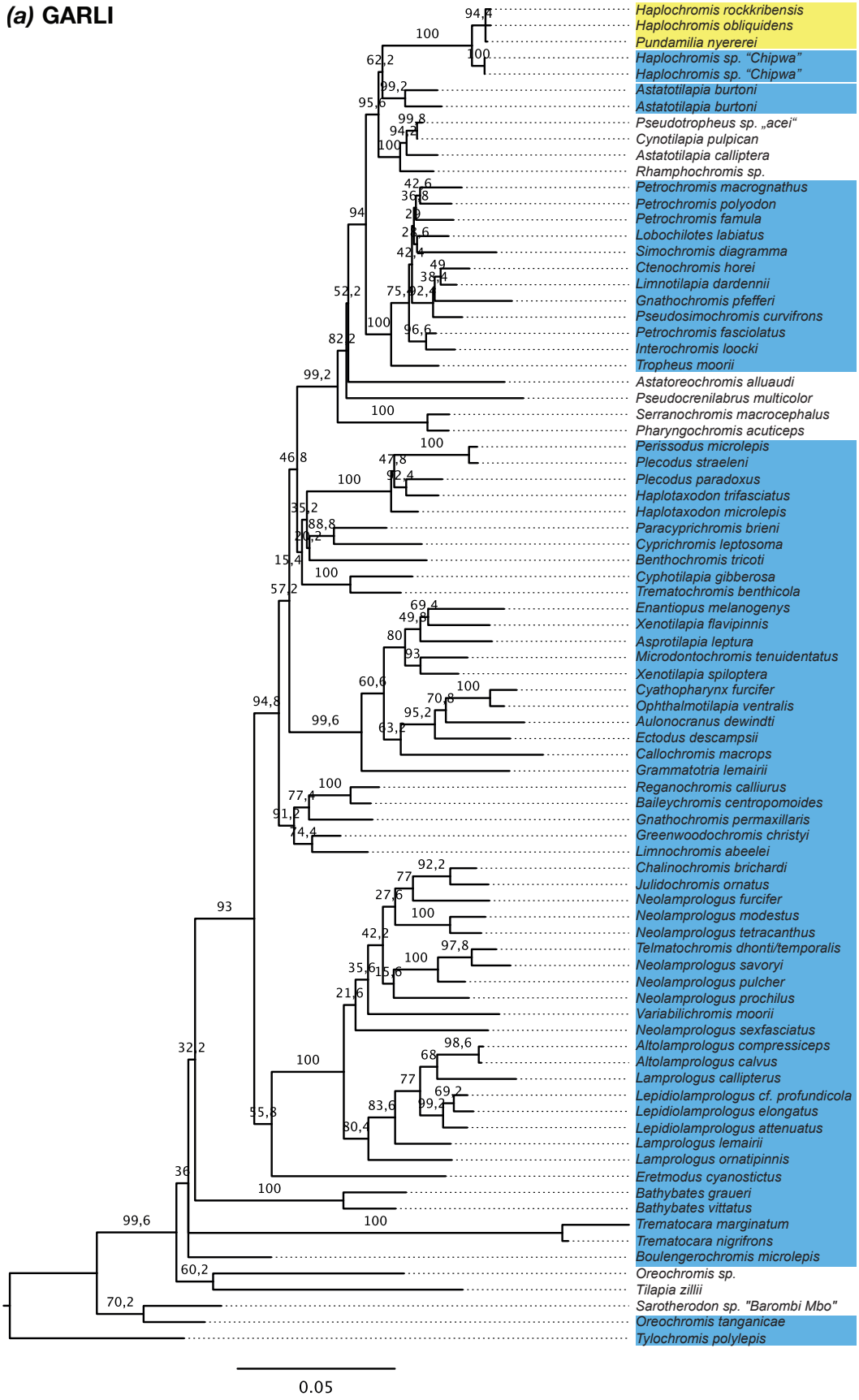
Supplementary table 3: List of the 182 haplochromine specimens and their mitochondrial control region (d-loop) accession numbers. Specified are the original publications and their sample information including haplotype number following Verheyen et al. 2003 and this study. Haplotypes used in figure 1(d) are indicated with an asterisk.

Species	Published in	Accession number	Locality	Collected by	SampleID	Haplotype in Verheyen et al. 2003 / this study
Haplochromis simpsoni	Nagl et al. 2000	AF213518	Lake Nabugabo	-	Gasi589	77
Haplochromis beadlei	Nagl et al. 2000	AF213519	Lake Nabugabo	-	Pabe593	77
Haplochromis laparogramma	Nagl et al. 2000	AF213520	Lake Victoria	-	Yila179	89
Haplochromis laparogramma	Nagl et al. 2000	AF213521	Lake Victoria	-	Yila335	80
Haplochromis laparogramma	Nagl et al. 2000	AF213522	Rusinga / Lake Victoria	-	Yila6937	25
Haplochromis lividus	Nagl et al. 2000	AF213523	Lake Victoria	-	Hali327	93
Haplochromis nubila	Nagl et al. 2000	AF213524	Lakes Nabugabo, Kayina and Kayania	-	Asnu	92*
Haplochromis chilotes	Nagl et al. 2000	AF213525	Rusinga / Lake Victoria	-	Pach	98
Haplochromis cinctus	Nagl et al. 2000	AF213526	Lake Victoria	-	Enci	77*
Haplochromis melanopterus	Nagl et al. 2000	AF213527	Lake Victoria	-	Lime	95
Neochromis nigricans	Nagl et al. 2000	AF213528	Lake Victoria	-	Neni	121
Haplochromis plagiodon	Nagl et al. 2000	AF213529	Lake Victoria	-	Papl	105
Haplochromis riponianus	Nagl et al. 2000	AF213530	Lake Victoria	-	Psri	102
Haplochromis fischeri	Nagl et al. 2000	AF213531	Lake Victoria	-	Ptsa	122
Haplochromis xenognathus	Nagl et al. 2000	AF213532	Anyanga / Lake Victoria	-	Ptxe6864	113
Haplochromis xenognathus	Nagl et al. 2000	AF213533	Anyanga / Lake Victoria	-	Ptxe6865	110
Haplochromis xenognathus	Nagl et al. 2000	AF213534	Mwanza Gulf / Lake Victoria	-	Ptxe326	109
Haplochromis xenognatus	Nagl et al. 2000	AF213535	Lake Victoria	-	Ptxe350	118*
Haplochromis nubilus	Nagl et al. 2000	AF213536	Lake Victoria	-	Asnu586	117
Prognathochromis venator	Nagl et al. 2000	AF213537	Lakes Nabugabo, Kayina and Kayania	-	Prve687	81
Prognathochromis venator	Nagl et al. 2000	AF213538	Lakes Nabugabo, Kayina and Kayania	-	Prve691	81
Haplochromis chilotes	Nagl et al. 2000	AF213539	Anyanga / Lake Victoria	-	Pach5721	79
Haplochromis chilotes	Nagl et al. 2000	AF213540	Lake Victoria	-	Pach5722	90
Haplochromis sp.'rockkribensis'	Nagl et al. 2000	AF213541	Lake Victoria	-	Haro486	108
Haplochromis sp.'rockkribensis'	Nagl et al. 2000	AF213542	Muhuru / Lake Victoria	-	Haro6745	75
Haplochromis sp.'velvetblack'	Nagl et al. 2000	AF213543	Lake Victoria	-	Havb21	115
Neochromis nigricans	Nagl et al. 2000	AF213544	Lake Victoria	-	Neni309	99
Neochromis nigricans	Nagl et al. 2000	AF213545	Lake Victoria	-	Neni817	96
Haplochromis plagiodon	Nagl et al. 2000	AF213546	Lake Victoria	-	Papl73	104
Haplochromis plagiodon	Nagl et al. 2000	AF213547	Lake Victoria	-	Papl160	91
Haplochromis plagiodon	Nagl et al. 2000	AF213548	Lake Victoria	-	Papl201	92
Haplochromis fischeri	Nagl et al. 2000	AF213549	Lake Victoria	-	Ptsa320	106
Haplochromis velifer	Nagl et al. 2000	AF213550	Lakes Nabugabo, Kayina and Kayania	-	Asve616	88
Haplochromis velifer	Nagl et al. 2000	AF213551	Lakes Nabugabo, Kayina and Kayania	-	Asve605	94
Haplochromis velifer	Nagl et al. 2000	AF213552	Lakes Nabugabo, Kayina and Kayania	-	Asve619	114
Haplochromis velifer	Nagl et al. 2000	AF213553	Lakes Nabugabo, Kayina and Kayania	-	Asve663	107
Haplochromis sp.'rockkribensis'	Nagl et al. 2000	AF213554	Lake Victoria	-	Haro6747	76*
Haplochromis sp.	Nagl et al. 2000	AF213555	WogoRiver / LakeRukwa	-	1514	27*
Haplochromis sp.	Nagl et al. 2000	AF213556	MyungaRiver / LakeRukwa	-	1605	28*
Haplochromis sp.	Nagl et al. 2000	AF213557	Kasenyi / Lake George	-	8831	73*
Haplochromis sp.	Nagl et al. 2000	AF213558	Kasenyi / Lake George	-	HT-8833	68
Haplochromis sp.	Nagl et al. 2000	AF213559	Kasenyi / Lake George	-	HT-87868786	5
Haplochromis sp.	Nagl et al. 2000	AF213560	Kasenyi / Lake George	-	HT-8801	64
Haplochromis sp.	Nagl et al. 2000	AF213561	Kasenyi / Lake George	-	HT-8837	1
Haplochromis sp.	Nagl et al. 2000	AF213562	Kasenyi / Lake George	-	HT-88348834	41
Haplochromis sp.	Nagl et al. 2000	AF213563	Kashaka / Lake George	-	HT-8924	43
Haplochromis sp.	Nagl et al. 2000	AF213564	Katwe / LakeEdward	-	HT-8880	26
Haplochromis sp.	Nagl et al. 2000	AF213566	Katwe / LakeEdward	-	HT-8879	71
Haplochromis sp.	Nagl et al. 2000	AF213567	Katwe / LakeEdward	-	HT-87688768	40
Haplochromis sp.	Nagl et al. 2000	AF213568	Katwe / LakeEdward	-	HT-8773	45
Haplochromis sp.	Nagl et al. 2000	AF213569	Katwe / LakeEdward	-	8777	46*
Haplochromis sp.	Nagl et al. 2000	AF213570	Katwe / LakeEdward	-	HT-8778	2
Haplochromis sp.	Nagl et al. 2000	AF213571	Bugoigo / LakeAlbert	-	HT-9049	66
Haplochromis sp.	Nagl et al. 2000	AF213572	Butiaba / LakeAlbert	-	HT-8990	69
Haplochromis sp.	Nagl et al. 2000	AF213573	Butiaba / LakeAlbert	-	HT-9003	44
Haplochromis sp.	Nagl et al. 2000	AF213574	Butiaba / LakeAlbert	-	HT-9019	42
Haplochromis sp.	Nagl et al. 2000	AF213575	LakeLutoto / Uganda	-	HT-8692	30
Haplochromis sp.	Nagl et al. 2000	AF213576	LakeLutoto / Uganda	-	HT-8694	31
Haplochromis sp.	Nagl et al. 2000	AF213577	LakeLutoto / Uganda	-	HT-8687	32
Haplochromis sp.	Nagl et al. 2000	AF213578	LakeChibwera / Uganda	-	HT-8947	62
Haplochromis sp.	Nagl et al. 2000	AF213579	LakeChibwera / Uganda	-	HT-8950	60
Haplochromis sp.	Nagl et al. 2000	AF213580	LakeChibwera / Uganda	-	HT-8948	61
Haplochromis sp.	Nagl et al. 2000	AF213581	LakeWamala / Lake VictoriaRegion	-	HT-8632	111

Haplochromis sp.	Nagl et al. 2000	AF213582	KatongaRiver / Lake VictoriaRegion	-	HT-8678	116
Haplochromis sp.	Nagl et al. 2000	AF213583	KatongaRiver / Lake VictoriaRegion	-	HT-8680	112
Haplochromis sp.	Nagl et al. 2000	AF213584	Kazinga Channel / L.Edward and George	-	HT-8741	70
Haplochromis sp.	Nagl et al. 2000	AF213585	Kazinga Channel / L.Edward and George	-	HT-8711	4
Haplochromis sp.	Nagl et al. 2000	AF213586	Kazinga Channel / L.Edward and George	-	HT-8718	3
Haplochromis sp.	Nagl et al. 2000	AF213587	Kazinga Channel / L.Edward and George	-	HT-87228722	39
Haplochromis sp.	Nagl et al. 2000	AF213588	MigoriRiver / Lake Victoria	-	HT-6701	87
Haplochromis sp.	Nagl et al. 2000	AF213589	Malagarazi River	-	HT-1006	na
Haplochromis sp.	Nagl et al. 2000	AF213590	Malagarazi River	-	HT-1011	na
Haplochromis sp.	Nagl et al. 2000	AF213591	Malagarazi River	-	HT-1510	na
Haplochromis sp.	Nagl et al. 2000	AF213592	Malagarazi River	-	HT-1531	na
Haplochromis sp.	Nagl et al. 2000	AF213593	Malagarazi River	-	HT-1590	na
Haplochromis sp.	Nagl et al. 2000	AF213594	Malagarazi River	-	HT-1591	na
Haplochromis sp.	Nagl et al. 2000	AF213595	Lupa River	-	HT-1597	na
Haplochromis sp.	Nagl et al. 2000	AF213596	Piti River	-	HT-1598	na
Haplochromis sp.	Nagl et al. 2000	AF213597	Piti River	-	HT-1546	na
Haplochromis sp.	Nagl et al. 2000	AF213598	Piti River	-	HT-1547	na
Haplochromis sp.	Nagl et al. 2000	AF213599	Pangani River	-	HT-1076	na
Haplochromis sp.	Nagl et al. 2000	AF213600	Pangani River	-	HT-1501	na
Haplochromis sp.	Nagl et al. 2000	AF213601	Wogo River / Lake Rukwa	-	HT-1636	na
Haplochromis sp.	Nagl et al. 2000	AF213602	Wogo River / Lake Rukwa	-	HT-1635	na
Haplochromis sp.	Nagl et al. 2000	AF213603	Wogo River / Lake Rukwa	-	HT-1515	na
Haplochromis sp.	Nagletal.2002	AF213604	Pangani River	-	HT-1530	na
Haplochromis sp.	Nagletal.2001	AF213605	Lake Chala	-	HT-1738	na
Haplochromis sp.	Nagletal.2003	AF213606	Lake Babati	-	HT-6249	na
Haplochromis sp.	Nagl et al. 2000	AF213607	Lake Manyara	-	HT-1537	na
Haplochromis sp.	Nagl et al. 2000	AF213608	Malagarazi River	-	HT-1601	na
Haplochromis sp.	Nagl et al. 2000	AF213609	Kazinga Channel / L. Edwardand George	-	HT-8746	na
Haplochromis sp.	Nagl et al. 2000	AF213610	Lake George	-	HT-8785	na
Haplochromis sp.	Nagl et al. 2000	AF213611	Lake George	-	HT-8903	na
Haplochromis sp.	Nagl et al. 2000	AF213612	Lake George	-	HT-8911	na
Haplochromis sp.	Nagl et al. 2000	AF213613	Malagarazi River	-	HT-1533	na
Haplochromis sp.	Nagl et al. 2000	AF213614	Malagarazi River	-	HT-1609	na
Astatoreochromis alluaudi	Nagl et al. 2000	AF213616	Lake Victoria	-	Asal6744	na
Astatoreochromis alluaudi	Nagl et al. 2000	AF213617	Lake Victoria	-	Asal5928	na
Pseudotropheus sp.'msobo'	Nagl et al. 2000	AF213622	Lake Malawi	-	Psms5170	na
Labeotropheus trewavasae	Nagl et al. 2000	AF213623	Lake Malawi	-	Latr5493	na
Haplochromis burtoni	Stiassny et al. 1994	AF400710	-	-	8153	na
Limnochromis auritus	Sturmbauer & Meyer 1992	AF400728	Lake Tanganyika	-	27749	na
Petrochromis orthognathus	Stiassny et al. 1994	AF400734	Lake Tanganyika	-	28818	na
Haplochromis astatodon	Verheyen et al. 2003	AY226611	Lake Kivu	E.Verheyen	K114	7
Haplochromis astatodon	Verheyen et al. 2003	AY226611	Lake Kivu	E.Verheyen	K114	7*
Haplochromis insidiae	Verheyen et al. 2003	AY226627	Lake Kivu	E.Verheyen	K080	8
Haplochromis sp.nigroides / scheffersi	Verheyen et al. 2003	AY226629	Lake Kivu	E.Verheyen	K146	9
Haplochromis astatodon	Verheyen et al. 2003	AY226631	Lake Kivu	E.Verheyen	K119	10
Haplochromis astatodon	Verheyen et al. 2003	AY226632	Lake Kivu	E.Verheyen	K131	11
Haplochromis paucidens	Verheyen et al. 2003	AY226633	Lake Kivu	E.Verheyen	K112	12
Haplochromis paucidens	Verheyen et al. 2003	AY226640	Lake Kivu	E.Verheyen	K022	13
Haplochromis paucidens	Verheyen et al. 2003	AY226641	Lake Kivu	E.Verheyen	K034	14
Haplochromis sp.crebridens / olivaceu	Verheyen et al. 2003	AY226642	Lake Kivu	E.Verheyen	K036	15
Haplochromis astatodon	Verheyen et al. 2003	AY226643	Lake Kivu	E.Verheyen	K127	16
Haplochromis sp.crebridens / olivaceu	Verheyen et al. 2003	AY226646	Lake Kivu	E.Verheyen	K060	17
Haplochromis scheffersi	Verheyen et al. 2003	AY226647	Lake Kivu	E.Verheyen	K111	18
Haplochromis graueri	Verheyen et al. 2003	AY226648	Lake Kivu	E.Verheyen	K118	19
Haplochromis graueri	Verheyen et al. 2003	AY226649	Lake Kivu	E.Verheyen	K012	20
Haplochromis astatodon	Verheyen et al. 2003	AY226650	Lake Kivu	E.Verheyen	K115	21
Haplochromis astatodon	Verheyen et al. 2003	AY226651	Lake Kivu	E.Verheyen	K124	22
Haplochromis astatodon	Verheyen et al. 2003	AY226652	Lake Kivu	E.Verheyen	K076	23
Haplochromis astatodon	Verheyen et al. 2003	AY226654	Lake Kivu	E.Verheyen	K132	24
Haplochromis astatodon	Verheyen et al. 2003	AY226655	Lake Kivu	E.Verheyen	K51	25*
Haplochromis occultidens	Verheyen et al. 2003	AY226666	Lake Kivu	E.Verheyen	K030	33
Haplochromis graueri	Verheyen et al. 2003	AY226668	Lake Kivu	E.Verheyen	K001	36
Haplochromis astatodon	Verheyen et al. 2003	AY226669	Lake Kivu	E.Verheyen	K116	37
Haplochromis sp.crebridens / olivaceu	Verheyen et al. 2003	AY226670	Lake Kivu	E.Verheyen	K057	38
Haplochromis astatodon	Verheyen et al. 2003	AY226671	Lake Kivu	E.Verheyen	K135	47*
Haplochromis paucidens	Verheyen et al. 2003	AY226687	Lake Kivu	E.Verheyen	K056	48

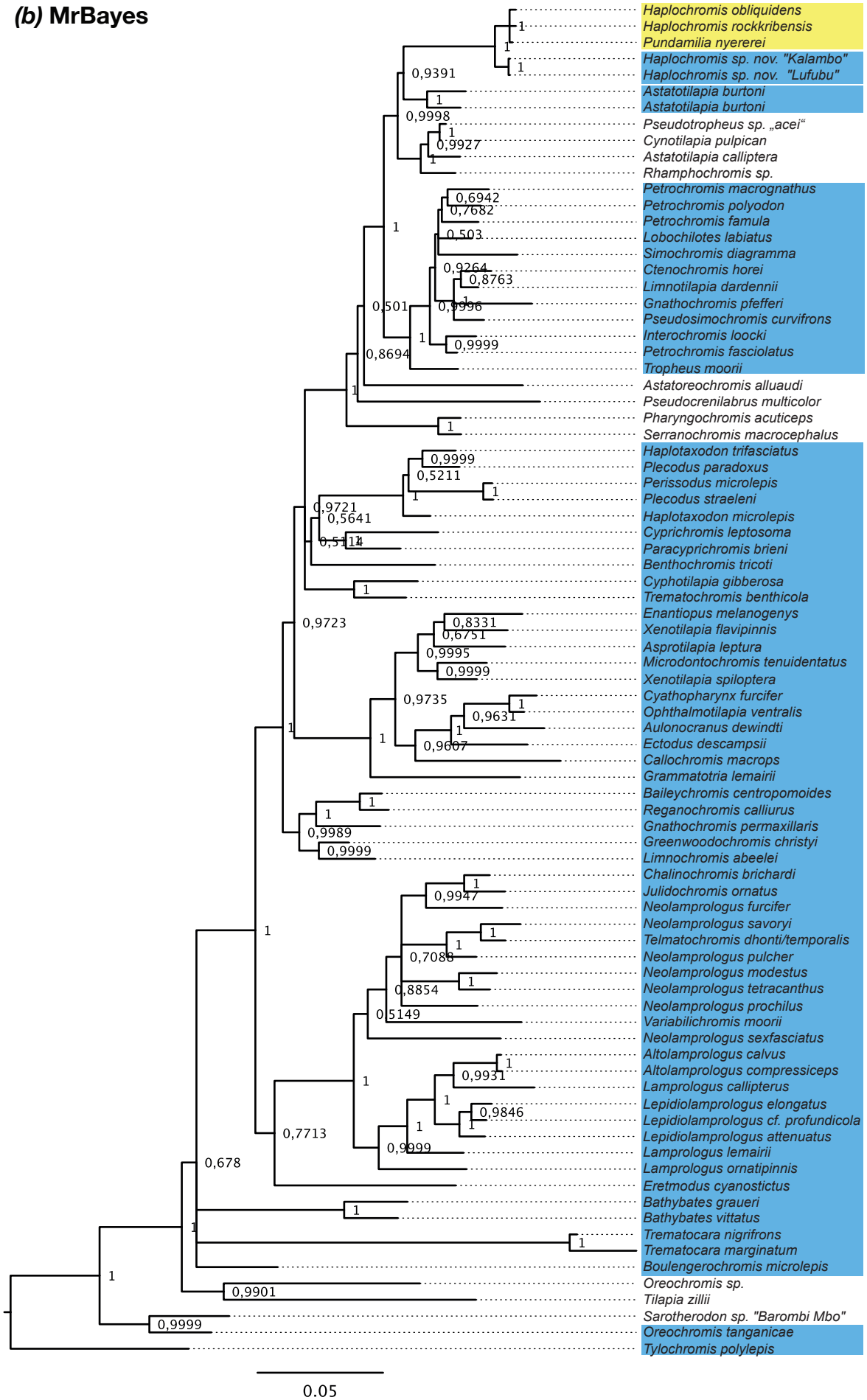
Haplochromis nigroides	Verheyen et al. 2003	AY226688	Lake Kivu	E.Verheyen	K028	49
Haplochromis astatodon	Verheyen et al. 2003	AY226691	Lake Kivu	E.Verheyen	K152	50
Haplochromis astatodon	Verheyen et al. 2003	AY226692	Lake Kivu	E.Verheyen	K138	51
Haplochromis paucidens	Verheyen et al. 2003	AY226694	Lake Kivu	E.Verheyen	K058	53
Haplochromis microchrysomelas	Verheyen et al. 2003	AY226695	Lake Kivu	E.Verheyen	K113	54
Haplochromis astatodon	Verheyen et al. 2003	AY226697	Lake Kivu	E.Verheyen	K120	55
Haplochromis microchrysomelas	Verheyen et al. 2003	AY226699	Lake Kivu	E.Verheyen	K142	56*
Haplochromis paucidens	Verheyen et al. 2003	AY226712	Lake Kivu	E.Verheyen	K174	57
Haplochromis crebridens	Verheyen et al. 2003	AY226714	Lake Kivu	E.Verheyen	K177	58
Haplochromis adolfifrederici	Verheyen et al. 2003	AY226715	Lake Kivu	E.Verheyen	K169	59
Haplochromis crebridens	Verheyen et al. 2003	AY226716	Lake Kivu	E.Verheyen	K063	74
Haplochromis sp.	Verheyen et al. 2003	AY226719	Cohoha / Bugesera Lakes	J.Snoeks	D9	82*
Haplochromis sp.	Verheyen et al. 2003	AY226720	Cohoha / Bugesera Lakes	J.Snoeks	B4	83
Haplochromis sp.	Nagl et al. 2000	AY226723	Rweru / Bugesera Lakes	-	R1	84
Haplochromis sp.	Verheyen et al. 2003	AY226726	Cohoha / Bugesera Lakes	J.Snoeks	D8	85
Haplochromis sp.	Verheyen et al. 2003	AY226727	Kachera / Uganda	E.Schraml	9803	6
Haplochromis sp.	Verheyen et al. 2003	AY226734	Victoria Nile	E.Schraml	9791	29
Haplochromis sp.	Verheyen et al. 2003	AY226735	Mugogo / Uganda	E.Schraml	9784	32
Haplochromis sp.	Verheyen et al. 2003	AY226736	aquarium trade	E.Schraml	9808	63
Haplochromis squamipinnis	Verheyen et al. 2003	AY226747	Lake Edward	E.Schraml	9813	65
Haplochromis sp.	Verheyen et al. 2003	AY226752	Nyamusingire / Uganda	E.Schraml	9765	67
Haplochromis sp.	Verheyen et al. 2003	AY226758	Nakivali / Uganda	E.Schraml	9721	72
Haplochromis sp.	Verheyen et al. 2003	AY226759	Lake Victoria	E.Schraml	9707	77
Haplochromis sp.	Verheyen et al. 2003	AY226761	Nawampasa / Lake Kyoga	E.Schraml	9788	78
Haplochromis sp.	Verheyen et al. 2003	AY226762	Lake Victoria	E.Schraml	9801	86
Haplochromis sp.	Verheyen et al. 2003	AY226763	Lake Victoria	E.Schraml	9713	91*
Haplochromis sp.	Verheyen et al. 2003	AY226764	Lake Victoria	E.Schraml	9706	92
Haplochromis sp.	Verheyen et al. 2003	AY226765	Lake Victoria	E.Schraml	9715	92
Haplochromis sp.	Verheyen et al. 2003	AY226766	Nawamapasa / Lake Kyoga	E.Schraml	9789	97*
Haplochromis sp.	Verheyen et al. 2003	AY226767	Lake Victoria	E.Schraml	9812	100
Haplochromis sp.	Verheyen et al. 2003	AY226768	Mulehe / Kabale Lakes	E.Schraml	9764	101*
Haplochromis sp.	Verheyen et al. 2003	AY226769	Lake Victoria	E.Schraml	9704	101
Haplochromis sp.	Verheyen et al. 2003	AY226779	Lake Victoria	E.Schraml	9703	103
Haplochromis sp.	Verheyen et al. 2003	AY226781	Bunyoni / Kabale Lakes	-	9727	119
Haplochromis sp.	Verheyen et al. 2003	AY226784	Bunyoni / Kabale Lakes	E.Schraml	9741	120
Haplochromis burtoni	Verheyen et al. 2003	AY226785	Cohoha / Bugesera Lakes	J.Snoeks	B6	na
Astatoreochromis alluaudi	Verheyen et al. 2003	AY226787	Cohoha / Bugesera Lakes	J.Snoeks	E9	na
Haplochromis gracilior	Verheyen et al. 2003	AY226788	Lake Kivu	-	K008	na
Haplochromis gracilior	Verheyen et al. 2003	AY226789	Lake Kivu	-	K009	na
Haplochromis gracilior	Verheyen et al. 2003	AY226790	Lake Kivu	-	K010	na
Thoracochromis brauschi	Verheyen et al. 2003	AY226791	Lac Fwa	Paul	9792	na
Serranochromis sp.WWS-2003	Verheyen et al. 2003	AY226792	Lake Mweru-Wantipa	T.Reuter	9793	na
Haplochromis stappersii	Salzburgeretal.2005	AY929941	Malagarazi River	L.DeVos	5-6 / 25 / 92	M3*
Haplochromis sp.	Salzburgeretal.2005	AY929992	Tanzania	L.Seegers	93 / 8	LR2*
Haplochromis sp.	Salzburgeretal.2005	AY930015	Tanzania	L.Seegers	92 / 12	LR1*
Cyrtocara moorii	Sturmbauer & Meyer 1992	U12554	Lake Tanganyika	-	30882	na
Haplochromis stappersii "Malagarasi1"	this study	KJ955382	Malagarazi River / Burundi	G.Banyankimbona	MRAC1840	M1*
Haplochromis stappersii "Malagarasi2"	this study	KJ955384	Malagarazi River / Burundi	G.Banyankimbona	MRAC1847	M1*
Haplochromis stappersii "Malagarasi3"	this study	KJ955385	Malagarazi River / Burundi	G.Banyankimbona	MRAC12034	M1*
Haplochromis stappersii "Malagarasi4"	this study	KJ955383	Malagarazi River / Burundi	G.Banyankimbona	MRAC12087	M2*
Haplochromis sp. "Chipwa"	this study	KJ955386	Kalambo River / Zambia	W.Salzburger	CH4	HLT*
Haplochromis sp. "Chipwa"	this study	KJ955387	Lufubu River / Zambia	W.Salzburger	LU2	HLT*
Haplochromis stappersii "Rusizi"	this study	KJ955381	Gatumbamarsh, Rusizi River	G.Banyankimbona	MRAC6334	RR*

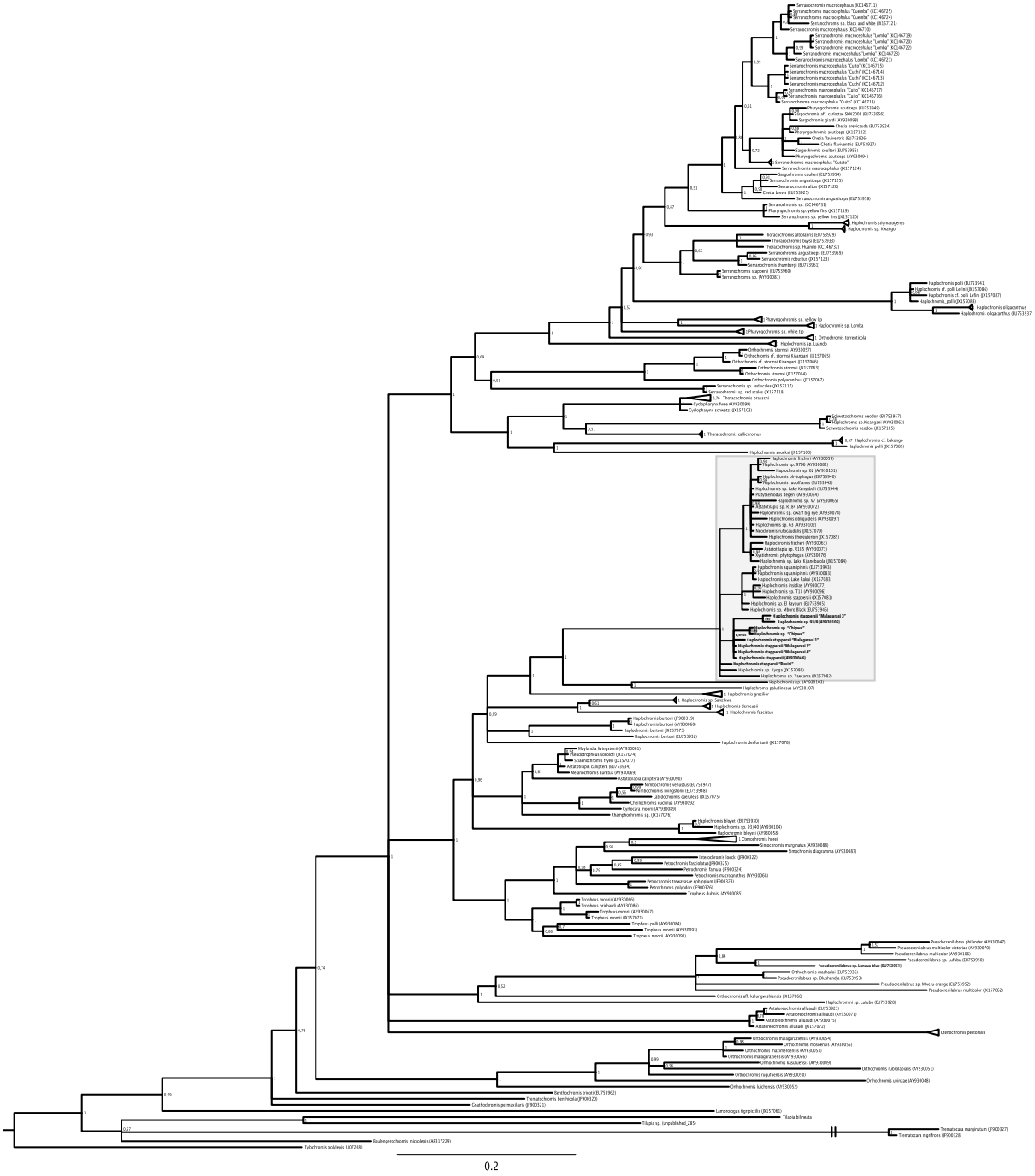
***(a)* GARLI**



Maximum likelihood (*a*) and Bayesian (*b*) tree based on the concatenated dataset (table S1). All bootstrap support values and posterior probabilities are plotted. The geographical origin of the specimen is indicated in color (blue = Lake Tanganyika; yellow = Lake Victoria; other locations are not further indicated).

(b) MrBayes

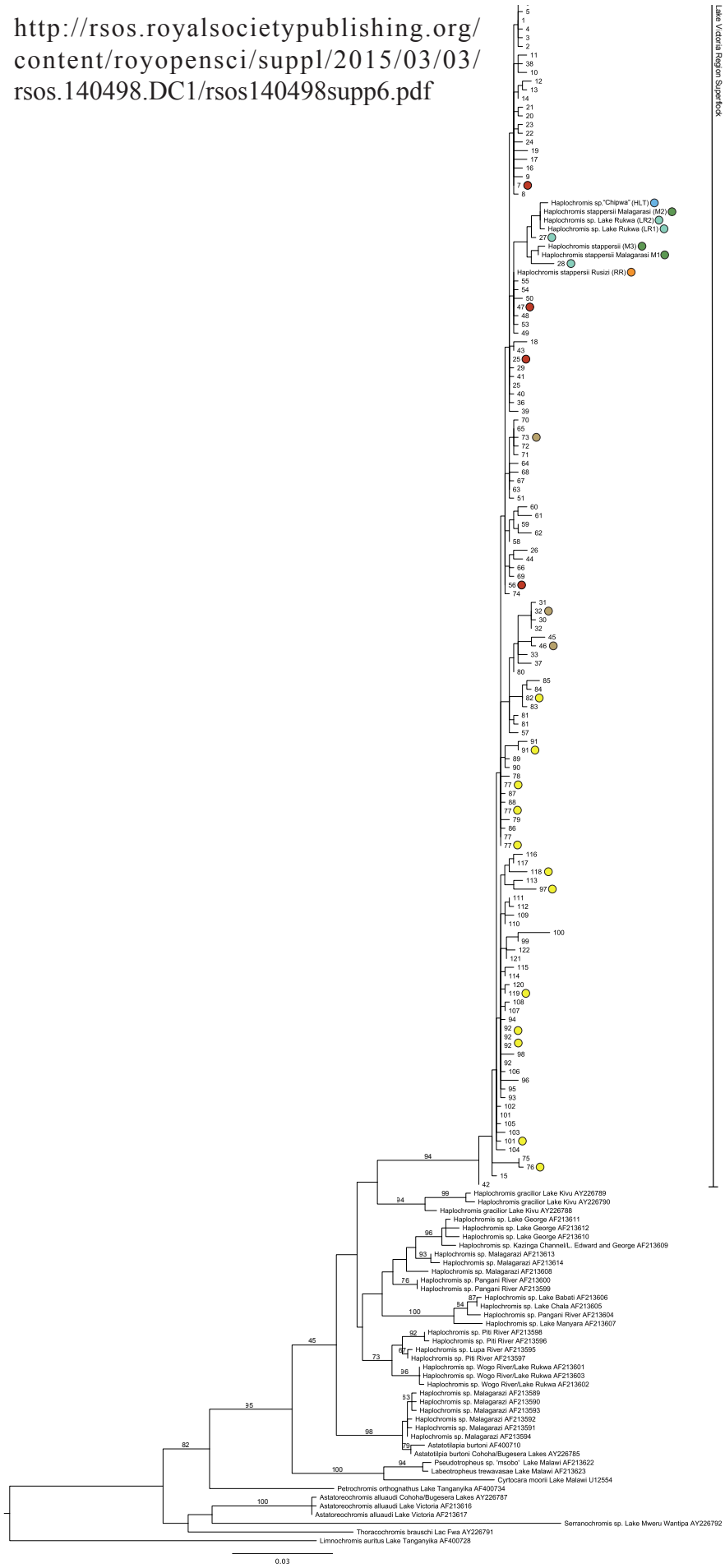

Haplochromis obliquidens
Haplochromis rockkribensis
Pundamilia nyererei
Haplochromis sp. nov. "Kalambo"
Haplochromis sp. nov. "Lufubu"
Astatotilapia burtoni
Astatotilapia burtoni
Pseudotropheus sp. „acei“
Cynotilapia pulpican
Astatotilapia calliptera
Rhamphochromis sp.
Petrochromis macrognathus
Petrochromis polyodon
Petrochromis famula
Lobochilotes labiatus
Simochromis diagramma
Ctenochromis horei
Limnotilapia dardennii
Gnathochromis pfefferi
Pseudosimochromis curvifrons
Interochromis loocki
Petrochromis fasciolatus
Tropheus moorii
Astatoreochromis alluaudi
Pseudocrenilabrus multicolor
Pharyngochromis acuticeps
Serranochromis macrocephalus
Haplotaxodon trifasciatus
Plecodus paradoxus
Perissodus microlepis
Plecodus straeleni
Haplotaxodon microlepis
Cyprichromis leptosoma
Paracyprichromis brieni
Benthochromis tricoti
Cyphotilapia gibberosa
Trematochromis benthicola
Enantiopus melanogenys
Xenotilapia flavipinnis
Asprotilapia leptura
Microdontochromis tenuidentatus
Xenotilapia spiloptera
Cyathopharynx furcifer
Ophthalmotilapia ventralis
Aulonocranus dewindti
Ectodus descampsii
Callochromis macrops
Grammatotria lemairii
Baileychromis centropomoides
Reganochromis calliurus
Gnathochromis permaxillaris
Greenwoodochromis christyi
Limnochromis abeelei
Chalinochromis brichardi
Julidochromis ornatus
Neolamprologus furcifer
Neolamprologus savoryi
Telmatochromis dhonti/temporalis
Neolamprologus pulcher
Neolamprologus modestus
Neolamprologus tetracanthus
Neolamprologus prochilus
Variabilichromis moorii
Neolamprologus sexfasciatus
Altolamprologus calvus
Altolamprologus compressiceps
Lamprologus callipterus
Lepidiolamprologus elongatus
Lepidiolamprologus cf. profundicola
Lepidiolamprologus attenuatus
Lamprologus lemairii
Lamprologus ornatipinnis
Eretmodus cyanostictus
Bathybates graueri
Bathybates vittatus
Trematocara nigrifrons
Trematocara marginatum
Boulengerochromis microlepis
Oreochromis sp.
Tilapia zillii
Sarotherodon sp. "Barombi Mbo"
Oreochromis tanganicae
Tylochromis polylepis
0,9391
0,9998
0,9927
0,6942
0,7682
0,503
0,9264
0,8763
0,501
0,9996
0,9999
0,8694
0,9999
0,5211
0,9721
0,5641
0,9723
0,8331
0,6751
0,9995
0,9999
0,9735
0,9631
0,9607
0,9989
0,9999
0,9947
0,7088
0,8854
0,5149
0,9931
0,9846
0,7713
0,9999
0,678
0,9901
0,9999
0.05



Supplementary figure 2: MrBayes 50% majority rule consensus tree with branch lengths based on the ND2 data set (table S2). Posterior probabilities ≥ 0.5 are plotted. The grey box represents the Lake Victoria Region superflock.

http://rsos.royalsocietypublishing.org/content/royopensci/suppl/2015/03/03/rsos.140498.DC1/rsos140498supp5.pdf

http://rsos.royalsocietypublishing.org/content/royopensci/suppl/2015/03/03/rsos.140498.DC1/rsos140498supp6.pdf



Chapter 4

The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change

M. Emília Santos, Ingo Braasch, Nicolas Boileau, Britta S. Meyer, Loïc Sauteur, Astrid Böhne, Heinz-Georg Belting, Markus Affolter, Walter Salzburger

Nature Communications (2014) 5, 5149
doi:10.1038/ncomms6149

M.E.S., I.B. and W.S. designed the study; M.E.S and W.S. collected the samples; M.E.S. performed the RNAseq, gene expression, comparative genomics and zebrafish functional analysis; N.B. performed the sequencing of *fhl2* paralogs coding region and analysed its rates of evolution; B.S.M. collected the 454 sequence data; and B.S.M. and W.S. performed the phylogenetic analysis; A.B. performed the SINE consensus alignments and analysed the transposable element composition of *fhl2b* genomic region; I.B. performed the zebrafish functional assays of the A. burtoni construct and *fhl2* paralogs synteny analysis; L.S., H.-G.B. and M.A. assisted with the zebrafish functional assays of the *A. burtoni, P. philander* and *N. sexfasciatus* construct; and M.E.S. and W.S. wrote the paper and all authors contributed to revisions.

nature COMMUNICATIONS

ARTICLE

Received 4 Feb 2014 | Accepted 4 Sep 2014 | Published 9 Oct 2014

DOI: 10.1038/ncomms6149 OPEN

The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change

M. Emília Santos[1,†], Ingo Braasch[2], Nicolas Boileau[1], Britta S. Meyer[1,†], Loïc Sauteur[3], Astrid Böhne[1], Heinz-Georg Belting[3], Markus Affolter[3] & Walter Salzburger[1]

The origin of novel phenotypic characters is a key component in organismal diversification; yet, the mechanisms underlying the emergence of such evolutionary novelties are largely unknown. Here we examine the origin of egg-spots, an evolutionary innovation of the most species-rich group of cichlids, the haplochromines, where these conspicuous male fin colour markings are involved in mating. Applying a combination of RNAseq, comparative genomics and functional experiments, we identify two novel pigmentation genes, *fhl2a* and *fhl2b*, and show that especially the more rapidly evolving b-paralog is associated with egg-spot formation. We further find that egg-spot bearing haplochromines, but not other cichlids, feature a transposable element in the *cis*-regulatory region of *fhl2b*. Using transgenic zebrafish, we finally demonstrate that this region shows specific enhancer activities in iridophores, a type of pigment cells found in egg-spots, suggesting that a *cis*-regulatory change is causally linked to the gain of expression in egg-spot bearing haplochromines.

[1] Zoological Institute, University of Basel, Basel 4051, Switzerland. [2] Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403-1254, USA. [3] Biozentrum, University of Basel, Basel 4056, Switzerland. † Present address: Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure, CNRS UMR 5242, 46 Allée d'Italie, 69364 Lyon Cedex 07, France (M.E.S.); Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany (B.S.M.). Correspondence and requests for materials should be addressed to M.E.S. (email: emilia.p.santos@gmail.com) or to W.S. (email: walter.salzburger@unibas.ch).

The *de novo* evolution of complex phenotypic traits poses a challenge to evolutionary biology[1–5]. While selection explains adaptation and speciation in an adequate manner[6], it is more difficult to conceive how selection would trigger the origin of evolutionary novelties such as insect wings, feathers, tetrapod limbs, flowers, the mammalian placenta, beetle horns or butterfly eye-spots[1,4,5,7,8]. The emergence of evolutionary innovations, that is, lineage-restricted traits linked to qualitatively new functions, involves the origin of new developmental modules that are responsible for the identity of these novel characters[4,5]. Most of the available evidence suggests that new developmental programs emerge largely through co-option of pre-existing regulatory gene networks via changes in their regulation and deployment ('old genes playing new tricks'[5]). Uncovering the mechanisms of how these developmental modules are co-opted or newly evolved is one of the primary goals of evo-devo research[2,3,5,7,8].

Anal fin egg-spots are an evolutionary innovation in the so-called 'haplochromines'[9] (Fig. 1a and Supplementary Fig. 1), the most species-rich group of cichlid fishes, best known for their spectacular adaptive radiations in the East African lakes Victoria and Malawi[10,11]. Adult males of ~1,500 cichlid species feature this pigmentation trait in the form of conspicuously coloured circular markings[9,11,12]. Haplochromine egg-spots vary substantially in colour, shape, number and arrangement between species (Fig. 1b), and even within species a certain degree of variation is observed. In some species, also females show egg-spots, which are then much less pronounced and colourful. The function of egg-spots has been implicated with the mating behaviour of the female-mouthbrooding haplochromines[12,13]. Immediately upon spawning, a haplochromine female gathers up her eggs into the mouth; the male then presents his egg-spots to which the female responds by snatching and bringing her mouth close to the male's genital opening; upon discharging sperm, the eggs become fertilized inside the female's mouth (Fig. 1c). The mother subsequently broods and carries her progeny in the oral cavities for several weeks after fertilization.

Here we are interested in the molecular basis of the anal fin egg-spots of haplochromine cichlids. The main advantages of the cichlid egg-spot system are that (i) the evolutionary innovation of interest emerged just a few million years ago and hence is recent compared with most other evolutionary novelties studied so far[9,10,14]; (ii) the phylogenetic context in which the novel trait evolved is known and living sister clades to the lineage featuring the novelty still exist[9,15,16]; and (iii) the genomes of two outgroup species lacking the trait and of three derived species featuring the trait are available. This allows us to study early events involved in the origin of an evolutionary innovation in an assemblage of phenotypically diverse, yet closely related and genetically similar species[14]. Using RNAseq, we identify two novel candidate pigmentation genes, the a- and b-paralogs of the four and a half LIM domain protein 2 (*fhl2*) gene, and show that both genes, but especially the more rapidly evolving b-copy, are associated with the formation of egg-spots. We then find that egg-spot bearing haplochromines—but not an egg-spot-less ancestral haplochromine and not the representatives from more basal

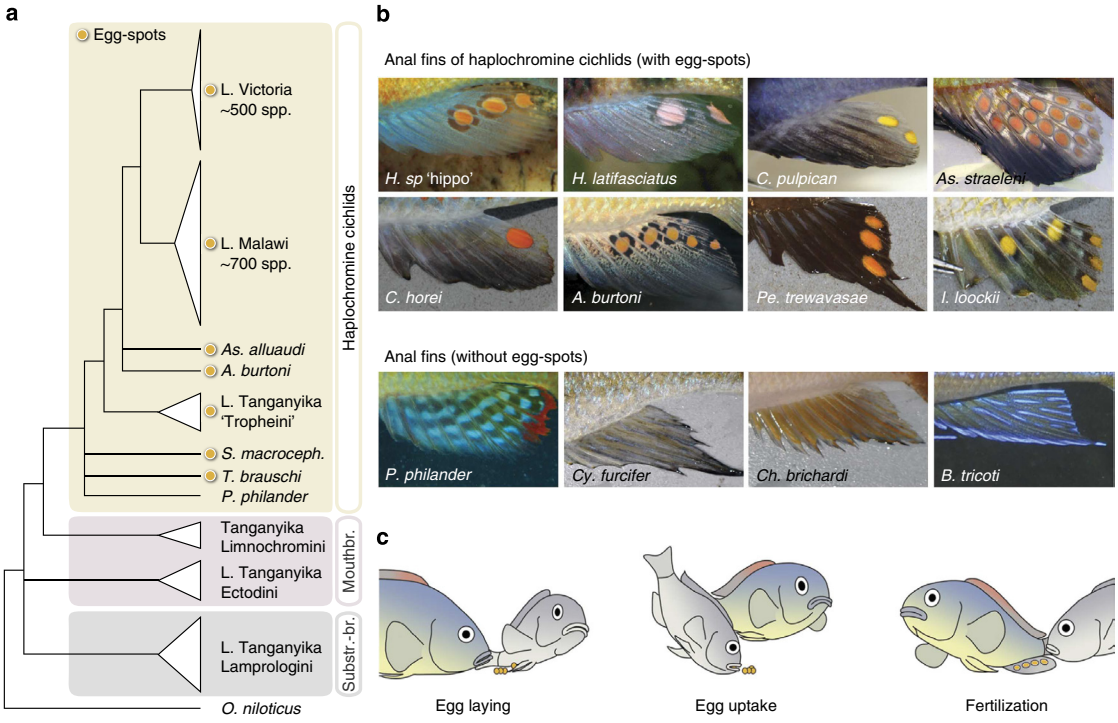


Figure 1 | The egg-spots of haplochromine cichlids. (**a**) Phylogeny of the East African cichlid fishes based on a new multimarker data set. The haplochromines are the most species-rich and derived group of cichlids in East Africa. One of the common features of haplochromines is the presence of egg-spots on the anal fin of males. Note that one of the ancestral lineages, represented here by *P. philander*, does not show this characteristic trait[9,33]. Substr-br, substrate brooders; mouthbr, mouthbrooders; spp.: species. (**b**) Examples of male anal fin patterns in East African cichlids. Haplochromine egg-spots (upper panel) vary in size, shape, number and colouration. Non-haplochromines and basal haplochromine *P. philander* (lower panel) do not show this trait. (**c**) A typical mating cycle of haplochromine cichlids.

cichlid lineages—exhibit a transposable element insertion in close proximity to the transcription initiation site of *fhl2b*. A functional assay with transgenic zebrafish reveals that only a haplochromine-derived genetic construct featuring the SINE (short interspersed repetitive element) insertion drove expression in a special type of pigment cells, iridophores. Together, our data suggest that a *cis*-regulatory change (probably in the form of a SINE insertion) is responsible for the gain of expression of *fhl2b* in iridophores, contributing to the evolution of egg-spots in haplochromine cichlids.

Results

***fhl2* paralogs: novel candidates for egg-spot morphogenesis**. As a first step, we performed an Illumina-based comparative transcriptomic experiment (RNAseq) between male (with egg-spots) and female (without egg-spots) anal fins in the haplochromine cichlid *Astatotilapia burtoni*. Two of the most differentially expressed genes according to RNAseq were the a- and b-paralogs of *fhl2* (~4 $\log_2$-fold and ~5 $\log_2$-fold differences, respectively; see Supplementary Table 2). These paralogs result from the teleost genome duplication[17] (Supplementary Fig. 2). The four and a half LIM domain protein 2 (Fhl2) is known as a transcriptional co-activator of the androgen receptor and the *Wnt*-signalling pathway[18,19]; Fhl2 plays a role in cell-fate determination and pattern formation, in the organization of the cytoskeleton, in cell adhesion, cell motility and signal transduction; furthermore, it regulates the development of heart, bone and musculature in vertebrates[20,21].

Expression of *fhl2a* and *fhl2b* is egg-spot specific. To confirm the results obtained by RNAseq, we performed quantitative real-time PCR (qPCR) experiments (Fig. 2a), this time also comparing egg-spot versus non-egg-spot tissue within male anal fins. In addition, we tested another haplochromine species, *Cynotilapia pulpican*, with a different egg-spot arrangement to exclude positional effects of gene expression on the anal fin. In both species, the two duplicates of *fhl2* were overexpressed in egg-spots (*A. burtoni*: *fhl2a*: $t_5 = 10.77$, $P = 0.0001$; *fhl2b*: $t_5 = 4.362$, $P = 0.0073$; *C. pulpican*: *fhl2a*: $t_4 = 5.031$, $P = 0.0073$; *fhl2b*: $t_4 = 9.154$, $P = 0.0008$). We then tested the expression of both *fhl2* paralogs in the four main developmental stages of egg-spot formation in *A. burtoni*[22] and compared it with other candidate pigmentation genes (including the previously identified xanthophore marker *csf1ra*, the melanophore marker *mitfa* and the iridophore marker *pnp4a*). We found that the expression of both *fhl2* paralogs increases substantially throughout anal fin and egg-spot development, and both genes showed higher expression levels compared with the other pigmentation genes (Fig. 2b); *fhl2b* shows the highest increase in expression exactly when egg-spots begin to form. Furthermore, we corroborate that the expression

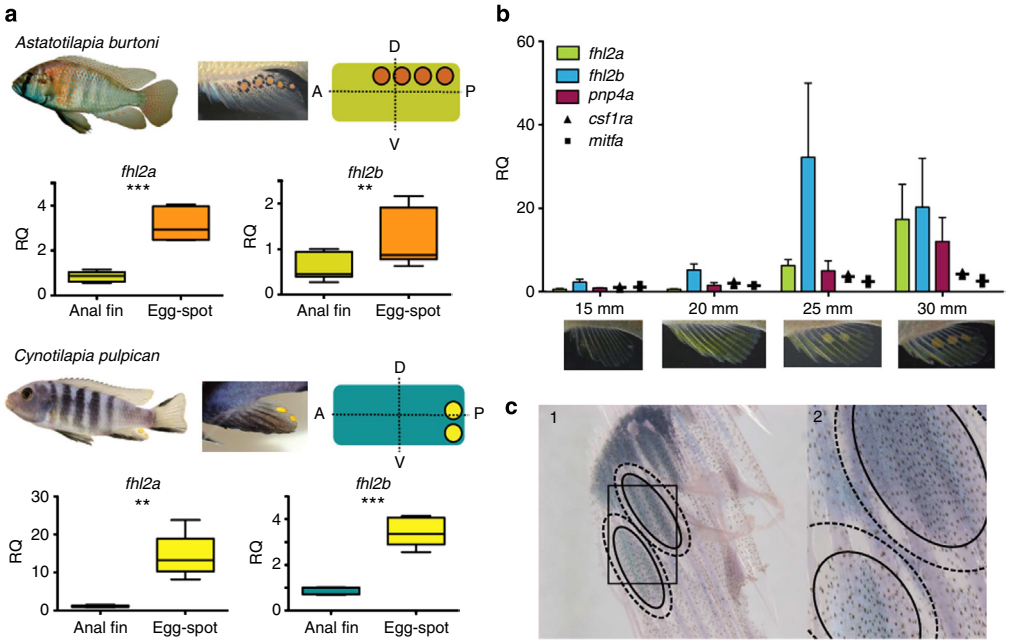


Figure 2 | The role of *fhl2a* and *fhl2b* in egg-spot formation. (**a**) qPCR experiments reveal that both genes are overexpressed in egg-spot compared with adjacent anal fin tissue in the haplochromine cichlids *A. burtoni* and *C. pulpican* (**$P < 0.01$; ***$P < 0.001$; RQ, relative quantity). Images of male fishes of the two species, their anal fins and a scheme showing the distribution of egg-spots are provided. (**b**) Expression profiles of *fhl2a* and *fhl2b* during the ontogenetic development of egg-spots in *A. burtoni* (note that egg-spots are absent in juveniles and only form when males become sexually mature; see ref. 22 for further details). The values on the *x* axis represent fish standard length in millimetres (three replicates per developmental stage were used). The error bars represent the s.e.m. *fhl2b* shows the largest increase in expression overall and its expression profile mimics the formation of egg-spots. Three other pigmentation genes (*pnp4a, csf1ra* and *mitfa*) were included for comparative reasons. *csf1ra* and *mitfa* show a much smaller increase in gene expression during egg-spot development than *fhl2a* and especially *fhl2b*, while *pnp4a* shows a constant increase in gene expression throughout the development of egg-spots. (**c**) RNA *in situ* hybridization experiments revealed that both *fhl2* paralogs (results only shown for *fhl2b*) are primarily expressed in the colourful inner circle of haplochromine egg-spots (defined by the solid line) and not in the transparent outer ring (defined by the dashed line). Expression was also observed in the proximal fin region, which also contains pigment cells. Panel 2 is a close-up from the region defined by the square in panel 1.

domain of both *fhl2a* and *fhl2b* matches the conspicuously coloured inner circle of egg-spots with RNA *in situ* hybridization (see Fig. 2c for results on *fhl2b*).

***fhl2a* and *fhl2b* evolved under purifying selection**. In general, phenotypic differences can arise via mutations affecting the function of proteins or via changes in gene regulation[5]. Therefore, we examined coding sequence evolution in the two *fhl2* paralogs to test for positive selection and potential change of function in a phylogenetically representative set of 26 East African cichlids. We found that the two *fhl2* genes are highly conserved in cichlids, with few amino-acid differences between species and an average genetic divergence (0.4% in *fhl2a* and 0.7% in *fhl2b*) that lies below the transcriptome-wide average of 0.95% (ref. 23). None of the observed amino-acid changes was correlated with the egg-spot phenotype (Supplementary Table 7).

Greater functional specialization of *fhl2b* in haplochromines. Usually, after a gene duplication event, the duplicates go through a period of relaxed selection, during which one of the two copies can diversify and acquire new functions[24]. We found that the b-copy of *fhl2* shows an elevated rate of molecular evolution compared with its paralog (*fhl2a*), which more closely resembles the ancestral sequence (Fig. 3a). An additional series of qPCR experiments in 12 tissues revealed that, in cichlids, *fhl2a* is primarily expressed in heart, bony structures and muscles, whereas *fhl2b* is highly expressed in the eye, and further in skin and the egg-spots of haplochromines (Fig. 3b,c). This is different to the gene expression profiles in medaka, where both duplicates are highly expressed in heart, skin and eye tissues; and in zebrafish, where the two paralogs are primarily expressed in heart, eye and (pharyngeal) jaw tissues, with *fhl2a* showing rather low levels of gene expression (Supplementary Figs 3 and 4). When compared with the other teleost fishes examined here, our results suggest that the haplochromine *fhl2a* retained most of the previously described functions, whereas the more rapidly evolving *fhl2b* obtained new expression patterns. Together, the gene expression profile and the pattern of sequence evolution make *fhl2b* a prime candidate gene for the morphogenesis of haplochromine egg-spots.

***fhl2b* shows an AFC-SINE insertion in species with egg-spot**. Since there were no changes in the coding regions of *fhl2a* and *fhl2b* that are specific to the egg-spot bearing haplochromines, we shifted our focus towards the analysis of putative regulatory elements, exploring the recently available genomes of five East African cichlids (including the egg-spot bearing haplochromines *A. burtoni*, *Pu. nyererei*, *Metriaclima zebra* and the egg-spot-less non-haplochromines *Neolamprologus brichardi* and *Oreochromis niloticus*). The non-coding region of *fhl2a* shows homology with other teleosts (*Oryzias latipes*, *Takifugu rubripes*, *Tetraodon nigroviridis* and *Gasterosteus aculeatus*) and we identified four conserved non-coding elements (CNEs) in all species examined (Supplementary Fig. 5a). These CNEs might thus represent conserved regulatory regions responsible for ancestral conserved functions of *fhl2a* in teleosts. We might be missing cichlid-specific regulatory regions in important upstream regions although, as our capacity to detect lineage-specific enhancers is limited owing to the small sample size for each lineage and the high background conservation level present in cichlids.

Concerning *fhl2b*, we did not find any CNE that is shared by cichlids and other teleosts (Supplementary Fig. 5b). Strikingly, however, we found a major difference that is shared by the three egg-spot bearing haplochromines: the presence of a transposable element upstream of *fhl2b*. Specifically, we identified a SINE belonging to the cichlid-specific AFC-SINEs (African cichlid family of SINEs[25]), which inserted ~800-bp upstream of the transcriptional start site of *fhl2b* (Supplementary Fig. 6). To confirm that this insertion is associated with the egg-spot phenotype, we sequenced the upstream region of *fhl2b* in 19 cichlid species. The insertion was indeed present in nine additional, egg-spot bearing haplochromine species, yet absent in all 10 non-haplochromines examined (Supplementary Table 8). Importantly, we found that one haplochromine species lacks the AFC-SINE element, namely *P. philander*. This species belongs to one of the basal lineage of haplochromines (Fig. 1a), which is characterized by the absence of egg-spots (Fig. 1b). This suggests that the AFC-SINE upstream of *fhl2b* is not characteristic to the entire haplochromine clade, but to those that feature egg-spots, thus linking the SINE insertion to the origin of this evolutionary innovation.

Haplochromine *fhl2b* regulatory region drives iridophore expression. A long-standing hypothesis proposes that ubiquitous genomic repeat elements are potential regulators of transcription, and could thereby generate evolutionary variations and novelties[26,27]. SINEs are known for their capability of 'transcriptional rewiring', that is, to change the expression patterns of genes by bringing along new regulatory sequences when inserted in close proximity to a gene's transcriptional initiation site[7,28]. In order to test whether the insertion of an AFC-SINE close to *fhl2b* functions as an enhancer of gene expression, we aimed for a functional experiment. We were particularly interested to find out whether there were changes in enhancer activity between AFC-SINE-positive haplochromines and other cichlids lacking both the insertion and the egg-spot phenotype. To this end, we designed reporter constructs containing the upstream region of *fhl2b* (~2 kb upstream to intron 1) of three cichlid species linked to the coding region of green fluorescent protein (GFP), and injected these constructs into zebrafish (*Danio rerio*) embryos to generate transgenic lines. We switched to the zebrafish system here, as no functioning transgenesis was available for haplochromine cichlids at the time the study was performed (owing to the small number of eggs per clutch associated with the characteristic female-mouthbrooding behaviour). The three constructs were derived from *A. burtoni* (haplochromine with egg-spots, AFC-SINE$^+$), *P. philander* (haplochromine without egg-spots, AFC-SINE$^-$) and *N. sexfasciatus* (lamprologine, AFC-SINE$^-$), respectively (Fig. 4a).

We were able to produce stable transgenic zebrafish lines for each of the three constructs to examine the expression of GFP. Importantly, we found striking differences in expression between the *A. burtoni* construct and the two constructs lacking the AFC-SINE. Of the three reporter lines, only the AFC-SINE$^+$ showed GFP expression in iridophores, a silvery-reflective type of pigment cells (Fig. 4b,c and Supplementary Fig. 7). This experiment demonstrates the presence of novel enhancer activities in the regulatory region of *fhl2b* in derived haplochromines and strongly suggests that these came along with the SINE insertion.

Iridophores and egg-spot development. The egg-spot phenotype has previously been associated with pigment cells containing pteridines (xanthophores)[16,22], whereas our new results indicate an auxiliary role of iridophores in egg-spot formation. We thus re-evaluated the adult egg-spot phenotype by removing the pteridine pigments of the xanthophores (Fig. 4e). We indeed found that *A. burtoni* egg-spots show a high density of iridophores, which is further corroborated by the increase in gene expression of the iridophore marker *pnp4a* during egg-spot

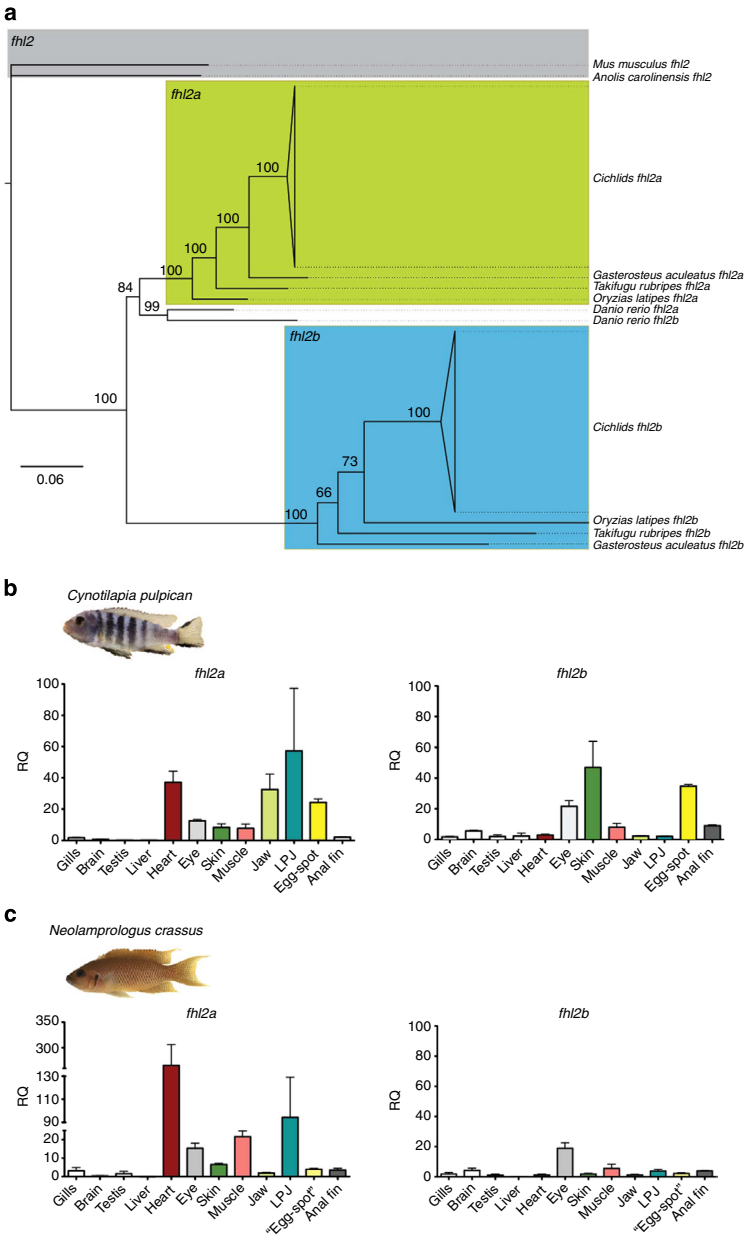


Figure 3 | Gene tree of the two *fhl2* paralogs and expression profiling in East African cichlid fishes. (**a**) Bayesian inference phylogeny of the orthology and paralogy relationships between cichlids, other teleosts (*O. latipes*, *D. rerio*, *Ta. rubripes* and *G. aculeatus*) and tetrapods (*Anolis carolinensis* and *Mus musculus*) *fhl2* sequences. This gene tree is important for generating functional hypotheses about both duplicates, and to infer the ancestral state of the *fhl2* gene before duplication. Our phylogeny indicates that *fhl2a* is more similar to the ancestral state, while *fhl2b* is apparently evolving faster in teleosts. Values at the tree nodes represent posterior probabilities. In Supplementary Fig. 2, we present a synteny analysis supporting the origin of teleost *fhl2* duplicates in the teleost genome duplication. (**b**) Relative quantity (RQ) of *fhl2a* and *fhl2b* gene expression in 12 tissues (three replicates per tissue) in *C. pulpican*, an egg-spot bearing haplochromine from Lake Malawi. The error bars represent the s.e.m. (**c**) RQ of *fhl2a* and *fhl2b* gene expression in 12 tissues in *N. crassus*, a substrate spawning lamprologine that has no egg-spots. In both species, gill tissue was used as reference; in *N. crassus*, 'egg-spots' corresponds to the fin region where haplochromines would show the egg-spot trait. In *C. pulpican* (**b**), *fhl2a* is highly expressed in heart, in pigmented tissues (eye, skin and egg-spot) and in craniofacial traits (oral jaw and lower pharyngeal jaw); *fhl2b* is mainly expressed in the pigmented tissues. *N. crassus* (**c**) shows a similar expression patterns for *fhl2a* and *fhl2b*, with the difference that *fhl2a* does not show high expression levels in jaw tissues, and *fhl2b* is not highly expressed in skin and fin tissue. These results suggest that *fhl2b* shows a higher functional specialization, and that it might be involved in the morphogenesis of sexually dimorphic traits such as pigmented traits including egg-spots. LPJ, lower pharyngeal jaw bone.

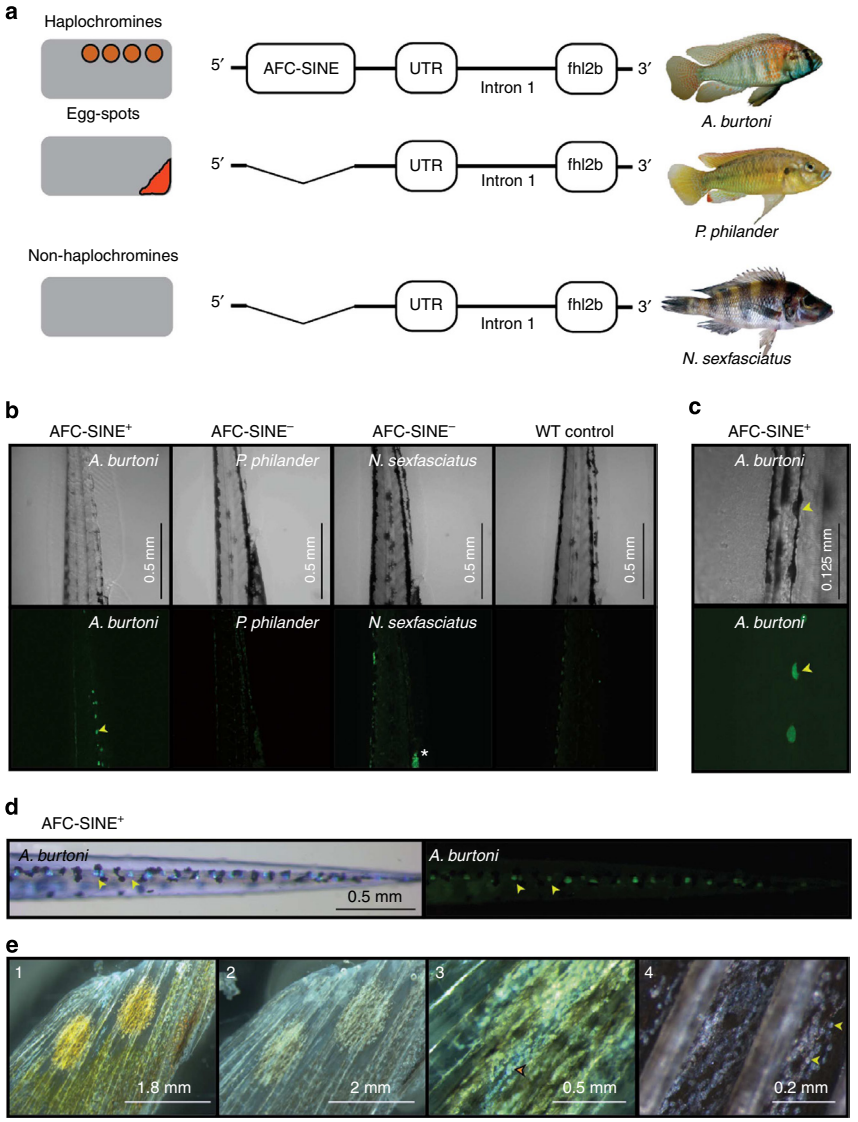


Figure 4 | The molecular basis of egg-spot formation. (**a**) The egg-spot bearing haplochromines feature an AFC-SINE insertion in close proximity to the transcriptional start site of *fhl2b*, which is absent in the ancestral and egg-spot-less genus *Pseudocrenilabrus* and in all non-haplochromines. The sequences from the three species shown here were the ones used to engineer the reporter constructs, where the *fhl2b* coding sequence was substituted by GFP. (**b**) In transgenic zebrafish, only the AFC-SINE$^+$ construct showed GFP expression in the iridophores, a type of pigment cells (one of them is indicated by a yellow arrow). The upper panel depicts bright-field images of 3-day-old zebrafish embryo trunks; the lower panel shows the respective embryos under ultraviolet light. The green signal in the AFC-SINE negative *N. sexfasciatus* line (marked with an asterisk) is auto-fluorescence from the yolk extension. (**c**) Higher magnification image from *A. burtoni* AFC-SINE$^+$ reporter construct driving GFP expression in the iridophores. Orientation in **b**,**c**: bottom: anterior, top: posterior. (**d**) Top-down view of a trunk of a 3-day-old AFC-SINE-positive zebrafish embryo. The left panel depicts a bright-field image where the iridophores of the dorsal stripe are illuminated by the incident light (yellow arrows). The right panel depicts GFP expression of the same embryo. The GFP signal co-localizes with iridophores. (**e**) Cellular basis of egg-spots: this series of images shows that egg-spots are made up of xanthophores, iridophores and scattered melanophores. Image 1 shows an *A. burtoni* fin with two egg-spots. Image 2 shows the same fin without pteridine pigments (xanthophores are not visible anymore). Images 3 and 4 are higher magnification images of the egg-spots without pteridine under slightly different light conditions confirming that egg-spots have a high density of iridophores (examples of this cell type are highlighted with arrows). UTR, untranslated region.

formation (Fig. 2b). With the exception of the proximal region of the anal fin, the number of iridophores is greatly reduced in the fin tissue surrounding egg-spots (Supplementary Fig. 8a). Interestingly, this proximal region is the only area of the anal fin besides the egg-spots where we observed *fhl2* expression with RNA *in situ* hybridization (see Fig. 2c for *fhl2b*), once more linking *fhl2* expression with iridophores (and less so with xanthophores, which are very rare in this region). In the non-haplochromine *N. crassus*, which features a yellow anal fin pattern containing xanthophores, we did not find iridophores in

the xanthophore-rich region (Supplementary Fig. 9), suggesting that the xanthophore/iridophore pattern is unique to haplochromine egg-spots. Importantly, we also observed that iridophores appear early in the newly forming egg-spot of haplochromines, that is, before the first xanthophores start to aggregate (Supplementary Fig. 8b).

In zebrafish, stripe development is initiated by iridophores, which serve as morphological landmarks for stripe orientation in that they attract further pigment cells such as xanthophores by expressing the *csf1* ligand gene[29,30]. Interestingly, it has previously been shown that a gene encoding a Csf1 receptor known for its role in xanthophore development in zebrafish, *csf1ra*, is expressed in haplochromine egg-spots[16]. We thus examined the expression of the ligand *csf1b* and show that its relative level of gene expression doubles during egg-spot development, and that this increase coincides with the emergence of the phenotype (Supplementary Fig. 10). This leads us to suggest that a similar pigment cell type interaction mechanism might be involved in egg-spot patterning as the one described for zebrafish[29,30]. The specific mode of action of fin patterning in haplochromine cichlids, and how Fhl2b interacts with the Csf1/Csf1r system, remains to be studied in the future.

Contribution of *fhl2a* in egg-spot formation. The role of the more conserved and functionally constrained a-paralog of *fhl2* in egg-spot development cannot be dismissed. Its temporally shifted increase in gene expression compared with *fhl2b* (Fig. 2b) suggests that *fhl2a* most likely acts as a more downstream factor involved in pigment pattern formation. We were nevertheless interested in uncovering the regulatory region responsive for this expression pattern. The first intron of *fhl2a* shows two CNEs that are common across percomorph fish (Supplementary Fig. 5). Using the same strategy as described above, we generated a transgenic zebrafish line containing exon 1 and intron 1 of *A. burtoni* linked to GFP. This construct drove expression in heart in zebrafish embryos, which is consistent with the reported function of *fhl2a* in tetrapods[20], whereas there was no indication of a pigment cell related function for this reporter construct (Supplementary Fig. 7e). An alignment between the genomic regions of the two *fhl2* paralogs shows that there were no CNEs in common and generally very little homology between them, suggesting that the regulation of the expression of *fhl2a* in egg-spots might proceed in a different way (Supplementary Fig. 11).

Discussion

In this study, we were interested in the genetic and developmental basis of egg-spots, an evolutionary innovation of the most species-rich group of cichlids, the haplochromines, where these conspicuous colour markings on the anal fins of males play an important role in mating[11–13] (Fig. 1).

We first performed a comparative RNAseq experiment that led to the identification of two novel candidate pigmentation genes, the a- and b-paralogs of the four and a half LIM domain protein 2 (*fhl2*) gene. We then confirmed, with qPCR and RNA *in situ* hybridization, that the expression domain of both duplicates indeed matches the conspicuously coloured inner circle of egg-spots (Fig. 2). Especially the more rapidly evolving b-copy of *fhl2* emerged as strong candidate gene for egg-spot development, as its expression profile mimics the formation of egg-spots (Figs 2b and 3). Interestingly, we found that the egg-spot bearing haplochromines, but not other cichlids, feature a transposable element in the *cis*-regulatory region of *fhl2b*. Finally, making use of transgenic zebrafish, we could show that a *cis*-regulatory change in *fhl2b* in the ancestor of the egg-spot bearing haplochromine cichlids (most likely in the form of the AFC-SINE insertion) resulted in a gain of expression in iridophores, a special type of pigment cells found in egg-spots (Fig. 4). This in turn might have led to changes in iridophore cell behaviour and to novel interactions with pigmentation genes (*csf1b*, *csf1ra* and *pnp4a*), thereby contributing to the formation of egg-spots on male anal fins. The specific mode of action of the SINE insertion, and how the *fhl2b* locus interacts with these other pigmentation genes remains elusive at present. Addressing these questions would require functional studies in haplochromines, which are, however, hampered by the specific mechanisms involved in the trait complex of interest (mouthbrooding makes it notoriously difficult to obtain enough eggs—in a controlled manner—to make such experiments feasible).

Our results are also suggestive of an important role of the a-copy of *fhl2* in cichlid evolution. With our qPCR experiments, we provide strong evidence that *fhl2a* is involved in jaw tissue in zebrafish (Supplementary Fig. 3) and, importantly, in the pharyngeal jaw apparatus of cichlids (Fig. 3b,c), another putative evolutionary innovation of this group. The pharyngeal jaw apparatus is a second set of jaws in the pharynx of cichlids that is functionally decoupled from the oral jaws and primarily used to process food[11,12,15]. Interestingly, *fhl2a* has previously been implicated in the evolution of fleshy lips in cichlids[31], which is yet another ecologically relevant trait. From a developmental perspective, the main tissues underlying these traits—the craniofacial cartilage (the jaw apparatus) and pigment cells (egg-spots)—have the same origin, the neural crest, which itself is considered an evolutionary key innovation of vertebrates[32]. It thus seems that the function of *fhl2* in cichlids may have been split into (a) an ecologically important, that is, naturally selected, scope of duties, and (b) a role in colouration and pigmentation more likely to be targeted by sexual selection.

Taken together, our study permits us to propose the following hypothesis for the origin of cichlid egg-spots: In one of the early, already female-mouthbrooding, haplochromines the insertion of a transposable element of the AFC-SINE family in the *cis*-regulatory region of *fhl2b*, and its associated recruitment to the iridophore pigment cell pathway, mediated the evolution of egg-spots on the anal fins—possibly from the so-called *perfleckmuster* common to many cichlids[16]. The conspicuous anal fin spots were fancied by haplochromine females, which—just like many other cichlids and also the ancestral and egg-spot-less haplochromine genus *Pseudocrenilabrus*—have an innate bias for yellow/orange/red spots that resemble carotenoid-rich prey items[33], leading to the fixation of the novel trait. In today's haplochromines, egg-spots seem to have a much broader range of functions related to sexual selection[34].

Most of the currently studied evolutionary innovations comprise relatively ancient traits (for example, flowers, feathers, tetrapod limb, insect wings and mammalian placenta), thereby making it difficult to scrutinize their genetic and developmental basis. Here we explored a recently evolved novelty, the anal fin egg-spots of male haplochromine cichlids. We uncovered a regulatory change in close proximity to the transcriptional start site of a novel iridophore gene that likely contributes to the molecular basis of the origin of egg-spots in the most rapidly diversifying clade of vertebrates. This, once more, illustrates the importance of changes in *cis*-regulatory regions in morphological evolution[2].

Methods

Samples. Laboratory strains of *A. burtoni*, *C. pulpican*, *Astatoreochromis alluaudi*, *Pu. nyererei*, *Labidochromis caeruleus*, *Pseudotropheus elegans* and *N. crassus* were kept at the University of Basel (Switzerland) under standard conditions (12 h light/12 h dark; 26 °C, pH 7). Before dissection, all specimens were euthanized with MS 222 (Sigma-Aldrich, USA) following an approved procedure (permit no. 2317

issued by the cantonal veterinary office Basel). Individuals of all other specimens were collected in the southern region of Lake Tanganyika (Zambia) under the permission of the Lake Tanganyika Unit, Department of Fisheries, Republic of Zambia, and processed in the field following our standard operating procedure[15]. Tissues for RNA extraction were stored in RNAlater (Ambion, USA), and tissues for genomic DNA extraction were stored in ethanol and shipped to the University of Basel.

RNA and DNA extractions. Isolation of RNA was performed according to the TRIzol protocol (Invitrogen, USA) after incubating the dissected tissues in 750 µl of TRIzol at 4 °C overnight or, alternatively, for 8–16 h (in order to increase the RNA yield after long-term storage). The tissues were then homogenized with a Bead-Beater (FastPrep-24; MP Biomedicals, France). Subsequent DNase treatment was performed with DNA-Free kit (Ambion). RNA quantity and quality was determined with a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA). cDNA was produced using the High Capacity RNA-to-cDNA kit (Applied Biosystems, USA). Genomic DNA was extracted using a high salt extraction method (modified from ref. 35).

Phylogenetic analyses. DNA extraction of 18 specimens of East African cichlid fishes was conducted as described above. For the amplification of nine nuclear markers (*rag*, *gapdhs*, *s7*, *bmp4*, *ednrb1*, *mitfa*, *tyr*, *hag* and *csfr1*), we used the primer sets published in ref. 36. The sequences of *M. zebra, O. niloticus* and *N. brichardi* were extracted from the respective genome assemblies (http://www.broadinstitute.org/models/tilapia). The data for *Astatoreochromis alluaudi, Thoracochromis brauschi* and *Serranochromis macrocephalus* were collected with Sanger sequencing following the method described in ref. 36, all other data were generated by amplicon sequencing with 454 GS FLX system at Microsynth, Switzerland, following the manufacturer's protocols[37,38]. Sequences were quality filtered using PRINSEQ (length: 150 bp minimum; low quality: mean ≥15; read duplicates)[39] and assembled with Burrows-Wheeler Aligner, Smith-Waterman alignment (BWA-SW) followed by visual inspection and consensus sequence generation in Geneious 6.1.6 (ref. 40). As a tenth marker, we included mitochondrial NADH dehydrogenase subunit 2 (ND2) sequences available on GenBank (see Supplementary Table 1 for accession numbers). Since the *ednrb1* gene sequence is not available in the *N. brichardi* genome assembly, we used the gene sequence from its sister species, *N. pulcher*, instead.

Sequences were aligned with MAFFT[41] and the most appropriate substitution model of molecular evolution for each marker was determined with JMODELTEST v2.1.3 (ref. 42) and BIC[43]. The partitioned data set (5,051 bp) was then subjected to phylogenetic analyses in MRBAYES v3.2.1 (ref. 44) and GARLI v2.0 (ref. 45). MRBAYES was run for 10,000,000 generations with two runs and four chains in parallel and a burn-in of 25%, GARLI was run 50 times followed by a bootstrap analysis with 500 replicates. SUMTREES v3.3.1 of the DENDROPY package v3.12.0 (ref. 46) was used to summarize over the replicates and to map bootstrap values to the ML topology.

Differential gene expression analysis using RNAseq. We used a transcriptomic approach (RNAseq) to identify genes differentially expressed between male and female anal fins of *A. burtoni*. Library construction and sequencing of RNA extracted from three male and three female anal fins (at the developmental stage of 30 mm; Fig. 2) was performed at the Department of Biosystems Science and Engineering, University of Basel and ETH Zurich. The samples were sequenced on an Illumina Genome Analyzer IIx. Each sample was sequenced in one lane and with a read length of 76 bp.

The reads were then aligned to an embryonic *A. burtoni* reference transcriptome assembled by Broad Institute (http://www.broadinstitute.org/models/tilapia). This transcriptome is not annotated and each transcript has a nomenclature where the first term codes for the parent contig and the third term codes for alternatively spliced transcripts (CompX_cX_seqX). The reference transcriptome was indexed using NOVOINDEX (www.novocraft.com) with default parameters. Using NOVOALIGN (www.novocraft.com), the RNAseq reads were mapped against the reference transcriptome with a maximum alignment (*t*) score of 30, a minimum of good-quality base pair per read (l) of 25 and a successive trimming factor (s) of 5. Reads that did not match these criteria were discarded. Since the reference transcriptome has multiple transcripts/isoforms belonging to the same gene, all read alignment locations were reported (rALL). The mapping results were reported (o) in SAM format. The output SAM file was then transformed into BAM format, sorted, indexed and converted to count files (number of reads per transcript) using SAMTOOLS version 0.1.18 (ref. 47). The count files were subsequently concatenated into a single data set—count table—and analysed with the R package EDGER[48] in order to test for significant differences in gene expression between male and female anal fins. The 10 most differentially expressed transcripts were identified by BLASTx[49] against GenBank's non-redundant database (Supplementary Table 2).

We selected two genes out of this list for in-depth analyses—*fhl2a* and *fhl2b*—for the following three reasons: (i) *fhl2b* was the gene showing the highest difference in expression between male and female anal fins; (ii) the difference in gene expression in its paralog, *fhl2a*, was also significantly high; and (iii) the functional repertoire of the Fhl2 protein family indicates that these might be strong candidates for the morphogenesis of a secondary male colour trait.

Differential gene expression analysis using qPCR. The expression patterns of *fhl2a* and *fhl2b* were further characterized by means of qPCR in three species, *A. burtoni*, *C. pulpican* and *N. crassus*. The comparative cycle threshold method[50] was used to calculate differences in expression between the different samples using the ribosomal protein L7 (*rpl7*) and the ribosomal protein SA3 (*rpsa3*) as endogenous controls. All reactions had a final cDNA concentration of 1 ng µl^{-1} and a primer concentration of 200 mM. The reactions were run on a StepOnePlusTM Real-Time PCR system (Applied Biosystems) using the SYBR Green master mix (Roche, Switzerland) with an annealing temperature of 58 °C and following the manufacturer's protocols. Primers were designed with the software GenScript Real-Time PCR (Taqman) Primer Design available at https://www.genscript.com/ssl-bin/app/primer. All primers were designed to span over exons to avoid gDNA contamination (see Supplementary Table 3 for details). Primer efficiencies of the experimental primers (*fhl2a* and *fhl2b*) were comparable to the efficiency of the endogenous controls *rpl7* and *rpsa3*.

We conducted the following experiments: qPCR experiment 1: Egg-spots were separated from the anal fin tissue in six male *A. burtoni* and five male *C. pulpican*. Relative quantity values were calculated for each sample, and the differential expression between anal fin (reference) and egg-spot tissue was analysed with a paired *t*-test using GraphPad Prism version 5.0a for Mac OS X (www.graphpad.com). qPCR experiment 2: *fhl2a, fhl2b, csf1ra, mitfa, pnp4a* and *csf1b* expression was measured in RNA extracted from *A. burtoni* fins at four different developmental stages[22]. Here, *csf1ra* was included as xanthophore marker[16], *mitfa* and *pnp4a* as melanophore and iridophore markers[51], respectively, and *csf1b* because of its role in pigment pattern organization in zebrafish[29,30]. We used three biological replicates for each developmental stage, and each replicate consisted of a sample pool of three fins, except for the youngest stage at 15 mm, where we pooled five fins. The first developmental stage was used as reference tissue. qPCR experiment 3: *fhl2a* and *fhl2b* expression was measured in RNA extracted from different tissues from three males from *C. pulpican* and *N. crassus* (gills, liver, testis, brain, heart, eye, skin, muscle, oral jaw, pharyngeal jaw and egg-spot). Although *N. crassus* does not have egg-spots, we separated its anal fin into an area corresponding to egg-spots in haplochromines and a section corresponding to anal fin tissue (the 'egg-spot' region was defined according to the egg-spot positioning in *A. burtoni*). Expression was compared among tissues for each species using gills as reference tissue. The same experiment was performed for *D. rerio* and *O. latipes* (two teleost outgroups), using *ef1a* and *rpl13a* (ref. 52), as well as *rpl7* and *18sRNA* (ref. 53) as endogenous controls, respectively.

Cloning of *fhl2a* and *fhl2b* and RNA *in situ* hybridization. *A. burtoni fhl2a* and *fhl2b* coding fragments were amplified by PCR (for primer information, see Supplementary Table 3) using Phusion Master Mix with High Fidelity buffer (New England BioLabs, USA) following the manufacturer's guidelines. These fragments were cloned into pCR4-TOPO TA vector using the TOPO TA cloning kit (Invitrogen). Plasmid extractions were done with GenElute Plasmid Miniprep Kit (Sigma-Aldrich). RNA probes were synthetized with the DIG RNA labelling kit (SP6/T7) (Roche). The insertion and direction of the fragments was confirmed by Sanger sequencing using M13 primers (available with the cloning kit) and BigDye terminator reaction chemistry (Applied Biosystems) on an AB3130*xl* Genetic Analyzer (Applied Biosystems). *In situ* hybridization was performed in 12 fins from *A. burtoni* males, six for *fhl2a* and six for *fhl2b*. The protocol was executed as described in ref. 16, except for an intermediate proteinase *K* treatment (20 min at a final concentration of 15 µg ml^{-1}) and for the hybridization temperature (65 °C).

Synteny analysis of teleost *fhl2* paralogs. The Synteny Database (http://syntenydb.uoregon.edu[54]) was used to generate dotplots of the human *FHL2* gene (ENSG00000115641) region on chromosome Hsa2 and the genomes of medaka (Supplementary Fig. 2a) and zebrafish (Supplementary Fig. 2b). Double-conserved synteny between the human *FHL2* gene and the *fhl2a* and *fhl2b* paralogons in teleost genomes provide evidence that the teleost *fhl2* paralogs were generated during the teleost genome duplication.

***fhl2a/fhl2b* coding region sequencing and analysis.** We then used cDNA pools extracted from anal fin tissue to amplify and sequence the coding region of *fhl2a* and *fhl2b* in a phylogenetically representative set of 26 cichlid species (21 Tanganyikan species, three species from Lake Malawi and two species from the Lake Victoria basin). This taxon sampling included 14 species belonging to the haplochromines and 12 species belonging to other East African cichlid tribes not featuring the egg-spot trait (Supplementary Table 4). *fhl2a* and *fhl2b* coding regions were fully sequenced (from start to stop codon) in five individuals per species in order to evaluate the rate of molecular evolution among cichlids. For PCR amplification, we used Phusion Master Mix and cichlid-specific primers (for primer information, see Supplementary Table 3) designed with Primer3 (ref. 55). PCR products were visualized with electrophoresis in a 1.5% agarose gel using GelRed (Biotium, USA). In cases where multiple bands were present, we purified the correct size fragment from the gel using the GenElute Gel Extraction Kit (Sigma-

Aldrich). PCR products were enzymatically cleaned with ExoSAP-IT (Affymetrix, USA) and sequenced with BigDye 3.1 Ready reaction mix (Applied Biosystems)—after BigDye XTerminator purificaton (Applied Biosystems)—on an AB3130*xl* Genetic Analyzer. Sequences were corrected, trimmed and aligned manually in CODONCODE ALIGNER (CodonCode Corporation).

***fhl2* phylogenetic analysis.** *fhl2a* and *fhl2b* sequences from non-cichlid teleosts and *fhl2* sequences from tetrapods were retrieved from ENSEMBL[56] (species names, gene names and accession numbers are available in Supplementary Table 5). We then constructed gene trees based on these sequences and on a subset of the cichlid sequences obtained in the previous step (information available in Supplementary Table 4) in order to confirm the orthologous and paralogous relationships of both duplicates. Sequences were aligned with CLUSTALW2 (ref. 57) using default parameters. The most appropriate model of sequence evolution was determined with JMODELTEST as described above. Phylogenetic analyses were performed with MRBAYES (1 million generations; 25% burn-in).

Tests for positive selection in *fhl2a* and *fhl2b*. Using PAUP* 4.0b10 (ref. 58), we first compiled a maximum likelihood tree based on the mitochondrial *ND2* gene, including all species used for the positive selection analyses (see Supplementary Table 6 for species and GenBank accession numbers). We used the GTR + Γ model with base frequencies and substitution rate matrix estimated from the data (as suggested by JMODELTEST[42]). We then ran CODEML implemented in PAML version 4.4b to test for branch-specific adaptive evolution in *fhl2a* and *fhl2b* applying the branch-site model (free-ratios model with ω allowed to vary)[59,60]. The branch comparisons and results are shown in Supplementary Table 7.

Identification of CNEs. We then made use of the five available cichlid genomes[61] to identify CNEs that could explain the difference in expression of *fhl2a* and *fhl2b* between haplochromines and non-haplochromines (note that there are three haplochromine genomes available: *A. burtoni, Pu. nyererei, M. zebra*; and two genomes belonging to more ancestral cichlid lineages: *N. brichardi* and *Or. niloticus*). For this analysis, we also included the respective genomic regions of four other teleost species (*O. latipes*, *Ta. rubripes*, *Te. nigroviridis* and *G. aculeatus*). More specifically, we extracted the genomic scaffolds containing *fhl2a* and *fhl2b* from the available cichlid genomes using BLAST v. 2.2.25 and the BIOCONDUCTOR R package BIOSTRINGS[62] to extract 5–6 kb of sequence containing *fhl2a* and *fhl2b* from these scaffolds.

Comparative analyses of the *fhl2a* and *fhl2b* genomic regions were done with MVISTA (genome.lbl.gov/vista)[63] using the LAGAN alignment tool[64]; *A. burtoni* was used as a reference for the alignment. We applied the repeat masking option with *Ta. rubripes* (Fugu) as reference. CNEs were defined as any non-coding section longer than 100 bp that showed at least 70% sequence identity with *A. burtoni*.

Sequencing of the upstream region of *fhl2b*. In order to confirm whether the AFC-SINE insertion was specific to egg-spot bearing haplochromines, we amplified the genomic region upstream of the *fhl2b* open reading frame in 19 additional cichlid species (10 haplochromines and 9 non-haplochromines). PCR amplification was performed as described above. For sequencing, we used four different primers, the two used in the amplification reaction and two internal primers, one haplochromine specific and another non-haplochromine specific. For detailed information about species and primers, see Supplementary Table 8.

Alignment of AFC-SINES from the *A. burtoni* genome. SINE elements were identified using the SINE insertion sequence 5′ of the *fhl2b* gene of *A. burtoni* as query in a local BLASTn search[49] with default settings against the *A. burtoni* reference genome. Blast hits were retrieved using custom scripts and extended to a region of 200-bp upstream and downstream of the identified sequence. Sequences were aligned using MAFFT v. 6 (ref. 41) with default settings and allowing for adjustment of sequence direction according to the reference sequence. The alignment was loaded into CODONCODE ALIGNER for manual correction and end trimming. Sequences shorter than 50 bp were excluded from the alignment. The final alignment contained 407 sequences that were used to build the *A. burtoni* SINE consensus sequence using the consensus method implemented in CODONCODE ALIGNER with a percentage-based consensus and a cutoff of 25%. The AFC-SINE element in the *fhl2b* promoter region was compared with the consensus sequence and available full-length AFC-SINE elements of cichlids in order to determine whether it was an insertion or deletion in haplochromines (Supplementary Table 8).

Characterization of *fhl2b* upstream genomic region in cichlids. The *fhl2b* genomic regions of the five cichlid genomes (*A. burtoni*, *M. zebra*, *Pu. nyererei*, *N. brichardi*, and *O. niloticus*) were loaded into CODONCODE ALIGNER and assembled (large gap alignments settings, identity cutoff 70%). Assemblies were manually corrected. Transposable element sequences were identified using the Repeat Masking function of REPBASE UNIT (http://www.girinst.org/censor/index.php) against all sequence sources and the bl2seq function of BLASTn[49]. Supplementary Fig. 6 shows a scheme of the transposable element composition of this genomic region in several cichlid species.

CNEs construct cloning and injection in zebrafish. We designed three genetic constructs containing the AFC-SINE and intron 1 of *fhl2b* of three cichlid species (*A. burtoni*, *P. philander* and *N. sexfasciatus*) (Fig. 4) and one containing the 5′-untranslated region, exon 1 and intron 1 of *A. burtoni fhl2a*. The three fragments were amplified with PCR as described above (see Supplementary Table 3 for primer information). All fragments were cloned into a pCR8/GW/TOPO vector (Invitrogen) following the manufacturer's specifications. Sequence identity and direction of fragment insertion were confirmed via Sanger sequencing (as described above) using M13 primers. All plasmid extractions were performed with GenElute Plasmid Miniprep Kit (Sigma-Aldrich). We then recombined these fragments into the Zebrafish Enhancer Detection ZED vector[65] following the protocol specified in ref. 66. Recombination into the ZED plasmid was performed taking into consideration the original orientation of the *fhl2b* genomic region. The resulting ZED plasmids were then purified with the DNA clean and concentrator − 5 Kit (Zymo Research, USA). Injections were performed with 1 nl into one-/two-cell stage zebrafish (*D. rerio*) embryos (*A. burtoni* construct was injected in wild-type strains AB and ABxEK, *P. philander* and *N. sexfasciatus* constructs were injected in wild-type strain ABxEK) with 25 ng µl^{-1} plasmid and 35 ng µl^{-1} Tol2 transposase mRNA. By outcrossing to wild-type zebrafish, we created five F2 stable transgenic lines for the *A. burtoni* construct, two F1 stable transgenic lines for the *P. philander* construct, and finally one F1 stable transgenic line for the *N. sexfasciatus* construct. Fish were raised and kept according to standard procedures[67]. Zebrafish were imaged using a Leica point scanning confocal microscope SP5-II-matrix and Zeiss LSM5 Pascal confocal microscope.

Fixation and dehydration of cichlid fins. In order to determine the pigment cell composition of egg-spots (and especially whether they contain iridophores in addition to xanthophores), we dissected *A. burtoni* anal fins. To better understand the morphological differences between non-haplochromine and haplochromine fins, we further dissected three *N. crassus* anal fins. To visualize iridophores, we removed the pteridine pigments of the overlying xanthophores by fixating the fin in 4% paraformaldehyde–PBS for 1 h at room temperature and washing it in a series of methanol:PBS dilutions (25%, 50%, 75% and 100%). Pictures were taken after 6 days in 100% methanol at − 20 °C.

References

1. Pigliucci, M. What, if anything, is an evolutionary novelty? *Philos. Sci.* **75,** 887–898 (2008).
2. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134,** 25–36 (2008).
3. Wagner, G. P. & Lynch, V. J. Evolutionary novelties. *Curr. Biol.* **20,** R48–R52 (2010).
4. Wagner, A. The molecular origins of evolutionary innovations. *Trends Genet.* **27,** 397–410 (2011).
5. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, 2001).
6. Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323,** 737–741 (2009).
7. Lynch, V. J., Leclerc, R. D., May, G. & Wagner, G. P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43,** 1154–1159 (2011).
8. Beldade, P. & Brakefield, P. M. The genetics and evo-devo of butterfly wing patterns. *Nat. Rev. Genet.* **3,** 442–452 (2002).
9. Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evol. Biol.* **5,** 17 (2005).
10. Kocher, T. D. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5,** 288–298 (2004).
11. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol. Ecol.* **18,** 169–185 (2009).
12. Fryer, G. & Iles, T. *The Cichlid Fishes of the Great Lakes of Africa: their Biology and Evolution* (Oliver & Boyd, 1972).
13. Wickler, W. 'Egg-dummies' as natural releasers in mouth-breeding cichlids. *Nature* **194,** 1092–1093 (1962).
14. Santos, M. E. & Salzburger, W. How cichlids diversify. *Science* **338,** 619–621 (2012).
15. Muschick, M., Indermaur, A. & Salzburger, W. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr. Biol.* **22,** 2362–2368 (2012).
16. Salzburger, W., Braasch, I. & Meyer, A. Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. *BMC Biol.* **5,** 51 (2007).
17. Braasch, I. & Postlethwait, J. in *Polyploidy and Genome Evolution* (eds Soltis, D E. & Soltis, P. S.) 341–383 (Springer, 2012).

18. Müller, J. M. *et al.* FHL2, a novel tissue-specific coactivator of the androgen receptor. *EMBO J.* **19,** 359–369 (2000).
19. Brun, J. *et al.* The LIM-only protein FHL2 controls osteoblast mesenchymal cell differentiation through non-canonical Wnt signalling. *Bone* **50,** S76 (2012).
20. Johannessen, M., Møller, S., Hansen, T., Moens, U. & Van Ghelue, M. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. *Cell. Mol. Life Sci.* **63,** 268–284 (2006).
21. Kadrmas, J. L. & Beckerle, M. C. The LIM domain: from the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* **5,** 920–931 (2004).
22. Heule, C. & Salzburger, W. The ontogenetic development of egg-spots in the haplochromine cichlid fish Astatotilapia burtoni. *J. Fish Biol.* **78,** 1588–1593 (2011).
23. Baldo, L., Santos, M. E. & Salzburger, W. Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome Biol. Evol.* **3,** 443–455 (2011).
24. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290,** 1151–1155 (2000).
25. Takahashi, K., Terai, Y., Nishida, M. & Okada, N. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol. Biol. Evol.* **15,** 391–407 (1998).
26. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165,** 349–357 (1969).
27. Britten, R. J. & Davidson, E. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev. Biol.* **46,** 111–138 (1971).
28. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9,** 397–405 (2008).
29. Frohnhöfer, H. G., Krauss, J., Maischein, H.-M. & Nüsslein-Volhard, C. Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. *Development* **140,** 2997–3007 (2013).
30. Patterson, L. B. & Parichy, D. M. Interactions with iridophores and the tissue environment required for patterning melanophores and xanthophores during zebrafish adult pigment stripe formation. *PLoS Genet.* **9,** e1003561 (2013).
31. Manousaki, T. *et al.* Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Mol. Ecol.* **22,** 650–669 (2013).
32. Shimeld, S. M. & Holland, P. W. Vertebrate innovations. *Proc. Natl Acad. Sci. USA* **97,** 4449–4452 (2000).
33. Egger, B., Klaefiger, Y., Theis, A. & Salzburger, W. A sensory bias has triggered the evolution of egg-spots in cichlid fishes. *PLoS ONE* **6,** e25601 (2011).
34. Theis, A., Salzburger, W. & Egger, B. The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. *PLoS ONE* **7,** e29878 (2012).
35. Miller, S., Dykes, D. & Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16,** 1215 (1988).
36. Meyer, B. S. & Salzburger, W. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. *Mol. Ecol. Resour.* **12,** 1097–1104 (2012).
37. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
38. Binladen, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2,** e197 (2007).
39. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).
40. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
41. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
42. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).
43. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).
44. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61,** 539–542 (2012).
45. Zwickl, D. J. *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion* (University of Texas, 2006).
46. Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26,** 1569–1571 (2010).
47. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
48. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
49. Altschul, F. S., Gish, W., Miller, W., Myers, W. E. & Lipman, J. D. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215,** 403–410 (1990).
50. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.* **29,** e45 (2001).
51. Curran, K. *et al.* Interplay between Foxd3 and Mitf regulates cell fate plasticity in the zebrafih neural crest. *Dev. Biol.* **344,** 107–118 (2010).
52. Tang, R., Dodd, A., Lai, D., McNabb, W. C. & Love, D. R. Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT–PCR normalization. *Acta. Biochim. Biophys. Sin.* **39,** 384–390 (2007).
53. Zhang, Z. & Hu, J. Development and validation of endogenous reference genes for expression profiling of medaka (*Oryzias latipes*) exposed to endocrine disrupting chemicals by quantitative real-time RT–PCR. *Toxicol. Sci.* **95,** 356–368 (2007).
54. Catchen, J. M., Conery, J. S. & Postlethwait, J. H. Automated identification of conserved synteny after whole-genome duplication. *Genome Res.* **19,** 1497–1505 (2009).
55. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132,** 365–386 (2000).
56. Flicek, P. *et al.* Ensembl 2012. *Nucleic Acids Res.* **40,** D84–D90 (2012).
57. Larkin, M. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
58. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, 2003).
59. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).
60. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148,** 929–936 (1998).
61. Brawand, D. *et al.* The genomic substrate for adaptive radiation in Africa cichlid fish. *Nature* **513,** 375–381 (2014).
62. Pages, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: String objects representing biological sequences, and matching algorithms. R-package, http://www.bioconductor.org/packages/release/bioc/html/Biostrings.html.
63. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32,** W273–W279 (2004).
64. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13,** 721–731 (2003).
65. Bessa, J. *et al.* Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev. Dyn.* **238,** 2409–2417 (2009).
66. Bessa, J. & Gómez-Skarmeta, J. L. Making reporter gene constructs to analyze cis -regulatory elements. *Mol. Methods Evol. Genet.* **772,** 397–408 (2011).
67. Westerfield, M. *The Zebrafish Book. A guide for the Laboratory Use of Zebrafish (Danio rerio)* (University of Oregon Press, 2000).

Acknowledgements

We thank A. Theis, T. Bosia, B. Egger, G. Banyankimbona, E. Schraml, M. Muschick and H. Gante for their assistance with the photographs; P. Beldade, J. Ganz, H. Hoekstra, J.H. Postlethwait, G. Wagner and the members of the Salzburger Laboratory for their discussions; J. Bessa for sending us the ZED vector; V. Laudet for providing medaka and zebrafish; J. Ganz, University of Oregon and the Imaging Core Facility of the Biozentrum, University of Basel for their support with confocal microscopy; the University of Oregon zebrafish facility, A. Indermaur and A. Rüegg (Basel) for fishkeeping; and the BROAD Institute and the Cichlid Genome Consortium for the cichlid genomes. This study was supported by grants from the Fundação para a Ciência e Tecnologia (FCT, Portugal) to M.E.S.; the Volkswagen Stiftung to I.B. and A.B.; the 'Fellowship for Excellence' from the International PhD Program in Molecular Life Sciences and the Werner-Siemens-Foundation (Aug) to L.S.; the Swiss National Science Foundation (SNF) to M.A.; and the European Research Council (Starting Grant 'INTERGENADAPT"), the University of Basel, the National Geographic Society and the SNF to W.S. The work of I.B. was further supported by National Institutes of Health grant R01 RR020833 (R01 OD011116) to John H. Postlethwait.

Author contributions

M.E.S., I.B. and W.S. designed the study; M.E.S and W.S. collected the samples; M.E.S. performed the RNAseq, gene expression, comparative genomics and zebrafish functional analysis; N.B. performed the sequencing of *fhl2* paralogs coding region and analysed its rates of evolution; B.S.M. collected the 454 sequence data; and B.S.M. and W.S. performed the phylogenetic analysis; A.B. performed the SINE consensus alignments and analysed the transposable element composition of *fhl2b* genomic region; I.B. performed the zebrafish functional assays of the *A. burtoni* construct and *fhl2* paralogs synteny

analysis; L.S., H.-G.B. and M.A. assisted with the zebrafish functional assays of the *A. burtoni*, *P. philander* and *N. sexfasciatus* construct; and M.E.S. and W.S. wrote the paper and all authors contributed to revisions.

Additional information

Accession codes: All nucleotide sequences reported in this study have been deposited in GenBank/EMBL/DDBJ under the accession codes KM263618 to KM264016. All the short reads have been deposited in GenBank/EMBL/DDBJ Sequence Read Archive (SRA) under the BioProject ID PRJNA25755.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

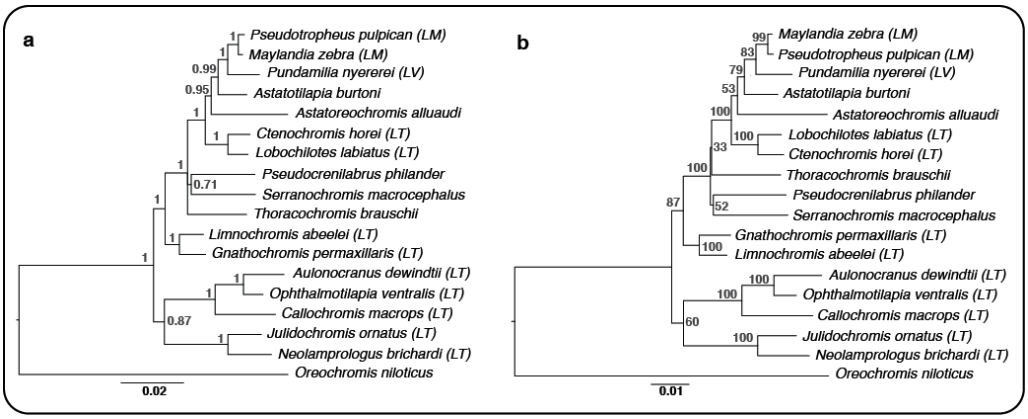
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Santos, M. E. *et al.* The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change. *Nat. Commun.* 5:5149 doi: 10.1038/ncomms6149 (2014).

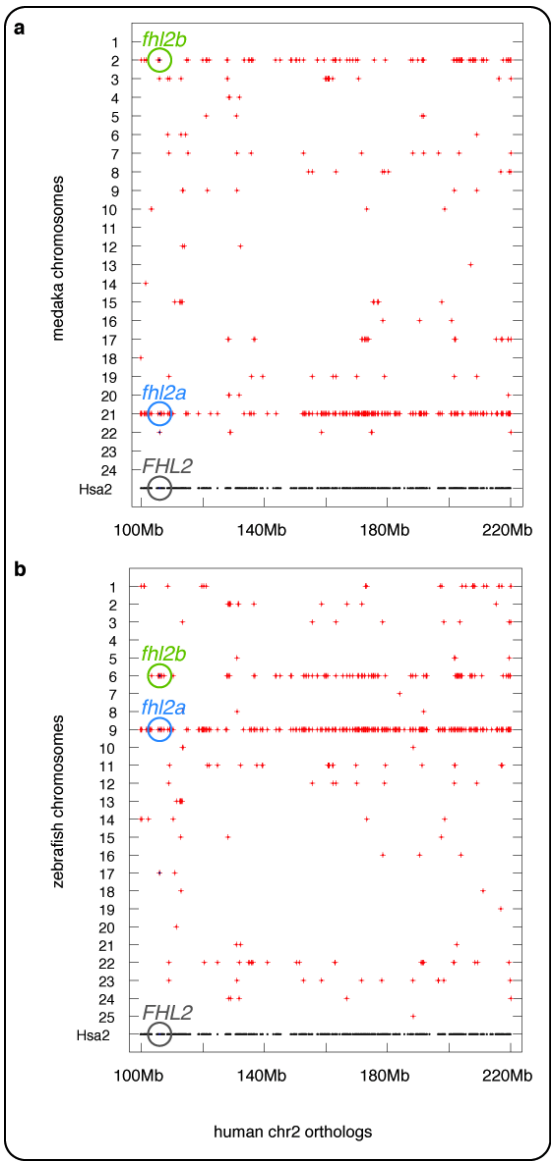
Supplementary Material

The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change

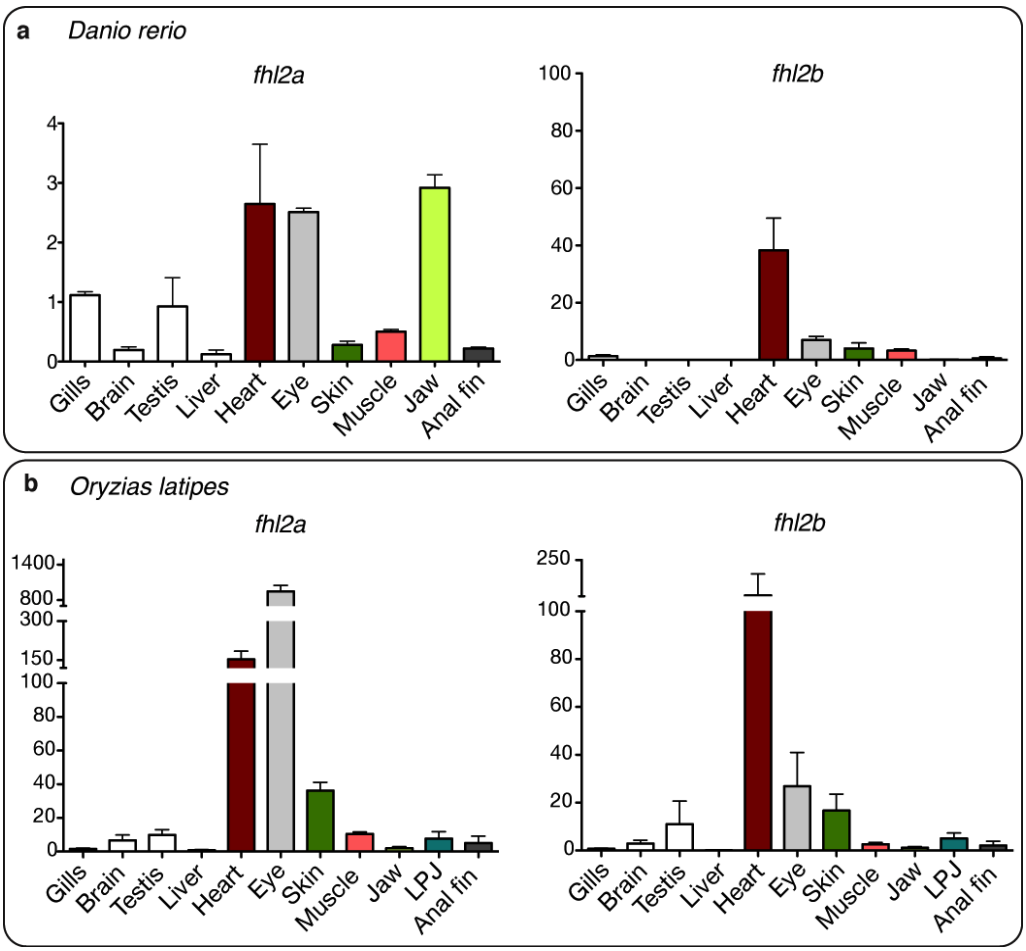
M. Emília Santos, Ingo Braasch, Nicolas Boileau, Britta S. Meyer, Loïc Sauteur, Astrid Böhne, Heinz-Georg Belting, Markus Affolter, Walter Salzburger



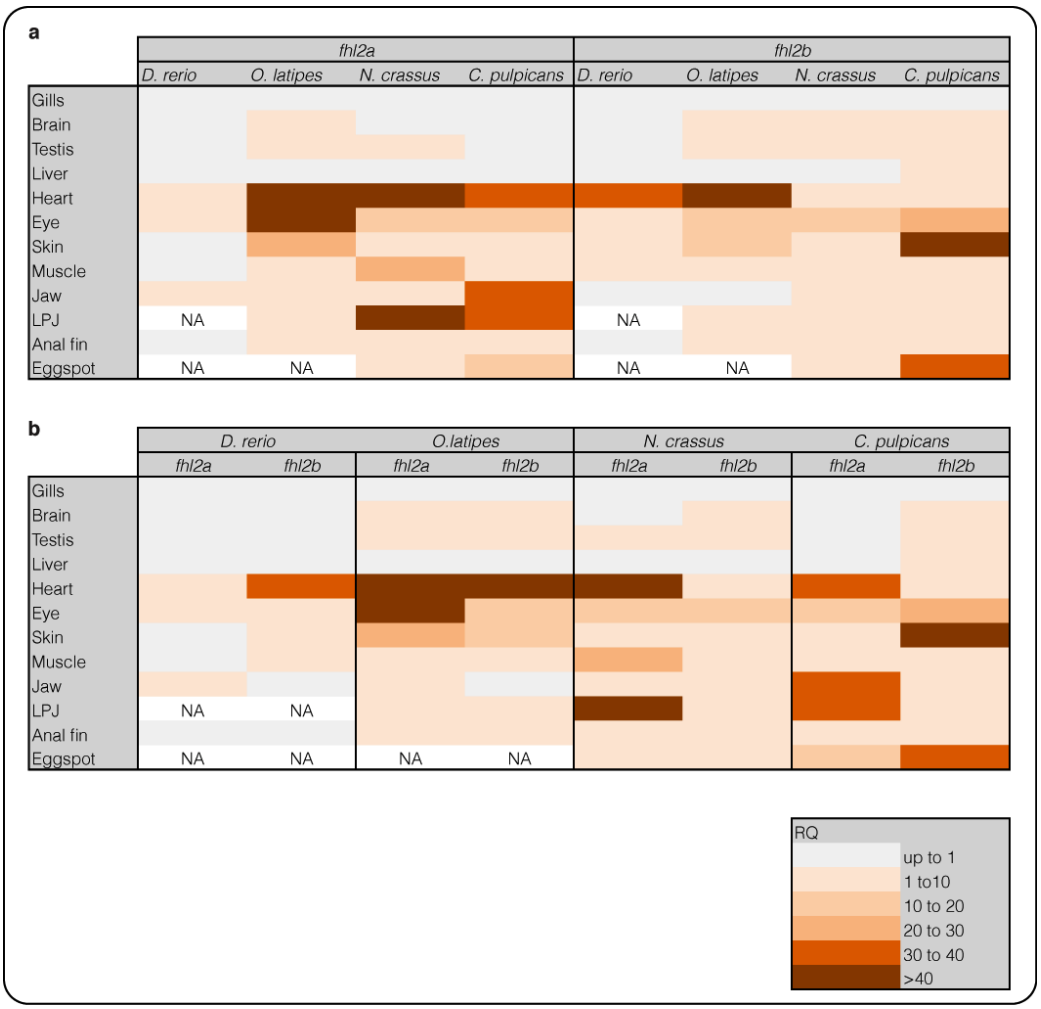
Supplementary Figure 1 | Phylogeny of East African cichlids based on a new multi-marker dataset. (**a**) Bayesian inference phylogeny with MrBayes. (**b**) Maximum likelihood phylogeny with GARLI and 500 bootstrap replicates. While most of the branches are supported with high posterior probabilities (**a**) and bootstrap values (**b**), the phylogenetic relationships among the more ancestral haplochromines – including *Pseudocrenilabrus philander* – are poorly supported and differ between the analyses. LM: Lake Malawi, LV: Lake Victoria, LT: Lake Tanganyika



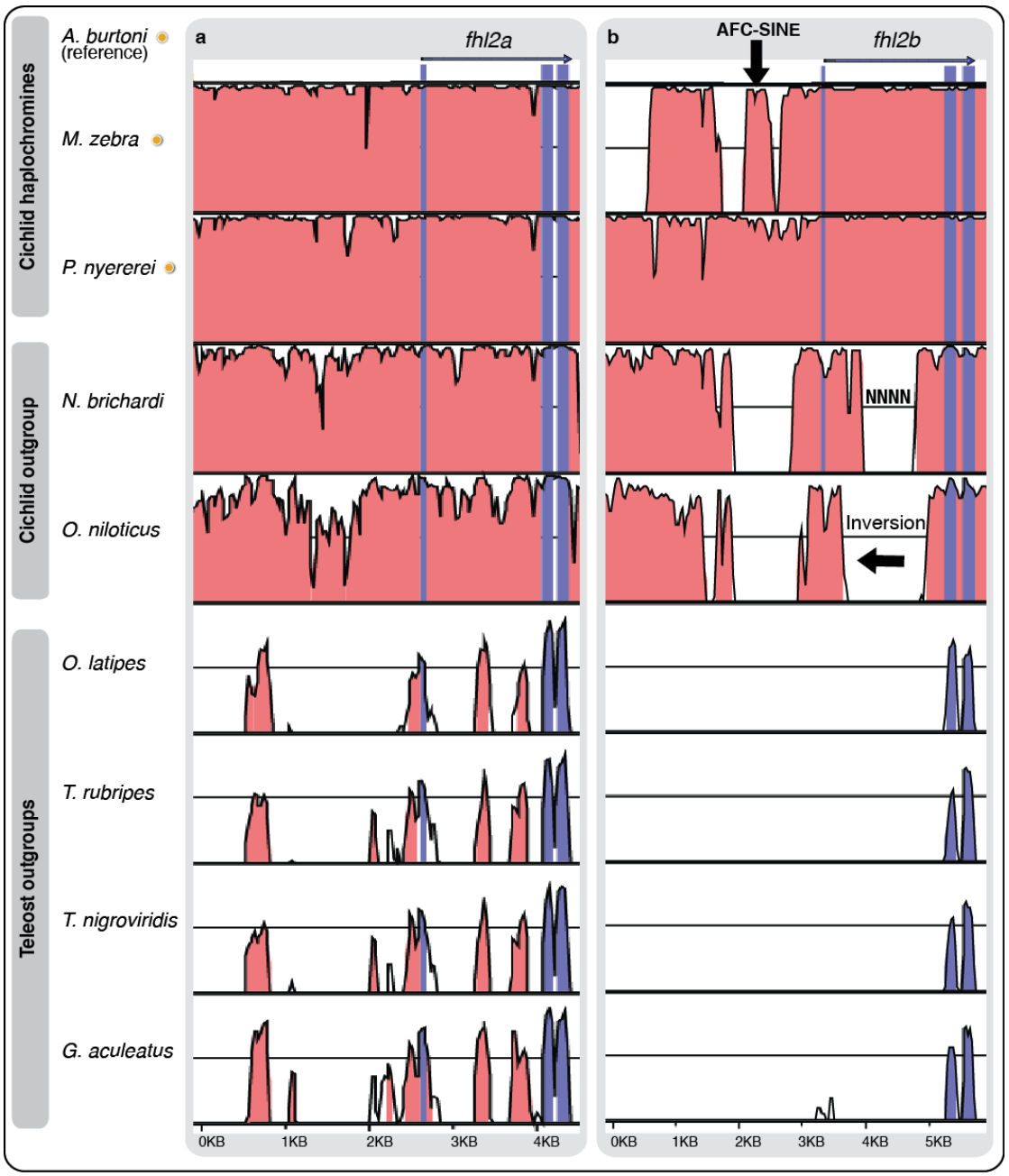
Supplementary Figure 2 | Synteny analysis of teleost *fhl2* paralogs. Dotplots of the human FHL2 gene region on human chr2 (100-220Mb) shows double conserved synteny to the two *fhl2* paralogons in (**a**) medaka on chromosomes Ola21 (*fhl2a*) and Ola2 (*fhl2b*) and in (**b**) zebrafish on chromosomes Dre9 (*fhl2b*) and Dre6 (*fhl2b*). These chromosomes were previously shown to be derived from the ancestral chromosome *c* and duplicated during the teleost genome duplication[1,2].



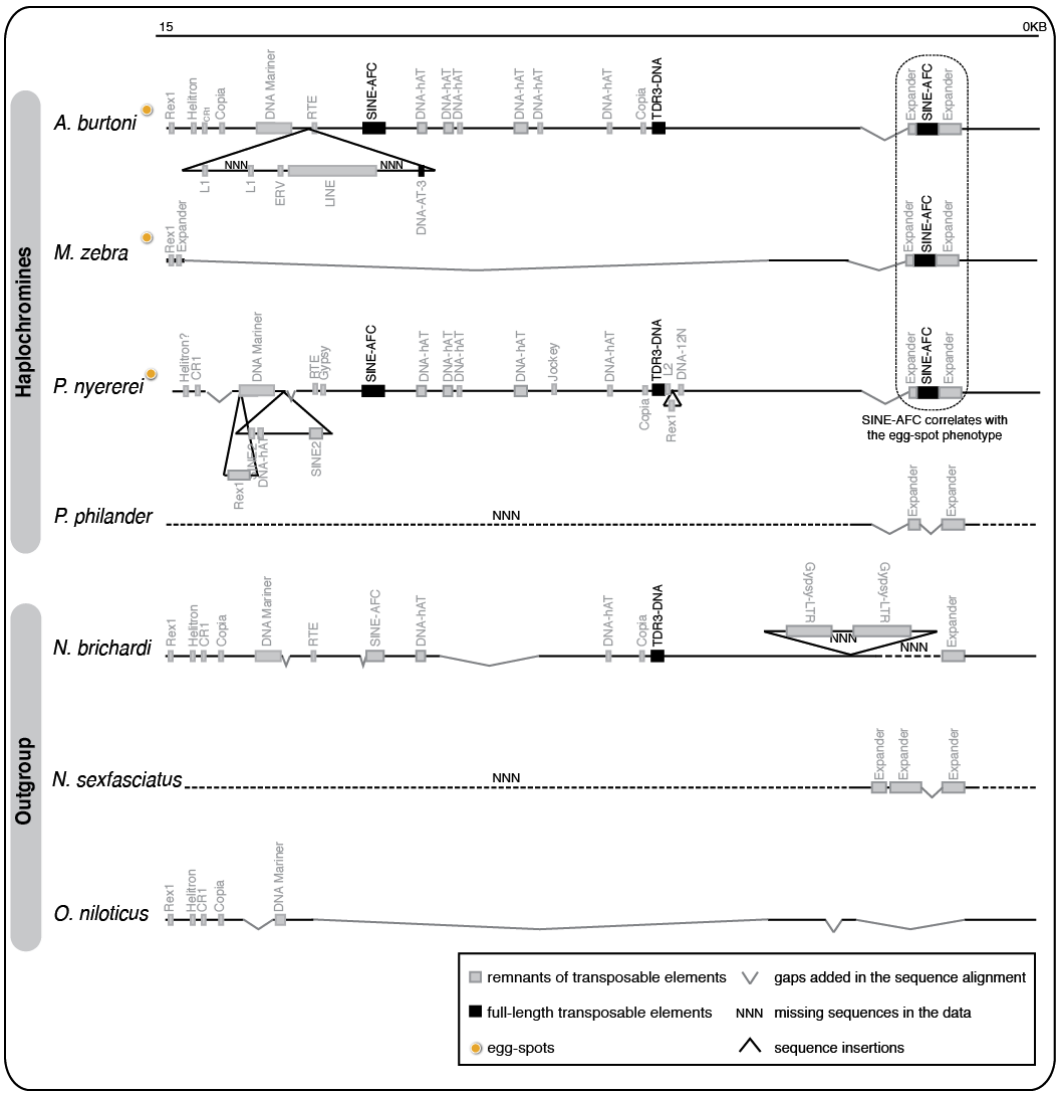
Supplementary Figure 3 | Gene expression profiling in the teleost *Danio rerio* and *Oryzias latipes*. (**a**) Relative quantification (RQ) of *fhl2a* and *fhl2b* gene expression in ten tissues in *D*. *rerio* (three replicates per tissue) (**b**) RQ of *fhl2a* and *fhl2b* gene expression in eleven tissues in *O*. *latipes* (three replicates per tissue). In both species, gill tissue was used as reference. The error bars represent the standard error of the mean (SEM). In *D*. *rerio* (**a**) expression of *fhl2a* is higher in heart, eye, and oral jaw, although the expression of this gene copy is overall very low, especially when compared to the level of *fhl2a* expression in cichlids and *O*. *latipes*. Contrary to the scenario in cichlids (Fig. 3), in *D*. *rerio fhl2b* is mainly expressed in the heart. In *O*. *latipes* (**b**) both duplicates are highly expressed in heart, skin and eye tissues. *fhl2a* does not show high expression levels in the pharyngeal jaw (unlike cichlid *fhl2a)*. In this species both copies show a similar expression profile. These results suggest that the divergence history between the duplicates was different in the different lineages of teleosts, where divergence in expression profile is stronger in cichlids (see Figure 3).



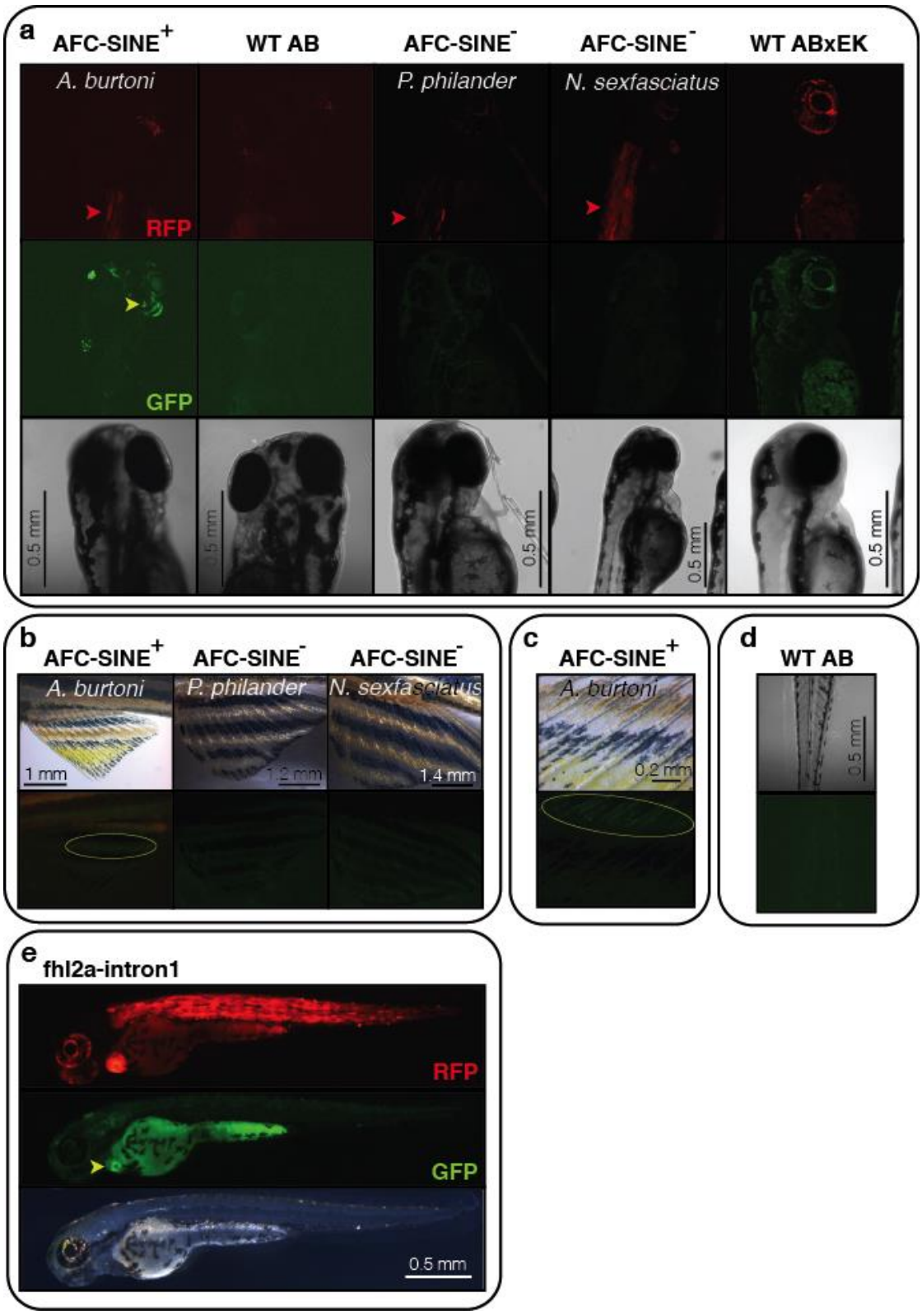
Supplementary Figure 4 | *fhl2* duplicates expression heatmap. The relative quantification values (RQ) from the qPCR experiment 3 (see methods, Fig. 3 and Supplementary Fig. 4) were categorized and color coded accordingly. Colour code and RQ value categories can be found in the bottom of this figure. The heatmap in (**a**) is grouped by gene, whereas the heatmap in (**b**) is grouped by species. The grouping of expression data by gene (**a**) suggests that both paralogs seems to have gained a new function in cichlids (*fhl2a* in jaw and LPJ and *fhl2b* in skin and egg-spot tissue). The grouping of the gene expression data by species (**b**) suggests that the expression profiles of *D. rerio* and *O. latipes* are similar, whereas the expression patterns of cichlids are more divergent.


Cichlid haplochromines
Cichlid outgroup
Teleost outgroups
A. burtoni (reference)
M. zebra
P. nyererei
N. brichardi
O. niloticus
O. latipes
T. rubripes
T. nigroviridis
G. aculeatus
a
fhl2a
b
AFC-SINE
fhl2b
NNNN
Inversion
0KB
1KB
2KB
3KB
4KB
5KB

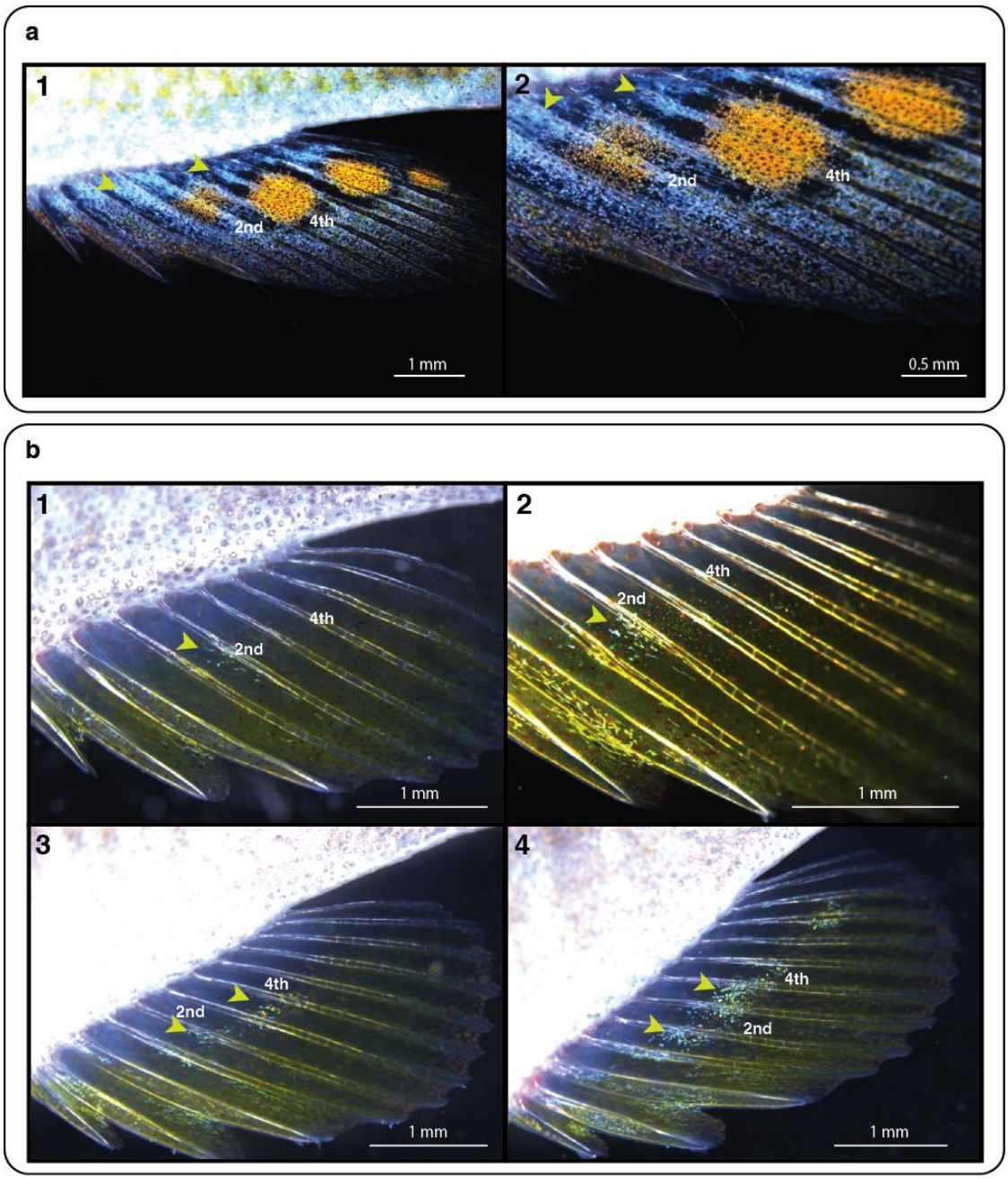
Supplementary Figure 5 | Genomic comparisons of *fhl2* paralogs of cichlids and other teleosts. *A. burtoni* genomic sequences were used as a reference for the alignments. The conserved regions (>70% identity) are marked in pink. Any conservation of the non-coding sequence across distant taxa is an indication of functional constraint and therefore of their potential role as enhancers/promoters. (**a**) mVISTA plot of a 5kb region of the *fhl2a* locus; the first three exons of *fhl2a* are shown in purple. We found four CNEs (conserved non-coding elements) and all of them are shared amongst cichlids; no major difference was detected among these CNEs. (**b**) mVISTA plot of a 5kb region of *fhl2b* locus; the first three exons of *fhl2b* are shown in purple. There are no conserved non-coding elements (CNEs) in common with the other examined teleosts, the only conserved regions are the exons. There are two regions that are conserved among haplochromines but not with non-haplochromines. One region corresponds to the first intron (positioned between 3.5-5kb), and the other region (positioned between 2-3kb) corresponds to a transposable element insertion (vertical arrow on top of the diagram), which is only present in haplochromines, the egg-spot bearing lineage. More specifically, we found that the egg-spot bearing haplochromines are characterized by an AFC-SINE insertion upstream of the *fhl2b* open reading frame. After close inspection we determined that the lack of conservation in intron one is a result of inversions (marked with a horizontal arrow) (in *O. niloticus* or in the other cichlids examined) and a missing sequence in the genome assembly of *N. brichardi* (marked with NNNN). This region most probably cannot explain the presence/absence of the egg-spot since the transcription factor binding sites would not be lost due to the inversion. These results indicate that a SINE element insertion is the likely explanation for the difference in *fhl2b* expression between haplochromines and non-haplochromines (Fig. 3).



Supplementary Figure 6 | Detailed characterization of the *fhl2b* upstream region in cichlid fishes. A more detailed characterization of the upstream genomic region of *fhl2b* shows that multiple transposon element insertions occurred in different cichlid species. It seems that this upstream region is prone to insertions (when compared to *fhl2a*), and that these insertions might disrupt the regulatory regions of this gene, explaining why *fhl2b* is more divergent in terms of gene expression than *fhl2a*. The AFC-SINE insertion is the only transposable element insertion that correlates with the egg-spot phenotype though. Interestingly, *M. zebra* (haplochromine) has several deletions in the *fhl2b* upstream region, but the AFC-SINE element is still conserved suggesting that this element is functionally important.

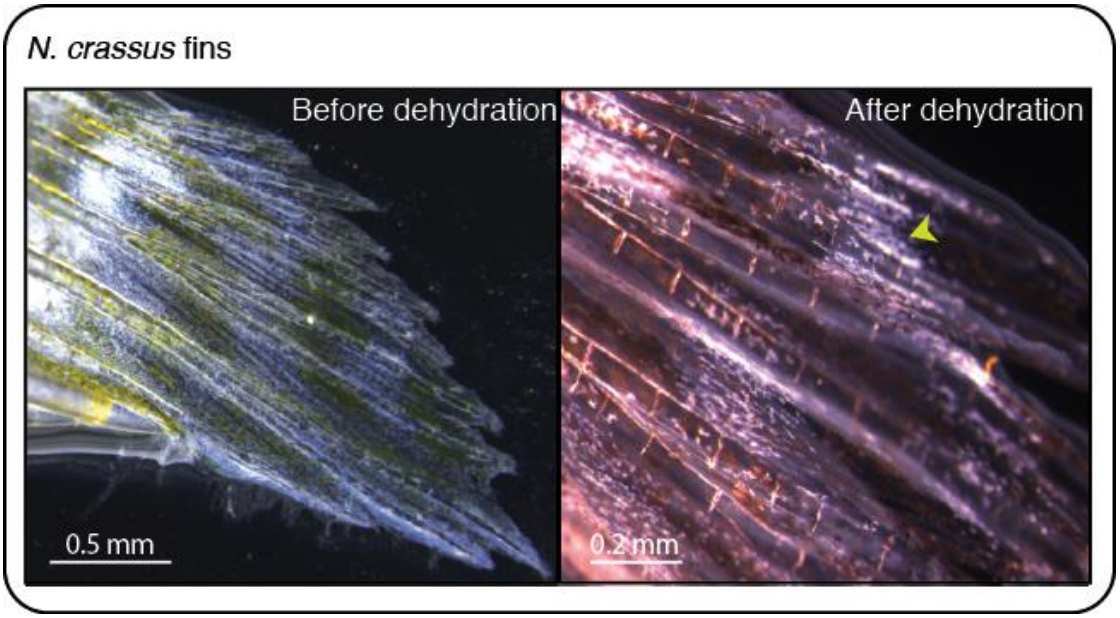

a
AFC-SINE+
WT AB
AFC-SINE-
AFC-SINE-
WT ABxEK
A. burtoni
P. philander
N. sexfasciatus
RFP
GFP
0.5 mm
b
AFC-SINE+
AFC-SINE-
AFC-SINE-
A. burtoni
P. philander
N. sexfasciatus
1 mm
1.4 mm
c
AFC-SINE+
A. burtoni
0.2 mm
d
WT AB
0.5 mm
e
fhl2a-intron1
RFP
GFP
0.5 mm

Supplementary Figure 7 | Molecular and cellular basis of egg-spot development. (**a**) The upper panel shows images of the head region of a 3 day-old zebrafish embryo under RFP filter, the middle panel is the same region under GFP light and the lower panel under bright field light. The RFP shows that transgenesis was effective (positive control, for more information see ref. 3). The *A*. *burtoni* AFC-SINE^{+} construct drives GFP expression in the iridophores of the embryo eye (yellow arrow), a pattern that is not seen with any of the AFC-SINE^{-} constructs (*P*. *philander* and *N*. *sexfasciatus*). The two wild-type strains (AB and ABxEK) used in this study were also imaged. (**b**) Only the AFC-SINE^{+} construct drives GFP expression in the iridophores of the adult anal fin in zebrafish (the stripe of iridophores is surrounded by yellow circle). (**c**) Higher magnification image of the anal fin of the AFC-SINE^{+} construct showing GFP expression in the iridophores. (**d**) Zebrafish wild-type strain AB. This image complements the main manuscript figure 4 where we only show the imaging for the wild-type strain ABxEK. (**e**) The construct containing the first exon and intron of *fhl2a* of *A*. *burtoni* drove expression in heart in zebrafish. Note however, that this experiment is not exactly comparable to the one with *fhl2b*, as the fhl2a construct did not contain the upstream region.

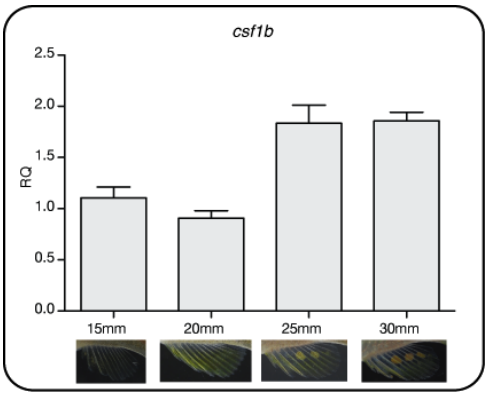


Supplementary Figure 8 | Iridophores in *A. burtoni* adult and developing fins.
(a) The upper panels show an anal fin of a juvenile *A. burtoni*. Panel 2 is a close up of panel 1. In addition to the high density of iridophores in the egg-spots (as shown in Figure 4e) we also find iridophores in high density in the proximal region of the fin (depicted with yellow arrows), which coincides with *fhl2a/b* expression patterns (see Figure 2c for *fhl2b*), suggesting that these genes are indeed iridophore genes. The first and second egg-spot are located in the second and fourth soft fin ray, respectively. Interestingly, iridophores are present in the fin rays, but absent from the inter-rays around the egg-spots, hinting at potential inhibitory relationships between different cellular types. Female anal fins (not shown) of *A. burtoni* show much smaller spots, which, in addition, emerge at later stages, and contain a much smaller number of xanthophores and iridophores. **(b)** The lower panels show different

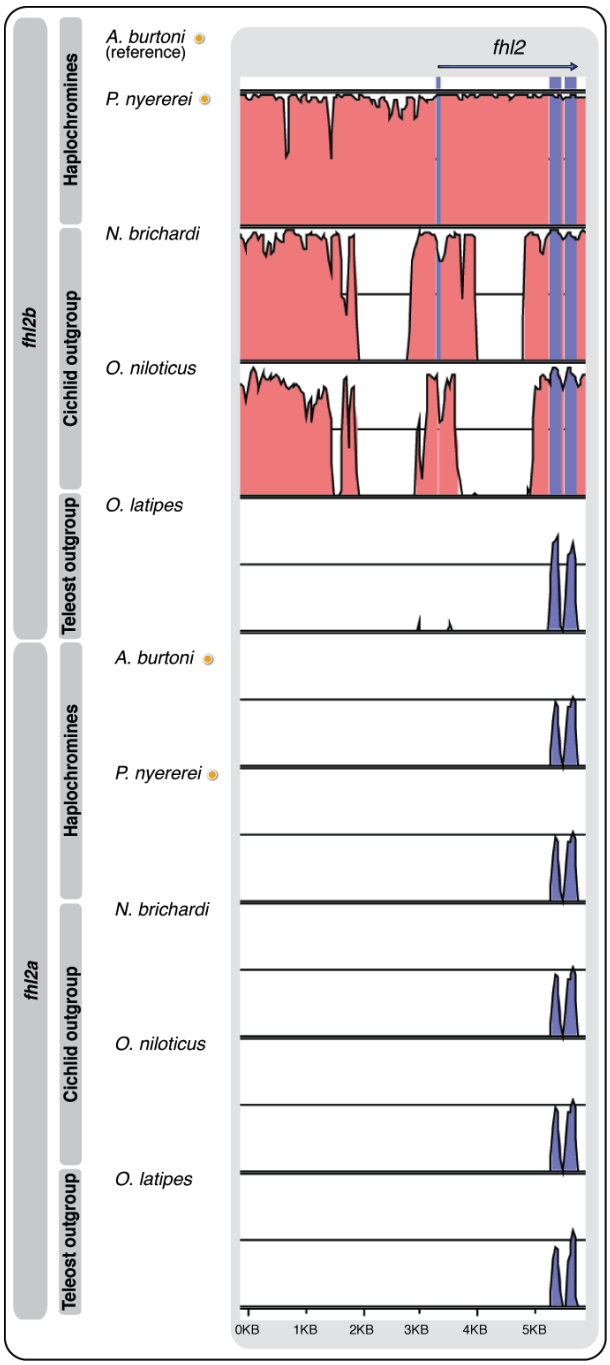
early stages of egg-spot development. The iridophores (blue cells depicted with yellow arrows) start accumulating in the second soft fin ray and then on the fourth (panel 1 and 2). At later stages (panel 3 and 4) xanthophores (yellow cells) start to accumulate where iridophores are present.



Supplementary Figure 9 | Characterization of pigment cells present in the non-haplochromine *N. crassus* fin. To better understand the morphological differences between haplochromine and non-haplochromine fins we dissected *N. crassus* anal fins. The fins of *N. crassus* show a yellow and silver transparent banded pattern (see left panel). The yellow bands are composed of xanthophores that are not visible after dehydration as the pteridine pigments are washed away. Contrary to fins with egg-spots, there are no iridophores in the xanthophore-rich regions. Instead the iridophores concentrate in the silver transparent band. Examples of iridophores are highlighted with a yellow arrow.



Supplementary Figure 10 | Expression of *csf1b* in egg-spot development. In addition to the expression profiles of *fhl2* duplicates (Fig. 2), *csf1ra*, *mitfa* and *pnp4a*, we studied the expression profile of *csf1b* (a key gene in mediating iridophores-xanthophore interaction in zebrafish) during the development of egg-spots in *A. burtoni*. The values on the x-axis represent fish standard length in millimeters. *csf1b* doubles in expression when the egg-spots emerge.



Supplementary Figure 11 | Genomic comparisons between *fhl2a* and *fhl2b* genomic region. The *A. burtoni fhl2b* genomic sequence was used as a reference for the alignment. As in Supplementary Fig. 3 the conserved regions (>70% identity) are marked in pink and conservation of the non-coding sequence across these two genes could be an indication of shared enhancers/promoters. There are no CNEs (conserved non-coding elements) shared between the two paralogs. This comparison shows that that the two regulatory regions are different.

Supplementary Table 1| Multimarker dataset used for phylogenetic analyses. The table lists GenBank accession numbers and the models of molecular evolution used in the analyses.

	Nuclear marker									
Species	*rag*	*gapdhs*	*s7*	*bmp4*	*ednrb1*	*mitfa*	*tyr*	*hag*	*csfr1a*	*nd2*
Astatotilapia burtoni	KM263618	KM263633	KM263648	KM263663	KM263678	KM263693	KM263708	KM263723	KM263738	AY930060
Astatoreochromis alluaudi	KM263630	KM263645	KM263660	KM263675	KM263690	KM263705	KM263720	KM263735	KM263750	JN628855
Pundamilia nyererei	KM263628	KM263643	KM263658	KM263668	KM263688	KM263703	KM263718	KM263733	KM263748	-
Metriaclima zebra	AGTA02035876	AGTA02043695	AGTA02046533	AGTA02003990	AGTA02044321	AGTA02044722	AGTA02007920	AGTA02058917	AGTA02053932	GU192142
Cynotilapia pulpican	KM263629	KM263644	KM263659	KM263667	KM263689	KM263704	KM263719	KM263734	KM263749	-
Ctenochromis horei	KM263623	KM263638	KM263653	KM263673	KM263683	KM263698	KM263713	KM263728	KM263743	JQ755337
Lobochilotes labiatus	KM263624	KM263639	KM263654	KM263671	KM263684	KM263699	KM263714	KM263729	KM263744	EF679250
Gnathochromis permaxillaris	KM263626	KM263641	KM263656	KM263672	KM263686	KM263701	KM263716	KM263731	KM263746	AY682522
Pseudocrenilabrus philander	KM263622	KM263637	KM263652	KM263674	KM263682	KM263697	KM263712	KM263727	KM263742	JX910862
Callochromis macrops	KM263620	KM263635	KM263650	KM263665	KM263680	KM263695	KM263710	KM263725	KM263740	CMU07242
Ophthalmotilapia ventralis	KM263621	KM263636	KM263651	KM263666	KM263681	KM263696	KM263711	KM263726	KM263741	AY337774
Aulonocranus dewindtii	KM263619	KM263634	KM263649	KM263664	KM263679	KM263694	KM263709	KM263724	KM263739	AY337782
Julidochromis ornatus	KM263627	KM263642	KM263657	KM263669	KM263687	KM263702	KM263717	KM263732	KM263747	EF462229
Limnochromis abeelei	KM263625	KM263640	KM263655	KM263670	KM263685	KM263700	KM263715	KM263730	KM263745	AY682535
Oreochromis niloticus	AERX01005120	AERX01010358	AERX01016503	AERX01008958	AERX01030771	AERX01008591	AERX01007002	AERX01042180	AERX01036679	GU477624
Neolamprologus brichardi	AFNY01022509	AFNY01036835	AFNY01033177	AFNY01048459	JF900291	AFNY01012153	AFNY01063785	AFNY01102376	AFNY01005183	AP006014
Thoracochromis brauschi	KM263632	KM263647	KM263662	KM263677	KM263692	KM263707	KM263722	KM263737	KM263752	AY930095
Serranochromis macrocephalus	KM263631	KM263646	KM263661	KM263676	KM263691	KM263706	KM263721	KM263736	KM263751	EF393690
sequenced basepairs	418	462	487	468	440	425	513	427	366	1045
used model in Garli/MrBayes (BIC)	JC	JC	HKY	HKY	HKY	HKY	HKY	K80	K80+G	TrN+G /GTR +G

Supplementary Table 2 | Top 10 differential expressed transcripts between female and male anal fins of *A. burtoni* and their identification as determined by BLASTx searches against the NCBI non-redundant database[4]. From these ten transcripts, three were identified as *fhl2* - four and a half LIM domain protein 2 - and these were among the most differentially expressed genes. Whilst tetrapods have one copy of *fhl2*, the majority of teleosts have two copies due to the extra whole genome duplication[5]. We aligned the three transcripts and observed that they are, instead of one gene, two duplicates - *fhl2a* and *fhl2b* (identification via protein homology with other teleosts and by phylogenetic inference). Comp19010_c0_seq1 and comp17680_c0_seq1 correspond to two different parts of *fhl2b* gene and comp2939_c0_seq1 corresponds to *fhl2a*. logFC stands for $\log_2$Fold-Change in gene expression between male fins and female fins. The short read sequences are deposited under the BioProject ID PRJNA257552.

Differential expressed transcripts			**BLAST Identification**		
Transcript	**logFC**	**p-value**	**Description**	**Accession**	**e-value**
comp19010_c0_seq1 (*fhl2b*)	-5.296931976	2.55E-12	PREDICTED: four and a half LIM domains protein 2-like (Oreochromis niloticus)	XP_003446591.1	2.00E-119
comp17680_c0_seq1 (*fhl2b*)	-5.106938251	4.63E-12	PREDICTED: four and a half LIM domains protein 2-like (Oreochromis niloticus)	XP_003446591.1	3.00E-84
comp11583_c0_seq1	-4.486139111	2.41E-09	PREDICTED: similar to ORF2-encoded protein, partial (Hydra magnipapillata)	XP_002155414.1	4.00E-77
comp2939_c0_seq1 (*fhl2a*)	-4.257228806	6.57E-09	PREDICTED: four and a half LIM domains protein 2-like (Oreochromis niloticus)	XP_003453001.1	0
comp35399_c0_seq1	-4.52411029	1.32E-08	PREDICTED: crystal protein-like (Danio rerio)	XP_002661384.2	5.00E-164
comp6540_c0_seq1	-4.09287457	3.01E-08	PREDICTED: apolipoprotein D-like (Oreochromis niloticus)	XP_003448594.1	2.00E-123
comp7947_c0_seq3	-4.108859622	7.42E-08	PREDICTED: vitronectin-like (Oreochromis niloticus)	XP_003458657.1	0
comp7947_c0_seq1	-3.902264412	1.79E-07	PREDICTED: vitronectin-like (Oreochromis niloticus)	XP_003458657.1	0
comp51734_c0_seq1	-4.539615034	1.79E-07	No significant similarity found	---NA---	NA
comp7947_c0_seq2	-3.875437404	2.10E-07	PREDICTED: vitronectin-like (Oreochromis niloticus)	XP_003458657.1	0

Supplementary Table 3| Subset of primers used in this study.

Primer	Sequence 5'-3'	TºC	Task
Fhl2a_qpcr_fw	AAC ACC AGG GAT CTT TCC TAC AAG	58	qPCR in cichlids
Fhl2a_qpcr_rev	GCA CTG GAA GCA CTT AAA GCA TT	58	qPCR
Fhl2b_qpcr_fw	AGC AAG GAT CTG TCG TAC AAG GA	58	qPCR
Fhl2b_qpcr_rev	AGA CCG GCT GCA CTT GTT G	58	qPCR
RPL7_qpcr_fw	GGA GAA GTC CCT CGG CAA AT	58	qPCR
RPL7_qpcr_rev	GGC GGG CTT GAA GTT CTT TC	58	qPCR
RsPA3_qpcr_fw	AGA CCA ATG ACC TGA AGG AAG TG	58	qPCR
RsPA3_qpcr_rev	TCT CGA TGT CCT TGC CAA CA	58	qPCR
Fhl2b_probe_fw	GGT CCT CGA CTG CTA CCA AG	64	*In situ* probe PCR
Fhl2b_probe_rev	TTG CAG TTG AAG CAA TCG TT	64	*In situ* probe PCR
Fhl2a_probe_fw	CAG ACG TCC TCA GAC AGG AA	64	*In situ* probe PCR
Fhl2a_probe_rev	TGC ATC GTT CCC TGA TCA TA	64	*In situ* probe PCR
Fhl2b_TEregion_fw	CTA CTG GTG TTG GCC AGA GG	62	AFC-SINE + intron1
Fhl2b_exon2_rev	GAG AAT AGC GTC TCA TAG CAC T	62	AFC-SINE + intron1
Mitfa_fw	GCC TCG CCA TCA ACA GTT GT		qPCR in cichlids
Mitfa_rev	TCA TGC CAG GAG CAG TGA ATT		qPCR in cichlids
Csf1ra_fw	CTC AGG GCC TCG ACT TTT TG		qPCR in cichlids
Csf1ra_rev	TTC CTC GCA GCC ACA TCT C		qPCR in cichlids
Pnp4a_fw	CAT GAC CCT GGA CTG TGC TC		qPCR in cichlids
Pnp4a_rev	CTG GCT GAT GTC CCA AAC AA		qPCR in cichlids
Csf1b_fw	CCC ATG CAG ACA CTC CAT CA		qPCR in cichlids
Csf1b_rev	TTT GCT CAA ACT CCT CCG TTC		qPCR in cichlids
Ltk_fw	CTC AGG ACA GTG CTG CCA AC		qPCR in cichlids
Ltk_rev	CAG GAT GGA TCC TCC CAA AG		qPCR in cichlids
Fhl2a_Drerio_qpcr_fw	CGG CTG CGC AGA AGT AAA G		qPCR in *D. rerio*
Fhl2a_Drerio_qpcr_rev	GTA TGG GTT GTC CTC ACG CA		qPCR in *D. rerio*
Fhl2b_Drerio_qpcr_fw	CAC GGG ACA GGG ATT GTT TA		qPCR in *D. rerio*
Fhl2b_Drerio_qpcr_rev	CCC GAA CAG AGA CTC CTT GC		qPCR in *D. rerio*
Fhl2a_Olatipes_qpcr_fw	CAC TGC AAG AAG CCC ATC AC		qPCR in *O. latipes*
Fhl2a_Olatipes_qpcr_rev	ACG AAG CAC TCT TTG TGC CA		qPCR in *O. latipes*
Fhl2b_Olatipes_qpcr_fw	CCA ACA CCT GTG AGG AAT GC		qPCR in *O. latipes*

Supplementary Table 4 | Sequencing of the *fhl2a* and *fhl2b* coding region sequencing in cichlids. The table lists specimens, primer sequences and annealing temperatures (the sequences are available under the accession numbers KM263753 to KM263998).

Species name	Lineage	Anal fin egg-spot	Nr of ind.(fhl2a/fhl2b)
Astatoreochromis alluaudi	Haplochromini	Yes	5 / 5
Astatotilapia burtoni	Haplochromini	Yes	5 / 5
Labidochromis caeruleus	Haplochromini	Yes	5 / 4
Pseudocrenilabrus multicolor	Haplochromini	Blotch	4 / 4
Pseudotropheus elegans	Haplochromini	Yes	4 / 5
Cynotilapia pulpican	Haplochromini	Yes	4 / 5
Ctenochromis horei	Haplochromini	Yes	5 / 5
Interochromis loocki	Haplochromini	Yes	4 / 2
Petrochromis famula	Haplochromini	Yes	5 / 5
Petrochromis fasciolatus	Haplochromini	Yes	5 / 5
Petrochromis polyodon	Haplochromini	Yes	4 / 5
Pseudosimochromis curvifrons	Haplochromini	Yes	5 / 5
Simochromis diagramma	Haplochromini	Yes	5 / 5
Tropheus moori	Haplochromini	Yes	5 / 5
Julidochromis marlieri	Lamprologini	No	5 / 5
Julidochromis ornatus	Lamprologini	No	5 / 5
Neolamprologus brichardi	Lamprologini	No	4 / 4
Neolamprologus pulcher	Lamprologini	No	4 / 5
Neolamprologus sexfasciatus	Lamprologini	No	5 / 5
Cyprichromis 'jumbo'	Cyprichromini	No	5 / 5
Cyprichromis leptosoma	Cyprichromini	No	5 / 5
Aulonocranus dewindtii	Ectodini	No	5 / 5
Callochromis macrops	Ectodini	Blotch	5 / 5
Cyathopharynx furcifer	Ectodini	No	4 / 5
Ophthalmotilapia ventralis	Ectodini	No	5 / 5
Xenotilapia flavipinnis	Ectodini	No	5 / 5
Primer	**Sequence 5'-3'**	**PCR annealing T°C**	**Task**
Fhl2b_CDS_fw	GGT CCT CGA CTG CTA CCA AG	65-68	Coding region PCR/Seq.
Fhl2b_CDS_rev	ATC CGG CTC GGG TTG TCT	65-68	Coding region PCR/Seq.
Fhl2a_CDS_fw	CTG CCA CAG ACT CCA CAC AG	65-68	Coding region PCR/Seq.
Fhl2a_CDS_rev	TGC ATC GTT CCC TGA TCA TA	65-68	Coding region PCR/Seq.

Supplementary Table 5 | Species (and GenBank accession numbers) used to infer the *fhl2* gene tree.

Species name	Lineage	*fhl2a*	*fhl2b*	*fhl2*
Astatoreochromis alluaudi	Cichlidae	KM263877	KM263753	NA
Astatotilapia burtoni	Cichlidae	KM263882	KM263758	NA
Pseudocrenilabrus philander	Cichlidae	KM263976	KM263853	NA
Cynotilapia pulpican	Cichlidae	KM263980	KM263857	NA
Ctenochromis horei	Cichlidae	KM263897	KM263773	NA
Petrochromis fasciolatus	Cichlidae	KM263958	KM263833	NA
Pseudosimochromis curvifrons	Cichlidae	KM263967	KM263843	NA
Simochromis diagramma	Cichlidae	KM263984	KM263862	NA
Tropheus moori	Cichlidae	KM263989	KM263867	NA
Julidochromis ornatus	Cichlidae	KM263925	KM263800	NA
Neolamprologus brichardi	Cichlidae	KM263935	KM263809	NA
Neolamprologus sexfasciatus	Cichlidae	KM263943	KM263818	NA
Cyprichromis 'jumbo'	Cichlidae	KM263906	KM263783	NA
Cyprichromis leptosoma	Cichlidae	KM263911	KM263788	NA
Aulonocranus dewindtii	Cichlidae	KM263887	KM263763	NA
Callochromis macrops	Cichlidae	KM263892	KM263768	NA
Cyathopharynx furcifer	Cichlidae	KM263902	KM263778	NA
Xenotilapia flavipinnis	Cichlidae	KM263994	KM263872	NA
Oryzias latipes	Adrianichthyidae	ENSORLG00000012482	ENSORLG00000001848	NA
Takifugu rubripes	Tetraodontidae	ENSTRUG00000013559	ENSTRUG00000008468	NA
Gasterosteus aculeatus	Gasterosteidae	ENSGACG00000003005	ENSGACG00000015048	NA
Danio rerio	Cyprinidae	ENSDARG00000042018	ENSDARG00000003991	NA
Anolis carolinensis	Polychrotidae	NA	NA	ENSACAG00000010422
Mus musculus	Muridae	NA	NA	ENSMUSG00000008136

Foreground Branch	Model	Background ω				Foreground ω				LRT	*P*-value	Site	BEB
	Site classes	0	1	2a	2b	0	1	2a	2b				
***fhl2a* locus**													
Haplochromini (*Psephi excluded)*	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	-3.26212	1	--	--
Haplochromini + Tropheini (*Psephi excluded)*	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	0	1	--	--
Haplochromini + Tropheini (*Psephi included)*	Null	0.00512	1.00000	0.00512	1.00000	0.00512	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	-0.548136	1	--	--
Tropheini	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	-0.245402	1	--	--
Pseele	Null	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000				
	Alternative	0.00546	1.00000	0.00546	1.00000	0.00546	1.00000	1.00000	1.00000	3.26212	0.07089741	--	--
Cichlids (including Tilapia)	Null	0.00490	1.00000	0.00490	1.00000	0.00490	1.00000	1.00000	1.00000				
	Alternative	0.00490	1.00000	0.00490	1.00000	0.00490	1.00000	1.00000	1.00000	0.00000	1	--	--
***fhl2b* locus**													
Haplochromini (*Psephi excluded)*	Null	0.04541	1.00000	0.04541	1.00000	0.04541	1.00000	1.00000	1.00000			86 (Q)	0.685
	Alternative	0.04532	1.00000	0.04532	1.00000	0.04532	1.00000	**76.33684**	**76.33684**	**7.805352**	**0.00520917**	150 (C)	0.954*
Haplochromini + Tropheini (*Psephi excluded)*	Null	0.04471	1.00000	0.04471	1.00000	0.04471	1.00000	1.00000	1.00000				
	Alternative	0.04588	1.00000	0.04588	1.00000	0.04588	1.00000	22.94605	22.94605	3.832012	0.05028254	--	--
Haplochromini + Tropheini (*Psephi included)*	Null	0.04471	1.00000	0.04471	1.00000	0.04471	1.00000	1.00000	1.00000				
	Alternative	0.04588	1.00000	0.04588	1.00000	0.04588	1.00000	22.94743	22.94743	3.832242	0.05027564	--	--
Tropheini	Null	0.04754	1.00000	0.04754	1.00000	0.04754	1.00000	1.00000	1.00000				
	Alternative	0.04768	1.00000	0.04768	1.00000	0.04768	1.00000	1.00000	1.00000	-0.004042	1	--	--
Pseele	Null	0.04498	1.00000	0.04498	1.00000	0.04498	1.00000	1.00000	1.00000			10 (C)	0.950*
	Alternative	0.04484	1.00000	0.04484	1.00000	0.04484	1.00000	635.39848	**635.39848**	**12.508918**	**0.00040501**	86 (L)	0.593
												150 (C)	0.998**
Cichlids (including Tilapia)	Null	0.03499	1.00000	0.03499	1.00000	0.03499	1.00000	1.00000	1.00000				
	Alternative	0.03998	1.00000	0.03998	1.00000	0.03998	1.00000	8.89460	8.89460	1.364914	0.242688	--	--

Supplementary Table 6 | Species (with corresponding GenBank accession numbers) used to infer the ND2 phylogeny for the positive selection analysis.

Species name	Cichlid Lineage	ND2 accession number
Astatoreochromis alluaudi	Haplochromini	AY930071.1
Astatotilapia burtoni	Haplochromini	AF317.266.1
Labidochromis caeruleus	Haplochromini	AY740383.1
Pseudocrenilabrus multicolor	Haplochromini	AY930106.1
Pseudotropheus tropheops	Haplochromini	AY740384
Cynotilapia pulpican	Haplochromini	-
Ctenochromis horei	Haplochromini	EU753935
Interochromis loocki	Haplochromini	JF900322
Petrochromis famula	Haplochromini	JF900324
Petrochromis fasciolatus	Haplochromini	JF900325
Petrochromis polyodon	Haplochromini	JF900326
Pseudosimochromis curvifrons	Haplochromini	GQ995777.1
Simochromis diagramma	Haplochromini	AY930087
Tropheus moori	Haplochromini	AY930093
Julidochromis marlieri	Lamprologini	AF398230
Julidochromis ornatus	Lamprologini	EF462229
Neolamprologus brichardi	Lamprologini	AF398227
Neolamprologus pulcher	Lamprologini	EF462244
Neolamprologus sexfasciatus	Lamprologini	HM623828
Cyprichromis 'jumbo'	Cyprichromini	AF317.266.1
Cyprichromis leptosoma	Cyprichromini	AF398224
Aulonocranus dewindtii	Ectodini	AY337782
Callochromis macrops	Ectodini	AY337795
Cyathopharynx furcifer	Ectodini	AY337781
Ophthalmotilapia ventralis	Ectodini	AY337774
Xenotilapia flavipinnis	Ectodini	AY337794
Oreochromis niloticus	Tilapini	AF317237
Oryzias latipes	non-cichlid teleost	NC_004387.1

Supplementary Table 7 | Testing for branch or site-specific positive selection on East African cichlids *fhl2a* and *fhl2b* with CODEML.

We found no evidence for branch or site-specific positive selection in the *fhl2a* copy while in *fhl2b*, we detected positive selection on three amino acids (positions 10, 86 and 150). Positive selection was only detected within one species, *Pseudotropheus elegans* (Pseele). Overall, *fhl2a* and *fhl2b* are under purifying selection, showing that coding sequences alone cannot explain the emergence or diversity of the egg-spot trait in haplochromines. All *fhl2a* and *fhl2b* coding sequences were found to be 837bp long (279 amino acids), except for three individuals from two species - *Tropheus moori* (843bp and 745bp, respectively) and *Simochromis diagramma* (840bp)

Briefly, under the null model, the foreground branch (branch of interest) has proportions of sites under neutral selection that may differ from those on the background branches. In the alternative model, positive selection is allowed on the foreground branch. For clarity's sake, we do not show the proportion of sites in each category, only the computed value of the corresponding *dn*/*ds* ratios (ω).
In both models and on both branches:

0: $\omega_0 < 1$

1: $\omega_1 = 1$

In the Null model:

2a: $\omega_2 = 1$ on foreground and $\omega_0 < 1$ on background.

2b: $\omega_2 = 1$ on foreground branch and $\omega_1 = 1$ on background.

In the Alternative model:

2a: $\omega_2 \geq 1$ on foreground and $\omega_0 < 1$ on background.

2b: $\omega_2 \geq 1$ on foreground and $\omega_1 = 1$ on background.

With ω_0: dn/ds < 1, ω_1: dn/ds = 1, ω_2: dn/ds ≥ 1. *LRT:* Likelihood Ratio Test computed as $2\times(\ln L_1 - \ln L_0)$ where L_1 is the Likelihood for the Alternative model and L_0 is the likelihood of the Null model. Under the Null model, the LRT follows a Chi-square distribution with 1 df. *Site:* positively selected amino-acid site with the amino acid change in brackets. *BEB:* Bayes Empirical Bayes.

Supplementary Table 8 | The AFC-SINE insertion is specific to egg-spot bearing haplochromines. In order to test whether the SINE insertion is correlated with the egg-spot phenotype we sequenced this transposable element region in 19 cichlid species, including both haplochromines and non-haplochromines. We also sequenced one haplochromine species that has no egg-spot but instead features a blotch on its anal fin: *Pseudocrenilabrus philander*. We confirmed that the AFC-SINE insertion is specific to the egg-spot bearing haplochromines, whereas the ancestral haplochromine *P. philander* does not have this insertion. The AFC-SINE element in the *fhl2b* promoter region was compared to the consensus sequence and available full-length AFC-SINE elements of cichlids. The insertion in the *fhl2b* promoter covers a full-length element and is flanked both 3' and 5' by five nucleotide long direct repeats, so called target site duplications. These duplications are the result of the element's insertion process via DNA strand break and repair[6], confirming that this is an insertion in haplochromines and not a deletion in non-haplochromines, since no remnants of these sites or the element are detected in the other species. The sequences are available under the accession numbers KM263999 to KM264016.

Species Name	**Lineage**	**Anal Fin Eggspot**	**AFC-SINE**
Astatotilapia burtoni	Haplochromini	Yes	Present
Astatotilapia stapersi	Haplochromini	Yes	Present
Astatoreochromis alluaudi	Haplochromini	Yes	Present
Pundamilia nyererei	Haplochromini	Yes	Present
Metriaclima zebra	Haplochromini	Yes	Present
Cynotilapia pulpican	Haplochromini	Yes	Present
Ctenochromis horei	Tropheini	Yes	Present
Lobochilotes labiatus	Tropheini	Yes	Present
Gnathochromis pfefferi	Tropheini	Yes	Present
Limnotilapia dardenii	Tropheini	Yes	Present
Pseudocrenilabus philander	Basal Haplochromini	No	Absent
Callochromis macrops	Ectodini	No	Absent
Ophthalmotilapia ventralis	Ectodini	No	Absent
Ophthalmotilatpia nasuta	Ectodini	No	Absent
Aulonocranus dewindtii	Ectodini	No	Absent
Perissodus microlepis	Perissodini	No	Absent
Neolamprologus sexfasciatus	Lamprologini	No	Absent
Julidochromis ornatus	Lamprologini	No	Absent
Lamprologus lemairii	Lamprologini	No	Absent
Limnochromis abeeli	Limnochromini	No	Absent
*Oreochromis niloticus**	Tilapini	No	Absent
Primer	**Sequence 5'-3'**	**TºC**	**Task**
Fhl2b_TEregion_fw	GAAGTCATGCAATGACAGACA	58-60	PCR TE region
Fhl2b_TEregion_rev	AATCCTCTGGGCAAAATGTGC	58-60	PCR TE region
Fhl2b_Hap_fw **	CTACTGGTGTTGGCCAGAGG	59-60	Seq TE region
Fhl2b_NonHap_fw **	TTAAAGTCATTAAAATGTCCCGATT	59-60	Seq TE region

*only available from genome not amplified in house

**primers used only for sequencing and not for product amplification, Hap only works for haplochromines and non-Hap only works for non-haplochromines

Supplemental References

1. Kasahara, M. *et al*. The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447,** 714–9 (2007).

2. Nakatani, Y., Takeda, H., Kohara, Y. & Morishita, S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res*. **17,** 1254–65 (2007).

3. Bessa, J. *et al*. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev*. *Dyn*. **238,** 2409–2417 (2009).

4. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. **33,** D501–4 (2005).

5. Meyer, A. & Van de Peer, Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* **27,** 937–945 (2005).

6. Wicker, T. *et al*. A unified classification system for eukaryotic transposable elements. *Nat*. *Rev*. *Genet*. **8,** 973–982 (2007).

Chapter 5

The role of parasites and the immune system in the adaptive radiation of Lake Tanganyika cichlids

Britta S. Meyer, Pascal I. Habluetzel, Anna Roose, Melinda Hoffman, Walter Salzburger and Joost A.M. Raeymaekers

to be submitted to: Molecular Ecology

Version: 25th November 2013

BSM sequenced the MHC genes; PIH, AR, MH, JAMR performed the parasitological survey; BSM, PIH, JAMR analysed the data; BSM, JAMR drafted the manuscript; JAMR and WS supervised the project.

INTRODUCTION

The species flocks of the East African cichlid fishes belong to the most powerful model systems to study the genetic basis of adaptation and diversification[1-3]. Especially the cichlid fishes of the Lake Tanganyika radiation are genetically and morphologically very diverse as this species flock consists of 12-16 different lineages[4, 5]. The processes, which could have lead to the emergence of so many species in such adaptive radiations, could have been manifold and are still debated intensively[4, 6-9]. The main idea is that their evolutionary success is due to an interaction of several extrinsic and intrinsic factors. Until recently allopatric speciation mechanisms were thought to be the prevalent trigger to promote speciation; and due to lake-level fluctuations in Lake Tanganyika, this mechanism seems likely and could have played potentially an important role. However it is also very likely that during the early lake formation many newly created habitats and therefore ecological niches could be filled. In the mechanism referred as "ecological speciation", divergent natural selection pressures on phenotypic traits (such as trophic apparatus, body size and shape, colouration and colour perception, acoustics and olfaction, as well as behaviour) contribute to the ecological divergence of the species[10, 11]. Besides the fact that sexual isolation can arise as a side effect, due to pleiotropic (linked) interaction between genes or traits, ecological speciation happens accelerated if actually the same traits are both under natural and sexual selection, e.g. if both the environment and the potential mate favour one specific trait ("magic trait" principle)[12]. Parasites represent an ecological pressure, which is predominant in all living animals. Since parasites influence survival and reproduction, it has been suggested that they could also influence adaptation, reproductive isolation and speciation (reviewed by[13]). These possibilities have been considered in a number of studies, including in cichlids. First, differences in parasite community composition have been shown for closely related cichlid species living in sympatry[14]. Accordingly, these species were also different for a specific set of genes involved in the immune system, the major histocompatibility complex (MHC) class II genes, suggesting an adaptive genetic basis for the differences in parasitism with the potential to promote reproductive isolation. Second, differences in parasitism have been described for allopatric colour morphs *within* a species[15]. This confirms that parasites represent one of the factors that might promote divergence in allopatry, or accelerate speciation upon secondary contact. Third, a study by [16] suggested that parasite-mediated sexual selection might contribute to the divergence of female mating preferences for male coloration, strengthening reproductive isolation.

While each of these studies hint at the importance of parasites at the micro-evolutionary level, no study has investigated parasite-driven speciation in cichlids at the macro-evolutionary level. Here, we explore the potential contribution of parasitism to the speciation processes of the Lake Tanganyika cichlid radiation. In this system, habitat and diet shifts have been proposed to be major drivers of adaptation and speciation[17]. However, habitat and diet shifts likely lead to the exposure to different parasites or differences in infection risk. As such, and in particular because parasites might influence natural as well as sexual selection, the Lake Tanganyika cichlid radiation might be to an important extent parasite-driven. We first evaluated whether there is covariation between trophic morphology, trophic level, diet, body shape, and parasitism within the LT cichlids. We expected different macroparasite communities among species, possibly between algae and invertebrate feeders (i.e. species that shifted in diet), or between generalists and sand and rock dwelling species (i.e. species that shifted in habitat). Second, we tested whether the species evolved immunogenetic differences by screening several loci of teleost MHC class II *B* genes. The resulting sequences were tested by means of molecular evolutionary analyses (e.g. *dN/dS* substitution rate ratio and genetic diversity within and among tribes).

MATERIAL AND METHODS

MHC screening

To access the genetic diversity within and among the different Lake Tanganyikan cichlid tribes and to get insights for the acting selection on the MHC genes in cichlids of Lake Tanganyika, a representative set of species was chosen focusing mainly on Lamprologini (11 species), Ectodini (five species) and Haplochromini (four Tropheini species, two basal riverine haplochromines and each four derived Lake Malawi and Lake Victoria haplochromines) (see table 1). Other tribes (as Eretmodini, Perissodini, Boulengerochromini, Cyphotilapiini, Cyprichromini, Limnochromini, Trematocarini and Oreochromini) were included as well but with only one or respectively two species. The sampling of these specimen was conducted in the years 2007 and 2013, independently from the parasite screening, using the standard sampling procedure[17]. From each species about 8 to 16 individuals were used for genotyping.

In teleost fishes, the MHC is spread over three chromosomes. One chromosome contains the MHC class I. Whereas the teleost MHC class II loci are spread across two different chromosomes, which are called class II*a* and class II b genes. Each of these regions encompasses two separate subclasses of genes, the MHC class II A and B genes, coding for different two different chains within the MHC molecules, which are typically arranged in pairs along the MHC. The classII*a* genes are more conserved in teleost fish, whereas the classII*b* genes are more diverse and have undergone several rounds of duplication[18]. Here, we focused on the Mhc class II genes (D) of the subclass *b* and only on the genes for the beta chain (B), which are located in the five genomic regions (B, C, D, E, F) defined by [18] and which were named DBB-DFB. For the amplification of these cichlid Mhc loci the forward primer TU383 (CTCTTCATCAGCCTCAGCACA) and the reverse primer TU377 (TGATTTAGACAGARKGKYGCTGTA) from the study of [19] were used. This primer pair is known to amplify the intron 1 and parts of exon 2 of up to 17 homology groups [19] (note that in another study up to eight expressed putative loci were find with this primer set [14]).

The PCR amplification of the MHC was conducted with a multiplex approach in a final volume of 25 µL of the Multiplex PCR Kit (Qiagen, Hombrechtikon, Switzerland), the MHC specific barcoded fusion primers (0.1 µM of each primer) (see conditions in Appendix 1, 3 of this thesis) and DNA of the different species. Standardised PCR conditions started with an initial heat activation phase (necessary for the HotStarTaq DNA Polymerase) of 95 °C, and continued with 35 amplification cycles consisting of 30 s of denaturation at 94 °C, 90 s of annealing phase at 60 °C and an extension phase of 90 s at 72 °C. The PCR was terminated with a final extension phase of 10 min at 72 °C. The PCR products were purified with the magnetic bead system of Agencourt AMPure XP (Beckman Coulter, Nyon, Switzerland). Before the pyrosequencing step (454 with GS FLX system, Roche, conducted by Microsynth, Balgach, Switzerland), quality of the PCR products was assessed using the 2100 Bioanalyzer (Agilent, Basel, Switzerland).

The generated raw reads were processed with Roche's demultiplexing and converting tools (sffinfo, sfffile). For quality filtering we applied a filter for too short reads (≤ 150 bp), only allowed 1% of ambiguous bases (N) and filtered out low quality sequences (Mean ≥ 15). These sequences were imported species-wise into Geneious (v3.7. Biomatters Ltd, www.geneious.com) and de novo assembled. This resulted in contigs with highly identical reads. If more than 3 bp of a read were different than the rest of the consensus, the reads were excluded. Strict (50 %-) consensus sequences (here called alleles per individual) were generated within

Geneious (v3.6), with the option of assigning the highest quality and calling a base "N", if it had a lower PHRED score than 20. These alleles were aligned using MAFFT (--auto)[20] and insertions of ambiguous "N" and misalignments were manually excluded. A blast search with the obtained alleles let to the exclusion of further sequences. Identical or highly similar (from low coverage contigs with less than 4 reads) alleles were collapsed to haplotypes using again the de novo algorithm within Geneious - thus with very strict settings (only allowing maximal 1 % mismatches and the maximal 2 % of gaps per read). These haplotypes were used for the following analyses.

To check how many different groups were amplified with the two primers, we generated three neighbor-joining (NJ) trees [21] using Jukes-Cantor (JC) substitution model[22]: with the intron 1 and exon 2 as an entity, further only of the exon and finally of the translated amino acid sequences using Geneious. The three tree estimates were compared to (i) investigate the power of resolution of the amino acids sequence and (ii) to investigate the influence of the intron in the process of group assignment. The resulting groups were numbered with Latin numbers. We applied the "homology group" (DBB-DFB-like) assignment (see in [23]) to our data set. Therefore we just combined the alignments of both studies and transferred their homology group assignment to our analysis.

The genetic diversity was estimated with the software package MEGA (v5.2.2) at different levels[24]. We mixed all haplotypes of all groups and calculated the evolutionary divergence within tribes. This was done on intron-exon (~450-500 bp) and exon (241 bp) level. Gaps and missing data was eliminated in a pairwise comparison. Both the absolute number of differences and the uncorrected p-distance were recorded.

To investigate the effects of the MHC diversity of species in reference to the parasites (see below) we performed among-species comparisons, where we calculated the genetic distance as a between group average. Therefore we estimated the distance (uncorrected p-distance) of the exon (first, second and third codon together; 241 positions). Finally, we calculated the distance of amino acid sequences using the Jones-Taylor-Thornton (JTT) model. The JTT model, an empirical substitution model, corrects for multiple substitutions based on a model for amino acid substitutions using the substitution-rate matrix[25]. We excluded species with a very low amount of number of MHC reads.

The genetic distance (as the uncorrected p-distance for exon and amino acid translation) was also evaluated in our defined "homology" groups. We also counted the number of haplotypes per species occurred within these groups. Further we estimated the dn/ds ratio of all sites with SLAC, a maximum likelihood analysis[26], and thus could the number and location of positively and negatively selected sites to see if our sequenced MHC genes are under selection and thus functional.

Parasitological survey

A subset of 21 species included in the genomic survey along with two additional species (*S. diagramma* and *C. furcifer*) were screened for metazoan ecto- and endoparasites. For this purpose, sampling was conducted at Toby's place on the Zambian shoreline of Lake Tanganyika, i.e. the same site as for the genomic survey. While most species were obtained in August 2012, *S. diagramma* and *H. microlepis* were captured in August 2011 and July 2013, respectively. One species, *A. burtoni*, was obtained in July 2013 at Kapata, which is about 20 km more southward. About ten individuals per species were caught by chasing fish into standing nets. After capture

the fish were kept in tanks of 0.8 m x 0.8 m x 1.2 m depth or 0.8 m x 0.8 m x 2 m depth. Before usage, tanks were cleaned, dried and filled with lake water.

All fish were dissected in the field within four days after capture. The day of dissection (0, 1, 2 or 3 days after capture) was recorded in order to keep track of changes in parasitological parameters while the fish were kept in the tanks. Individual fish were killed with an overdose of MS222. The parasitological survey consisted of three parts. First, the outer surface and the mouth cavity of the fish were inspected for ectoparasitic monogeneans and crustaceans (copepods, branchiurans, isopods), bivalves, any kind of helminthic cyst. Second, the four gill branches on the left were dissected and stored on 100% analytical ethanol (EtOH), and later on screened for ectoparasitic monogeneans, crustaceans (copepods and branchiurans), bivalves, any kind of helminthic cyst. Third, fish were screened for intestinal monogeneans, digeneans, acanthocephalans, nematodes, and any kind of helminthic cysts. To do so, stomach, intestines, gall and urinary bladder were dissected and inspected with in a petridish with lake water. Finally, the sex of the fish was determined by visual inspection of the genital papilla and gonad development.

The parasitological survey was performed with a stereomicroscope and different observers. Observers were recorded in order to keep track of observer bias. The outer surface and the mouth cavity of the fish was screened by a single observer. The gills were screened by two observers in 2011, two observers in 2012 and two observers in 2013. The intestines were screened by three observers in 2011, four observers in 2012, and a single observer in 2013.

All parasites were counted and identified to genus or class level and preserved as follows. Monogeneans were isolated using dissection needles and were either mounted on slides in ammonium picrate glycerine for further morphological research, or stored on 100 % EtOH. Acanthocephalans and nematodes were stored on 80 % EtOH, while intestinal monogeneans, branchiurans, copepods, any kind of helminthic cysts, bivalves and unknown groups were stored on 100 % EtOH.

Data analysis

The analyses of infection levels were performed in the statistical package R[27]. Prevalence and mean abundance were calculated for each group of parasites and each host species following the terminology of Rosza et al.[28]. Manova was used to test for differences in infection levels (quantified either as prevalence or mean abundance) between cichlid tribes for all parasite groups together. Subsequently, Kruskal-Wallis Anova's were used to test for differences in infection levels between the tribes for each parasite group separately.

The level of covariation between body shape, trophic morphology, diet and parasitism within LT cichlids was investigated by a Spearman rank correlation analysis between infection levels and data collected previously by Muschick et al.[17]. Specifically, body shape was included in the analysis as the two first principal components of body shape variation, as calculated by Muschick et al.[17] from a geomorphometric analysis. Likewise, trophic morphology was included using the two first principal components of lower pharyngeal jaw shape variation as calculated by Muschick et al. [17]. Diet was included as the two first principal components on proportional diet data, as well as the two first principal components calculated from these data. In addition to body shape, trophic morphology and diet, we also included carbon and nitrogen stable isotope signatures (δ13C and δ15N), which are a proxy for trophic ecology[17]. In particular, δ13C values in LT cichlids were found to be correlated with body shape clusters, whereas δ15N values

correlate with the shape of the lower pharyngeal jaw. As such the δ13C and δ15N respectively reflect variation between macrohabitats (e.g. benthic versus pelagic) and the relative trophic level of an organism.

To further investigate which source of information (body shape, trophic morphology, diet or isotopes) was most strongly linked to infection levels for all parasite groups simultaneously, we performed a redundancy analysis (RDA). RDA is a canonical extension to PCA in which the principal components produced are constrained to be linear combinations of a set of predictor variables[29]. It enables to identify the best ordination model that describes parasite community similarities among cichlid species. This analysis was performed with the R library vegan. Significance of the proportion of variation in infection levels explained by each source of information was calculated and tested for significance using 1000 random permutations. For each source of information separately, the RDA analysis was preceded with a forward selection procedure as implemented in the "packfor" package in R[30]. Forward selection corrects for highly inflated type I errors and overestimated amounts of explained variation.

MHC vs. parasites, trophic ecology, morphology and isotope signatures

To investigate which source of information (body shape, trophic morphology, diet, isotopes or infection levels) was significantly linked to MHC divergence between cichlid species, we performed a redundancy analysis (RDA; see above) using the R library vegan. Infection levels were included in the analysis either as prevalence or mean abundance. Since the set of cichlid species for which parasite data were available was smaller than for the other data, two separate analyses were performed. The MHC data were included in the analysis as the first ten axis obtained from a classical multidimensional scaling analysis (CMDS) on MHC-based genetic distances, either based on exon 2 or on amino acids using the JTT model (see above). Significance of the proportion of variation in MHC profiles explained by each source of information was calculated and tested for significance using 1000 random permutations. For each source of information separately, the RDA analysis was preceded with a forward selection procedure as implemented in the packfor package in R[30].

RESULTS

Sequencing success

In total we received, after our filtering procedure, 11,569 reads of MHC(-like) genes using pyrosequencing and the barcoded fusion-primers. The filtered mean read lengths was 289.62 ± 90.31 bp. Overall the sequencing coverage was best for the three major tribes, namely the Lamprologini, the Ectodini and Haplochromini (Table 1). Whereas sequencing nearly completely failed in the basal lineages of Lake Tanganyika cichlid radiation (*Oreochromis tanganicae, Bathybathes graueri* and *Boulengerochromis microlepis*. Other species with few reads were *Eretmodus cyanostictus and Gnathochromis pfefferi*. After we defined alleles within individuals (data not shown here) and species, we blasted these. This resulted in 1198 MHC alleles, 201 sequences of "*Haplochromis burtoni* dual specificity protein phosphatase CDC14A-like OR SINEs", 26 alleles without any BLAST hit, 25 sequences of "*Haplochromis burtoni* zinc finger protein 513-like" and 14 miscellaneous sequences. Following our strategy

of filtering, processing and collapsing of alleles, we resulted in 573 different haplotypes within 39 "species" (37 plus Lake Malawi and Victoria haplochromines, which are not distinguished separately, due to their low read coverage and high relatedness). Note, that in Table 1 are more haplotypes, as some closely related species share exact alleles.

All three NJ-tree estimates created the same clusters of MHC genes (called homology group from here on), only differing in relationship to each other. However, as we were not interested in the evolutionary history of these genes at this stage, we decided to choose the exon-based NJ-tree estimate to represent the grouping. In total, we can visually distinguish six major groups and some substructuring (groups 1-6; in Figure 1). A combination and comparison (with alignment and NJ-trees) to the dataset of Hablützel et al.[23] revealed that our used primer pair successfully amplifies two DEB-like groups, three DFB-like groups, and to some other unnamed groups (see Table 2). However we could not assign any of our haplotypes to DBB-like, DCB-like or DDB-like groups.

The average genetic diversity tribe is shown in Table 3. This table shows besides the haplotype number, that the overall genetic distance within the three tribes with more species (but also higher read coverage and more defined haplotypes) is very similar (ranging around 0.2).

As we were focusing in detail to the homology groups, defined in this study as group 1-6, we realised that not all loci were amplified in all tribes equally (Table 2). The average genetic distance within the groups ranged between 0.077 (group 2b) and 0.166 (group 1). (Note that the low value could be due to the limited number of reads, whereas the higher number could be due to the fact that we possibly summarized two subgroups two group 1, also see Figure 1). In all groups we could detect positively and negatively sites, which differed more or less between the defined groups: closely associated groups have the tendency to have more positions in common with both positive and negative selection signs (as in the case of group 6a and 6b). However there are positions, which are shared over several homology groups.

Infection patterns

Manova revealed significant differences between Lake Tanganyika cichlid tribes for the prevalence of metazoan ecto- and endoparasites (Wilks' lambda = 0.0066, $F_{28,40}$ = 1.96, P = 0.0322). These differences were mainly due to the prevalence of acanthocephalans which was high in Tropheini, intermediate in Ectodini, and low in Lamprologini and Perissodini (Table 4, Figure 2). There was no multivariate difference between the tribes for the mean abundance of parasites (Wilks' lambda = 0.027, $F_{28,40}$ = 1.12, P = 0.38), but at the univariate level there were differences for the mean abundance of acanthocephalans and *Cichlidogyrus* sp. (Table 4). Mean abundance for both groups of parasites was again high in Tropheini, intermediate in Ectodini, and low in Lamprologini and Perissodini (Figure 3). Accordingly, a two-dimensional (PCA-based) representation of parasite communities (Figure 4) revealed partially non-overlapping parasite communities in the Lamprologini and Tropheini, while parasite communities in the Ectodini show similarities with both the Lamprologini and the Tropheini.

Spearman rank correlations revealed that infection levels across cichlid species were correlated with trophic ecology, morphology or isotope signatures. The prevalence of acanthocephalans increased with δ13C and the proportion of sand and aufwuchs in the diet, and decreased with δ15N (Table 5; Figure 5). The prevalence of nematodes increased with δ15N, and decreased with the proportion of mollusks in the diet (Table 5; Figure 5). The prevalence of *Urogyrus* sp. increased with the proportion of aufwuchs in the diet, while the prevalence of cysts in the gills decreased with the proportion of fish in the diet. The prevalence of cysts in the fins

and *Cichlidogyrus* sp. were respectively correlated with the second principal component of body shape variation, and the second principal component of lower pharyngeal jaw shape variation. Spearman rank correlations with mean abundance confirmed these patterns (Table 6). In addition, there was an increase of *Enterogyrus* sp. with the proportion of fish in the diet, an increase of *Ergasilus* sp. with the proportion of arthropods, and an increase of cysts in the fins with the proportion of fish scales.

Forward selection followed by RDA identified a significant effect of the second principal component of body shape variation on the entire parasite community, accounting for 11% of the variation in prevalences (RDA: $F_{1,19}$=2.27; P = 0.028). A significant effect of the second principal component of lower pharyngeal jaw shape variation on the entire parasite community was identified when the analysis was performed on mean abundance (RDA: $F_{1,19}$=3.83; P = 0.0325). This model accounted for 17% of the variation. Other variables accounting for trophic ecology, morphology or isotope signatures were not selected in these models.

MHC vs. parasites, trophic ecology, morphology and isotope signatures

A two-dimensional (MDS-based) representation of the MHC divergence between cichlid species revealed partial overlap between Tropheini, Ectodini and Lamprologini at the exon 2 as well as the amino acid level (Figure 3). The smaller tribes (Perissodini, Cyphotilapiini, Cyprichromini, and Trematocarini) seem to have different MHC profiles, especially at the amino acid level. Forward selection identified significant effects of the prevalence of acanthocephalans and *Gyrodactylus* sp. on the MHC exon 2 divergence. The combined model explained 31% of the variation at the exon 2 level (RDA: $F_{2,12}$=2.66; P = 0.036). At the amino acid level, only the prevalence of acanthocephalans showed a significant effect, accounting for 13% of the variation (RDA: $F_{1,13}$=2.02; P = 0.02). Similar results were obtained with mean abundance. Variables quantifying trophic ecology, morphology or isotope signatures did not explain significant variation at the exon 2 or amino acid level (results not shown).

DISCUSSION

MHC diversity

Here we present the first MHC sequences for East African cichlids other than Haplochromini (including Tropheini) and Oreochromini (note: there are two sequences of Cyphotilapiini in GenBank). We thus show that the primers by Malaga-Trillo et al.[19] are able to cover the whole phylogenetic range of Lake Tanganyika cichlid tribes. In this study we successfully amplified at least five homology groups of MHC classII*b* B genes, of which all showed signs of acting selection pressures and none of them comprising a stop codon. This let us to the conclusion that the here investigated genes are expressed and consequently relevant for the immune system. In Halblützel et al.[23] not all these groups were amplified in that degree and thus we add a new dimension of MHC gene diversity for Lake Tanganyika cichlids. It is likely that we only see the tip of the iceberg, as we could not amplify DBB, DCB and DDB-like loci. As we determined selection on differing positively (potentially antigen binding sites) and conserved negatively positions, we assume that each homology group plays a different role in the immune-defence. We would like to extend the MHC genotyping to more MHC class II*b* B loci and to increase the coverage, this could happen with the primer set developed in Hablützel et al.[23]. Sampling more intensively within a tribe or closely related taxa one would additionally gain insights to the

dynamic processes underlying the creation and maintenance of this diverse multigene family (e.g. the birth-and-death-model, the frequency of pseudogenes, the frequency of duplications or gene-conversions, amount of inter and intralocus exchange)[31].

Our NJ trees produced of intron-exon and exon alignments revealed no extreme incongruence in the assignment of homology groups. However, a study by Figueroa et al. showed that the phylogeny of the exon with that of the flanking introns were mismatching[32], it would be worth to re-analyse our data in detail, in order to check for this phenomenon. The authors assume that this could be due to convergent evolution in the exons and/or homogenization processes on the introns[32]. These processes are known from human HLA studies, were the introns flanking the polymorphic exons of some human MHC genes have been become relatively conserved due to intra-locus recombination and thus homogenization[33].

In other studies on the MHC of East African cichlid, a high degree of "trans-species polymorphisms", the long-term retention of allelic lineages after speciation, could be detected [32, 34]. However, these studies only included a very small subsample of closely related cichlids, mainly from the relatively young haplochromines, where shared polymorphism due to incomplete lineage sorting are also known in other genes. Nevertheless, we also observed allele sharing within tribes (based on exon-intron data). This was the case for two derived haplochromine (*Astatotilapia burtoni* and *Haplochromis cf. stappersii*), some closely related lamprologines (e.g. *Neolamprologus tetracanthus* and *N. modestus* and *N. pulcher*, and between *Lamprologus callipterus* and *Lepidiolamprolgus elongatus*) and between a known hybrid species (*Telmatochromis dhonti*) and other lamprologines. It remains to be tested if this allele sharing also occurs across tribes, if we only look at the exon level or amino acid level. If also balancing selection pressure or only incomplete lineage sorting is responsible for this ancestral polymorphisms, is not clear.

Covariation between trophic morphology, body shape, trophic ecology and parasitism

Habitat adaptation and trophic adaptation have been proposed to be important drivers of the Lake Tanganyika cichlid radiation, because differences in habitat and diet between species are associated with morphological differences. For instance, Muschick et al.[17] observed that diet is associated with pharyngeal jaw morphology, a key trait for feeding on a specialized diet (from algae and biofilms to invertebrates and fish), while habitat is associated with differences in body shape. Likewise, carbon and nitrogen stable isotope signatures, which are a proxy for trophic ecology, are also correlated with morphology[17]. In particular, δ13C values in LT cichlids correlate with body shape clusters, whereas δ15N values correlate with the shape of the lower pharyngeal jaw. As such the δ13C and δ15N respectively reflect variation between macrohabitats (e.g. benthic versus pelagic) and the relative trophic level of an organism.

Despite these potentially strong drivers of adaptation in Lake Tanganyika cichlids, we anticipated that habitat and diet shifts would also imply exposure to different parasites or shifts in infection risk. Such shifts would likely happen simultaneously, so there would be no particular order of adapting first to a novel habitat or diet before adapting to novel parasites, or vice versa. A number of findings in this study show that habitat and diet are indeed correlated with parasitism. For instance, δ13C was correlated with the prevalence of acanthocephalans. This suggests that benthic cichlid fishes have a higher risk of infection with acanthocephalans than pelagic species. Likewise, δ15N was correlated with the prevalence of nematodes, suggesting that species at a higher trophic level have a higher risk of infection with nematodes. Various parasites were also correlated with the proportion of prey items as well as with morphology. Overall, body shape and lower pharyngeal jaw shape seemed to explain the most variation in infection levels. This is

perhaps not surprising given the indications that these traits enable cichlids to occupy different niches - hence running into different parasites. This also explains why different cichlid tribes harboured partially non-overlapping parasite communities.

MHC vs. parasites, trophic ecology, morphology and isotope signatures

Together, the above results imply that the potential contribution of parasites to the radiation of Lake Tanganyika cichlids should not be overlooked. Parasites are known to represent a strong selective force that can lead to behavioural as well as immunological adaptations. In theory, parasitism could be even more important than habitat and diet shifts, because parasites are known to not only influence survival, but also mate choice. This combines the effect of natural and sexual selection, and has the potential to accelerate speciation[35].

In fishes, there is growing evidence that parasites can play such a role, mediated in particular through the immune system[36]. If so, it is expected that species with similar infections have similar MHC profiles. We indeed observed a match between MHC profiles and infection with certain parasite groups, in particular acanthocephalans. However, these results were not corrected for neutral genetic divergence between species. We therefore cannot exclude at this stage that species with similar infections show similar MHC profiles because of phylogenetic relatedness. Another uncertainty at this stage is whether the observed MHC diversity is representative, and allows to quantify these relationships. The same is true for our assessment of parasite diversity, since our focus was only on metazoan macroparasites and since taxonomic identifications were done with a low resolution. The few detailed parasitological studies that exist for Lake Tanganyika all hint at a huge diversity of parasite species[37, 38].

Conclusion

We showed that different cichlid tribes harboured partially non-overlapping parasite communities as well as partially non-overlapping MHC diversity. The trophic-morphological axis of diversification in Lake Tanganyika cichlids is strongly correlated with infection levels of metazoan macroparasites. As such the potential contribution of parasitism and immunogenetic adaptation to this radiation should be not be overlooked.

ACKNOWLEDGEMENTS

We thank Lawrence Makasa, Danny Sinyinza, Gilbert Sheltons, Christian Sturmbauer, Bernd Egger, Olivia Roth, Isabel Keller and the staff of the Lake Tanganyika Research Station in Zambia for advice and help with fieldwork and logistics. Sandra Camey kindly assisted with dissections and parasite surveys. Moritz Muschick kindly shared ecological data. Research was supported by grants from the Research Foundation – Flanders (FWO grant project G.0553.10), the Janggen-Pöhn-Stiftung, the Flemish Interuniversity Council (VLIR), and the University of Leuven (KU Leuven Centre of Excellence PF/10/07). JAMR received a EU Marie Curie Fellowship (IEF 300256). BSM received funding by the University of Basel, the Freiwillige Akademische Gesellschaft Basel and the Burkhardt-Bürgin Stiftung, Basel.

REFERENCES

1. Seehausen, O. African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society of London - Series B: Biological Sciences* **273**, 1987-98 (2006).

2. Kocher, T.D. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics* (2004).

3. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology* **18**, 169-85 (2009).

4. Koblmüller, S., Sefc, K.M. & Sturmbauer, C. The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. *Hydrobiologia* **615**, 5-20 (2008).

5. Salzburger, W., Meyer, A., Baric, S., Verheyen, E. & Sturmbauer, C. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Systematic Biology* **51**, 113-35 (2002).

6. Todd Streelman, J. & Danley, P. The stages of vertebrate evolutionary radiation. *Trends in Ecology & Evolution* (2003).

7. Seehausen, O. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198-207 (2004).

8. Yoder, J.B. et al. Ecological opportunity and the origin of adaptive radiations. *J Evol Biol* (2010).

9. Wagner, C.E., Harmon, L.J. & Seehausen, O. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**, 366-9 (2012).

10. Rundell, R.J. & Price, T.D. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends Ecol Evol* **24**, 394-9 (2009).

11. Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737-41 (2009).

12. Servedio, M.R., Doorn, G.S.V., Kopp, M., Frame, A.M. & Nosil, P. Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution* **26**, 389-397 (2011).

13. Karvonen, A.S., O. The Role of Parasitism in Adaptive Radiations-When Might Parasites Promote and When Might They Constrain Ecological Speciation? *International Journal of Ecology* **2012**, 20 (2012).

14. Blais, J. et al. MHC adaptive divergence between closely related and sympatric African cichlids. *PLoS ONE* **2**, e734 (2007).

15. Raeymaekers, J.A. et al. Contrasting parasite communities among allopatric colour morphs of the Lake Tanganyika cichlid Tropheus. *BMC Evol Biol* **13**, 41 (2013).

16. Maan, M.E., Van Rooijen, A.M.C., Van Alphen, J.J.M. & Seehausen, O.L.E. Parasite-mediated sexual selection and species divergence in Lake Victoria cichlid fish. *Biological Journal of the Linnean Society* **94**, 53-60 (2008).

17. Muschick, M., Indermaur, A. & Salzburger, W. Convergent Evolution within an Adaptive Radiation of Cichlid Fishes. *Current Biology* **22**, 2362-2368 (2012).

18. Sato, A., Dongak, R., Hao, L., Shintani, S. & Sato, T. Organization of Mhc class II A and B genes in the tilapiine fish Oreochromis. *Immunogenetics* **64**, 679-690 (2012).

19. Málaga-Trillo, E. et al. Linkage relationships and haplotype polymorphism among cichlid Mhc class II B loci. *Genetics* **149**, 1527-37 (1998).

20. Katoh, K. & Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-80 (2013).

21. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-25 (1987).

22. Jukes, T.H. & Cantor, C.R. in Mammalian protein metabolism (ed. Munro, M.N.) 21-132 (Academic Press, 1969).

23. Hablützel, P.I., Volckaert, F.A., Hellemans, B. & Raeymaekers, J.A. Differential modes of MHC class IIB gene evolution in cichlid fishes. *Immunogenetics* **65**, 795-809 (2013).

24. Tamura, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9 (2011).

25. Jones, D.T., Taylor, W.R. & Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275-82 (1992).

26. Pond, S.L. & Frost, S.D. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-3 (2005).

27. R Development Core Team 2012. (R Foundation for Statistical Computing, Vienna, Austria, 2012).

28. Rozsa, L., Reiczigel, J. & Majoros, G. Quantifying parasites in samples of hosts. *J Parasitol* **86**, 228-32 (2000).

29. Legendre, P. & Legendre, L. Numerical Ecology (Elsevier, 2012).

30. Dray, S. et al. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* **82**, 257-275 (2012).

31. Nei, M., Gu, X. & Sitnikova, T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences* **94**, 7799-7806 (1997).

32. Figueroa, F., Mayer, W., Sültmann, H. & CO'hUigin. Mhc class II B gene evolution in East African cichlid fishes. *Immunogenetics* (2000).

33. Cereb, N., Hughes, A.L. & Yang, S.Y. Locus-specific conservation of the HLA class I introns by intra-locus homogenization. *Immunogenetics* **47**, 30-36 (1997).

34. Klein, J., Sato, A. & Nikolaidis, N. MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annu Rev Genet* **41**, 281-304 (2007).

35. Gavrilets, S. Fitness Landscapes and the Origin of Species (Princeton University Press, 2004).

36. Eizaguirre, C., Lenz, T.L., Kalbe, M. & Milinski, M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nat Commun* **3**, 621 (2012).

37. Vanhove, M.P.M., Snoeks, J., Volckaert, F.A.M. & Huyse, T. First description of monogenean parasites in Lake Tanganyika: the cichlid Simochromis diagramma (Teleostei, Cichlidae) harbours a high diversity of *Gyrodactylus* species (Platyhelminthes, Monogenea). *Parasitology* **138**, 364-380 (2011).

38. Gillardin, C., Vanhove, M.M., Pariselle, A., Huyse, T. & Volckaert, F.M. Ancyrocephalidae (Monogenea) of Lake Tanganyika: II: description of the first *Cichlidogyrus* spp. parasites from Tropheini fish hosts (Teleostei, Cichlidae). *Parasitology Research* **110**, 305-313 (2012).

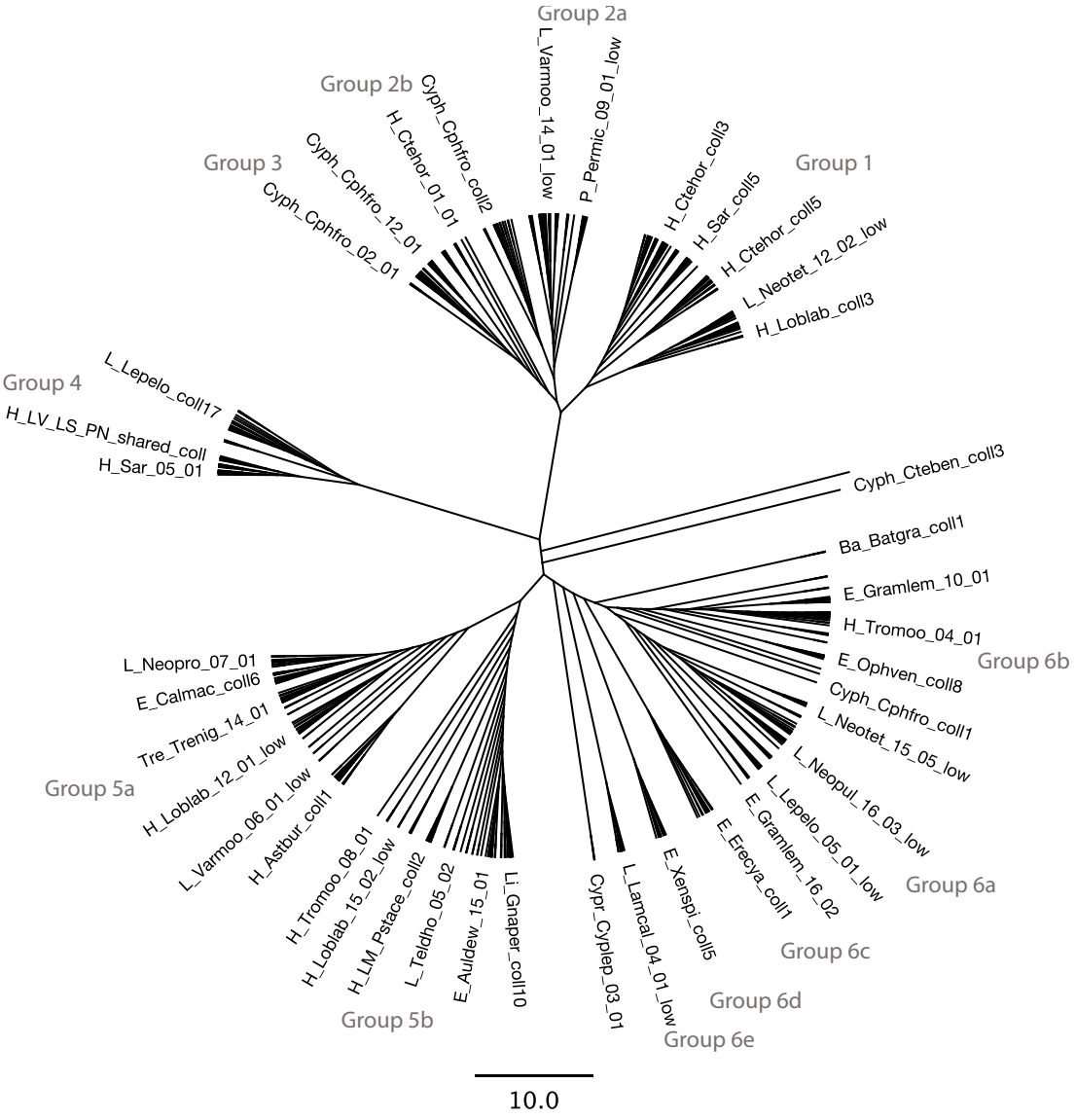


Figure 1. A NJ tree based on exon 2 sequence data of MHC genes. Groups are indicated with names 1-6. Subgroups are named a-e. This naming corresponds to Table B.

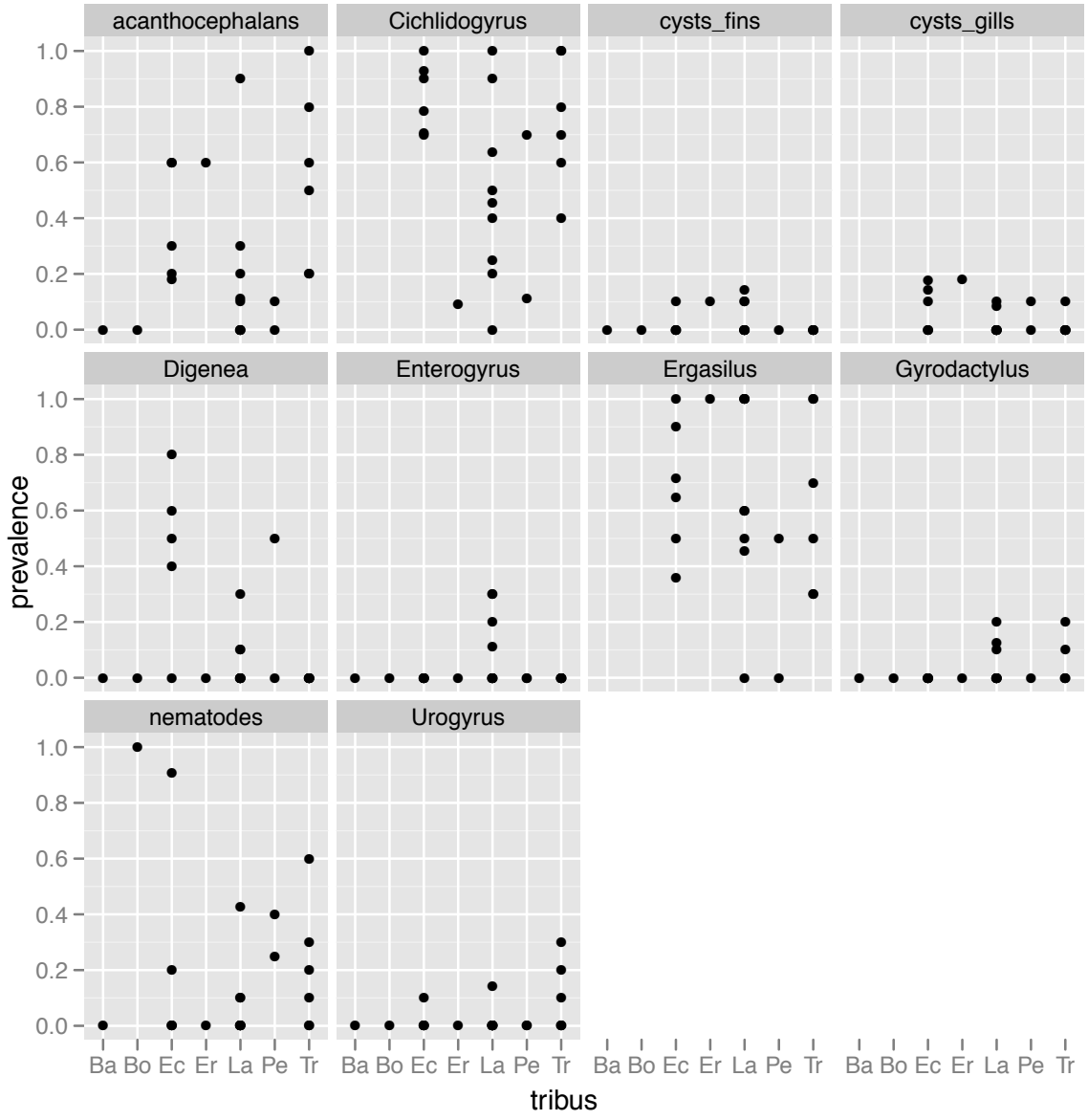


Figure 2. Prevalence of ten groups of endo- and ectoparasites by cichlid tribe. Ba: Bathybatini; Bo: Boulengerochromini; Ec: Ectodini; Er: Eretmodini; La: Lamprologini; Pe: Perissodini; Tr: Tropheini.

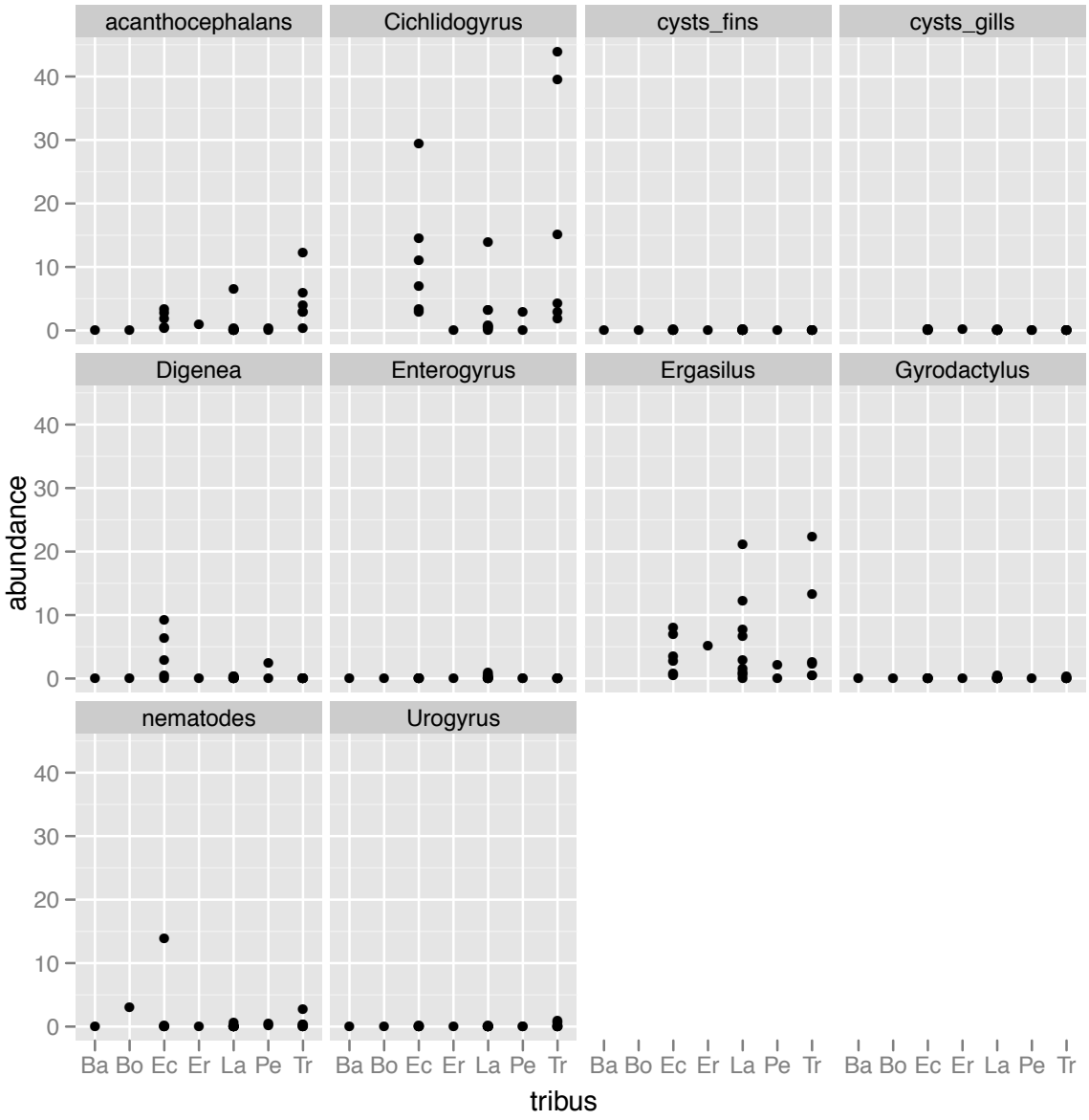
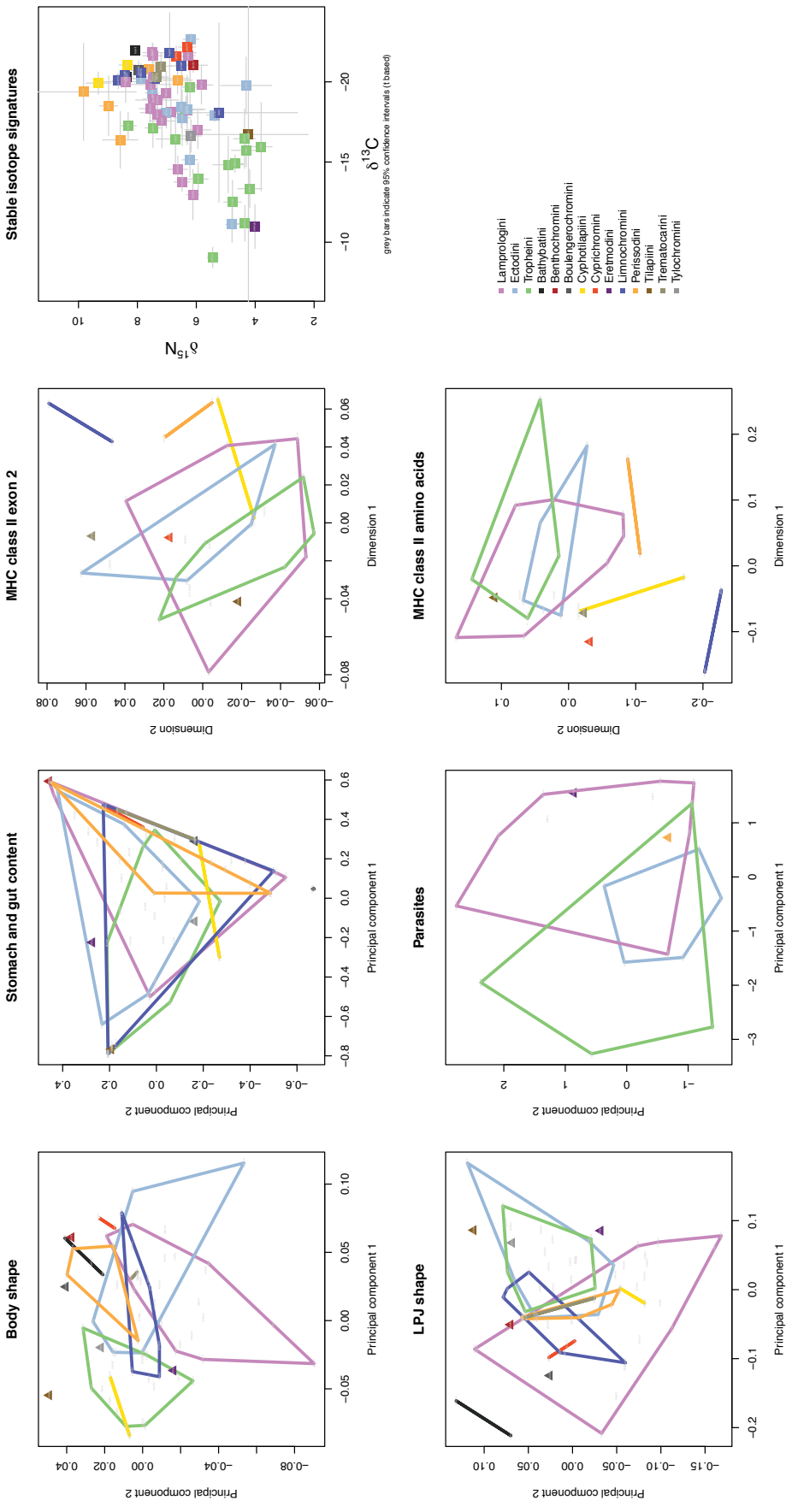


Figure 3. Mean abundance of ten groups of endo- and ectoparasites by cichlid tribe. Ba: Bathybatini; Bo: Boulengerochromini; Ec: Ectodini; Er: Eretmodini; La: Lamprologini; Pe: Perissodini; Tr: Tropheini.


Stable isotope signatures
δ15N
δ13C
grey bars indicate 95% confidence intervals (1 based)
Lamprologini
Ectodini
Tropheini
Bathybatini
Benthochromini
Boulengerochromini
Cyphotilapiini
Cyprichromini
Eretmodini
Limnochromini
Perissodini
Tilapiini
Trematocarini
Tylochromini
MHC class II exon 2
MHC class II amino acids
Stomach and gut content
Parasites
Body shape
LPJ shape
Dimension 1
Dimension 2
Principal component 1
Principal component 2

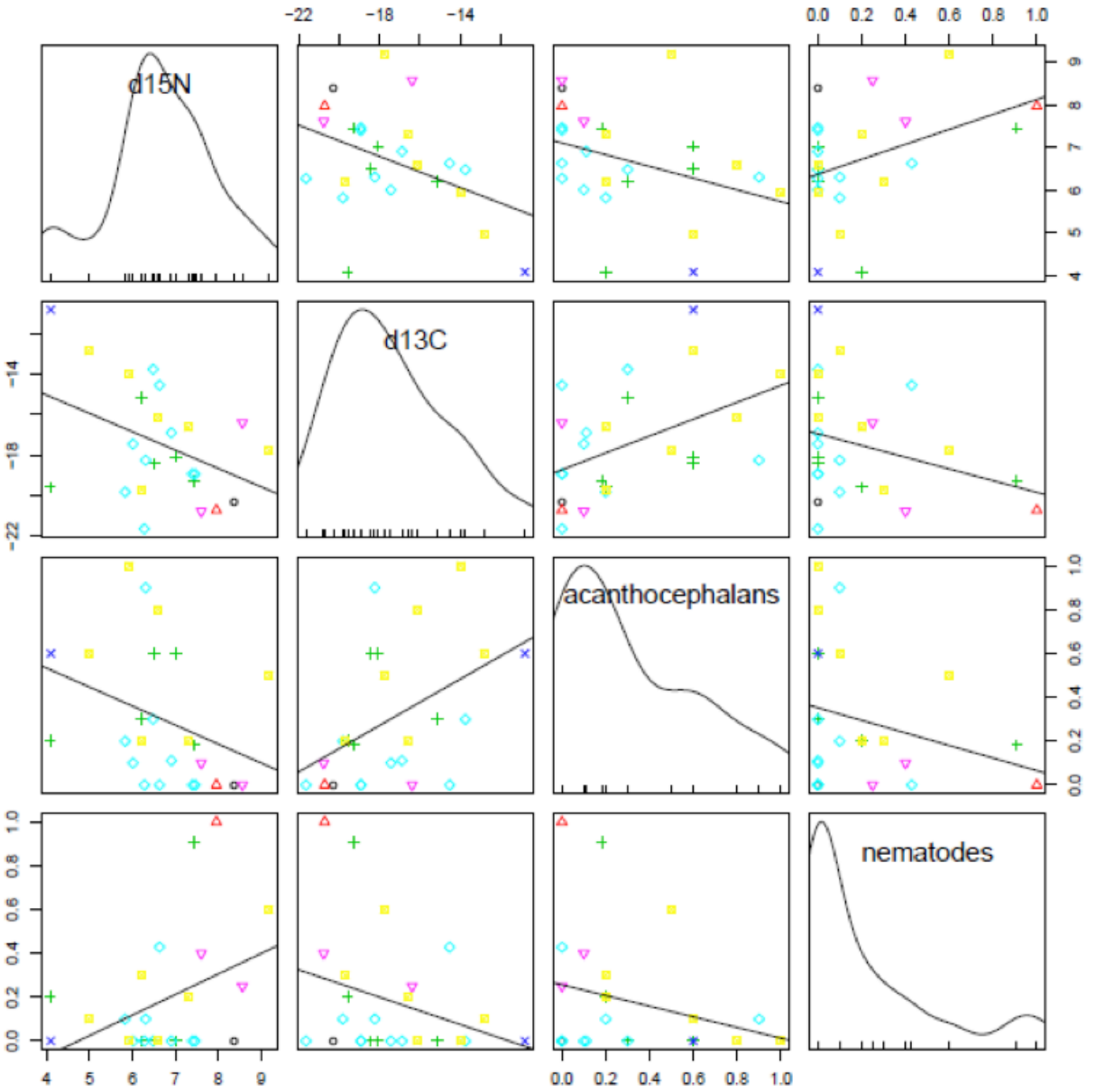


Figure 5. Scatterplot matrix of isotopes signatures (δ15N and δ13C) versus parasitological parameters (prevalence of acanthocephalans and nematodes). The various symbols represent different tribes (circles: Bathybatini; upper triangles: Boulengerochromini; +: Ectodini; x: Eretmodini; diamonds: Lamprologini; lower triangles: Perissodini; squares: Tropheini).

Next before:

Figure 4. Two-dimensional representations of morphology, trophic ecology, infection levels, MHC-based genetic divergence, and isotope signatures. Principal component analysis (PCA) of (A) body shape and (B) LPJ shape, (C) proportional stomach and gut contents, and (D) mean square-root transformed abundance of parasites. (E) classical multidimensional scaling (CMDS) plot based on between-species MHC classII*b* B exon 2 genetic distances. (F) CMDS plot based on between-species MHC classII*b* B amino acid distances. (G) Plot of stable isotope data (δ15N versus δ13C) for Lake Tanganyika cichlids. Filled triangles in (A, B, D) represent tribes for which only one species was analyzed; grey bars in (G) indicate t-based 95% confidence intervals. Plot A, B, C and G were reproduced from Muschik et al. (2012; Figure S1).

Table 1. Sequencing success of MHC: Tribe name, species name, reads total from a bigger experiment, where MHC-experiment was part of, number of reads from MHC- or MHC- like genes, and number of defined haplotypes based on intron-exon assignment.

tribe	full name	reads total from big experiment	reads MHC (like)	number of haplotypes based on intron-exon
Bathybatini	*Bathybates graueri*	16602	26	3
Boulengerochromini	*Boulengerochromis microlepis*	8086	68	6
Cyphotilapiini	*Cyphotilapia frontosa*	11540	134	18
Cyphotilapiini	*Ctenochromis benthicola*	7751	113	8
Cyprichromini	*Cyprichromis leptosoma*	20194	740	31
Ectodini	*Aulonocranus dewindtii*	9241	244	4
Ectodini	*Callochromis macrops*	20465	701	22
Ectodini	*Ophthalmotilapia ventralis*	19782	574	24
Ectodini	*Xenotilapia spiloptera*	22662	682	18
Ectodini	*Grammatotria lemairii*	18482	450	23
Eretmodini	*Eretmodus cyanostictus*	14568	55	8
Haplochromini	*Ctenochromis horei*	17742	334	20
Haplochromini	*Gnathochromis pfefferi*	17533	47	1
Haplochromini	*Lobochilotes labiatus*	10587	253	24
Haplochromini	*Tropheus moorii*	11419	164	18
Haplochromini	*Haplochromis stappersii*	11093	106	8
Haplochromini	*Pseudocrenilabrus philander*	19873	428	41
Haplochromini	*Sargochromis macrocephalus*	15648	338	33
Haplochromini	*Lake Malawi, several*	11771	103	7
Haplochromini	*Astatotilapia burtoni*	15685	240	8
Haplochromini	*Lake Victoria, several*	24081	216	16
Lamprologini	*Altolamprologus compressiceps*	14396	47	7
Lamprologini	*Lamprologus callipterus*	14903	180	13
Lamprologini	*Lepidiolamprologus elongatus*	23818	1743	41
Lamprologini	*Neolamprologus modestus*	14564	749	11
Lamprologini	*Neolamprologus tetracanthus*	14387	856	20
Lamprologini	*Variabilichromis moorii*	27230	124	12
Lamprologini	*Julidochromis ornatus*	19643	74	8
Lamprologini	*Neolamprologus pulcher*	10933	255	18
Lamprologini	*Telmatochromis dhonti*	15823	331	29
Lamprologini	*Neolamprologus prochilus*	7760	94	6
Lamprologini	*Neolamprologus caudopunctatus*	3547	138	5
Limnochromini	*Gnathochromis permaxillaris*	12747	175	13
Limnochromini	*Limnochromis abeelei*	9760	111	7
Oreochromini	*Oreochromis tanganicae*	6876	108	9
Perissodini	*Perissodus microlepis*	17061	224	12
Perissodini	*Haplotaxodon microlepis*	13710	96	9
Trematocarini	*Trematocara nigrifrons*	17575	243	26
Tylochromini	*Tylochromis polylepis*	8112	5	0

group1	Group2a	Group2b	Group3	Group4	Group5a	Group5b	Group6a	Group6b	Group6c	Group6d	Group6e
reads for group	97	33	23	43	54	88	57	106	17	14	0
Bathybates graueri	0	0	0	0	0	0	0	0	0	0	0
Boulengerochromis microlepis	0	0	0	0	0	0	3	6	0	0	0
Cyphotilapia frontosa	0	0	1	9	1	1	0	2	6	0	0
Ctenochromis benthicola	0	1	1	2	1	1	1	7	0	0	0
Cyprichromis leptosoma	2	0	0	5	4	1	6	0	7	0	4	0
Aulonocranus dewindtii	2	0	1	0	0	0	1	0	0	0	0	0
Callochromis macrops	0	0	1	0	4	4	5	0	2	3	2	0
Ophthalmotilapia ventralis	5	5	0	0	2	4	2	0	6	0	1	0
Xenotilapia spiloptera	4	0	0	0	0	5	3	0	1	1	2	0
Grammatotria lemairii	3	1	1	0	5	4	3	4	6	1	0	0
Eretmodus cyanostictus	3	0	0	0	1	0	1	0	0	1	0	0
Ctenochromis horei	7	1	2	0	3	0	3	1	2	0	0	0
Gnathochromis pfefferi												
Lobochilotes labiatus	7	0	0	2	3	5	2	1	5	0	0	0
Tropheus moorii	3	1	0	2	4	3	2	0	2	0	0	0
Haplochromis stappersii	1	0	1	2	0	1	0	3	3	0	0	0
Pseudocrenilabrus philander	7	0	1	3	5	7	5	8	12	0	1	0
Sargochromis macrocephalus	7	1	3	1	4	6	3	2	5	0	2	0
Lake Malawi (no Rhaeso)	0	0	0	0	1	2	1	0	1	0	0	0
Astatotilapia burtoni	2	0	1	1	1	1	0	2	2	0	0	0
Lake Victoria	4	1	1	0	3	0	0	0	1	0	0	0
Altolamprologus compressiceps	2	0	0	0	2	0	0	1	0	0	0	0
Lamprologus callipterus	2	5	1	0	0	0	4	0	0	0	0	0
Lepidiolamprologus elongatus	2	3	3	2	4	5	3	8	11	2	0	4
Neolamprologus modestus	2	1	0	1	0	2	0	1	2	0	0	0
Neolamprologus tetracanthus	5	2	0	1	0	1	0	4	5	4	0	0
Variabilichromis moorii	4	1	0	0	0	2	0	2	2	0	0	0
Julidochromis ornatus	3	0	2	1	1	0	0	0	0	0	0	0
Neolamprologus pulcher	1	5	0	2	3	1	0	3	3	0	0	1
Telmatochromis dhonti	10	2	1	1	1	3	4	2	2	1	0	1
Neolamprologus prochilus	2	1	0	0	0	1	1	0	0	0	0	0
Neolamprologus caudopunctatus	0	1	2	0	1	0	0	0	0	0	0	0
Gnathochromis permaxillaris	0	0	0	3	0	3	6	0	0	0	1	0
Limnochromis abeelei	0	0	0	0	0	2	1	0	0	0	1	0
Oreochromis tanganicae	2	1	0	0	0	1	0	2	4	0	0	0
Perissodus microlepis	2	0	0	0	0	0	0	0	2	0	0	0
Haplotaxodon microlepis	3	0	0	2	0	0	2	0	1	0	0	0
Trematocara nigrifrons	1	2	0	3	0	13	0	1	0	0	0	0
Tylochromis polylepis	0	0	0	0	0	0	0	0	5	0	0	0

Table 2. Species and number of haplotypes in different homology groups (1-6), assignment by Hablützel et al. 2012 and Klein et al 1993, mean dn/ds ratio, list of position of positively and negatively selected sites (treshold of p-value 0.05).

Continued next page.

	group1	Group2a	Group2b	Group3	Group4	Group5a	Group5b	Group6a	Group6b	Group6c	Group6d	Group6e
similiar to Halblützel et al. 2012	undefined	DFB	DFB	undefined	undefined	DEB-like	DEB-like	DFB-like	DFB-like	DFB-like	DFB-like	DFB-like
homology group Klein et al. 1993	close to DFB similiar to unknown region (hg 3)	9	9	similiar to unknown region (hg 18)	Orni 459251/ 9	distantly similar to 2	?	1	1	1	1	1
uncorrected p-value (distance)	0.166	0.098	0.077	0.095	0.120	0.102	0.094	0.099	0.129	not analyzed	not analyzed	not analyzed
mean number of differences	38.0	23.0	17.6	22.1	27.9	23.1	22.8	17.2	22.7	not analyzed	not analyzed	not analyzed
mean dN/dS	1.10	1.08	1.61	1.46	2.04	1.34	2.15	1.90	2.09	not analyzed	not analyzed	not analyzed
positivley selected sites (<0.05)	8	64	66	34	1	6	15	8	8			
	62	66		63	26	7	62	21	9			
	73			76	40	8	76	24	17			
	76				63	17		33	21			
	77				66	33		34	24			
					68	34			33			
					70	61			34			
					76	76			43			
									63			
									64			
									65			
negativly selected sites (<0.05)	9	8	39	20	2	1		19	5			
	16	30	41		31	2		41	41			
	20	40	49		41	3		54	49			
	25	41			49	18		59	54			
	39	49			57	28						
	48				58	38						
	50					60						
	53											
	55											

Table continued next page

Table 3. Overview of number of haploytpes within different tribes, number of shared haplotypes within tribes, and the genetic distance both based on intron-exon and exon-only sequences (absolute difference and uncorrected p-distance).

	Number of haplotypes	shared haplotypes within tribe (intron/exon)	p-distance intron-exon	differences intron-exon	p-distance exon	differences exon
Ectodini	91	0	0.208	67.5	0.206	42.3
Lamprologini	159	10	0.215	67.2	0.208	42.2
Perissodini	21	0	0.182	59.9	0.196	46.1
Haplochromini	173	7	0.224	73.0	0.214	46.0
Cyphotilapiini	26	0	0.197	64.2	0.192	40.4
Limnochromini	20	0	0.177	58.8	0.186	40.9

Table 4. Non-parametric (Kruskal-Wallis) Anova on prevalence of metazoan ecto- and endoparasites between Lake Tanganyika cichlid tribes.

		Prevalence			Mean abundance		
		chi-squared	df	P	chi-squared	df	P
Endoparasites	Acanthocephala	12.7036	6	0.04799	15.0597	6	0.0198
	Nematoda	8.5214	6	0.2023	7.8547	6	0.2489
	Urogyrus	5.2505	6	0.5121	5.168	6	0.5225
	Enterogyrus	8.0579	6	0.2339	8.0503	6	0.2344
	Digenea	11.571	6	0.07225	11.6918	6	0.06921
	Gill cysts	6.6476	4	0.1557	5.1111	4	0.2761
	Fin cysts	5.957	6	0.428	5.1104	6	0.5297
Ectoparasites	*Gyrodactylus*	3.9404	6	0.6847	3.9079	6	0.6891
	Cichlidogyrus	8.6023	4	0.07185	12.0591	4	0.01692
	Ergasilus	3.9548	4	0.4122	2.1789	4	0.7029

Table 5. Spearman rank correlations across cichlid species between infection levels of various parasite groups, morphology, trophic ecology and isotope signatures. A) Prevalence vs. proportional stomach and gut content; B) Prevalence vs. principal components for body shape (body1 and body2), lower pharyngeal jaw shape (lpj1 and lpj2), proportional stomach and gut content (diet1 and diet2), and stable isotope signatures (δ13C and δ15N).

```
> mplete.obs",method="spearman");
```

	X1.x	X2.x	X1.y	X2.y	d13c	d15n	Comp.1	Comp.2
acanthocephalans	-0.37819	-0.01673	0.401969	0.004843	**0.443354**	**-0.43587**	-0.2646	0.235546
nematodes	0.120023	0.373561	-0.25633	0.301454	-0.30797	**0.413569**	0.080946	-0.15073
Urogyrus	-0.31175	0.074111	0.063698	0.128009	0.070436	0.014087	0.136584	-0.16966
Enterogyrus	0.160093	-0.30277	-0.25789	-0.15607	-0.06431	0.101147	0.296741	-0.10048
Digenea	0.376343	0.339019	0.18869	-0.03162	-0.01866	-0.08242	-0.22601	-0.07205
cysts_gills	0.053006	0.098217	0.032219	0.016629	0.190197	-0.0899	-0.16629	0.110689
cysts_fins	0.108219	**-0.51172**	0.183199	-0.27209	0.163101	-0.34321	-0.143	0.19634
Gyrodactylus	-0.18859	-0.07159	-0.06158	0.014625	-0.27403	0.026171	0.277876	0.231692
Cichlidogyrus	-0.03578	0.346861	-0.08202	**0.454191**	-0.17147	0.191537	0.145289	-0.3473
Ergasilus	-0.1546	-0.24745	0.190584	0.079965	0.12439	-0.03732	0.255	0.042204

```
> mplete.obs",method="spearman");
```

	sand	aufwuchs	plants	mollusks	arthropods	crustaceans	fish	fisheggs	fishscales
acanthocephalans	**0.452105**	**0.437102**	0.323319	0.310026	-0.04271	-0.36095	-0.12066	-0.033	-0.18336
nematodes	-0.36133	0.039811	-0.14231	**-0.44265**	0.006513	0.134536	-0.01413	-0.19056	-0.07755
Urogyrus	0.054116	**0.528514**	-0.15567	0.2121	0.044099	0.290338	0.044517	-0.00725	0.041723
Enterogyrus	-0.26427	-0.31221	-0.23754	0.111399	0.189566	0.213193	0.387099	0.243561	0.045285
Digenea	0.05379	-0.20913	0.09556	0.257537	-0.31466	-0.28745	-0.27332	-0.05351	0.163454
cysts_gills	-0.12047	-0.36457	0.160441	-0.08638	0.043132	0.13422	**-0.47554**	-0.18228	-0.06517
cysts_fins	-0.02121	-0.00786	0.184158	0.191573	0.088121	-0.04359	-0.35989	-0.13893	0.404298
Gyrodactylus	0.031689	0.179125	-0.06602	0.116536	0.299429	-0.00039	0.28391	-0.01717	0.035476
Cichlidogyrus	0.404137	0.059604	-0.20643	0.402507	-0.07766	0.220099	0.096436	0.306391	-0.27492
Ergasilus	0.036895	0.054758	-0.11645	0.329533	0.352291	0.099085	0.163566	0.003129	-0.14352

Table 6. Spearman rank correlations across cichlid species between infection levels of various parasite groups, morphology, trophic ecology and isotope signatures. A) Mean abundance vs. proportional stomach and gut content; B) Mean abundance vs. principal components for bodyshape (body1 and body2), lower pharyngeal jaw shape (lpj1 and lpj2), proportional stomach and gut content (diet1 and diet2), and stable isotope signatures (δ13C and δ15N).

```
> mplete.obs",method="spearman");
```

	X1.x	X2.x	X1.y	X2.y	d13c	d15n	Comp.1	Comp.2
acanthocephalans	**-0.45474**	0.149534	0.360899	0.129801	0.377563	-0.26749	-0.25259	0.263987
nematodes	0.093948	0.370678	-0.24603	0.310216	-0.29394	**0.426023**	0.064648	-0.1386
Urogyrus	-0.29142	0.084486	0.04408	0.143259	0.0551	0.008571	0.14081	-0.17877
Enterogyrus	0.135239	-0.31801	-0.28655	-0.10779	-0.07766	0.122518	0.316673	-0.06829
Digenea	0.372081	0.311449	0.177749	-0.04819	0.003628	-0.08032	-0.22128	-0.06167
cysts_gills	0.124411	0.030584	-0.05287	0.032139	0.109896	-0.12804	-0.11093	0.076202
cysts_fins	0.14779	**-0.47878**	0.156257	-0.25247	0.139323	-0.33407	-0.12701	0.155488
Gyrodactylus	-0.1354	-0.04231	-0.0277	-0.04001	-0.27311	0.006924	0.240033	0.223877
Cichlidogyrus	-0.22962	0.378343	-0.05479	**0.459665**	-0.12698	0.010872	0.148293	-0.25223
Ergasilus	-0.17367	-0.19804	0.031774	0.183678	0.061371	0.030468	**0.422634**	-0.01872

```
> mplete.obs",method="spearman");
```

	sand	aufwuchs	plants	mollusks	arthropods	crustaceans	fish	fisheggs	fishscales
acanthocephalans	0.287662	**0.48316**	0.339871	0.189243	-0.03377	-0.29621	-0.12828	-0.0741	-0.2598
nematodes	-0.34882	0.060176	-0.11517	**-0.4308**	0.011627	0.155029	-0.04184	-0.20458	-0.10345
Urogyrus	0.08051	**0.520178**	-0.16483	0.233274	0.040406	0.270823	0.044498	-0.00725	0.048024
Enterogyrus	-0.26413	-0.32739	-0.24011	0.11557	0.208884	0.176584	**0.414421**	0.234631	0.001728
Digenea	0.033542	-0.19692	0.097613	0.273821	-0.29694	-0.28305	-0.2828	-0.06031	0.183461
cysts_gills	-0.02879	-0.33065	0.103405	0.030697	0.05754	0.125161	**-0.47437**	-0.12049	0.044676
cysts_fins	-0.00548	0.001564	0.16408	0.235466	0.084671	-0.04104	-0.35838	-0.156	**0.425717**
Gyrodactylus	0.061389	0.201704	-0.04437	0.126846	0.276192	-0.00789	0.244686	-0.02942	0.063346
Cichlidogyrus	0.335459	0.247464	-0.16216	0.270849	-0.04218	0.396668	0.017359	0.328975	-0.2729
Ergasilus	0.049453	0.056331	-0.28163	0.319695	**0.437432**	0.226668	0.2033	0.095038	-0.13208

Discussion

The main aim of this doctoral thesis was to deepen our knowledge of the phylogenetic relationships within the species flock of cichlid fishes in East African Lake Tanganyika, a role model for adaptive radiation[1, 2]. This goal was successfully reached, as presented in the five chapters of my thesis. In the following I would like to briefly summarize and review my main results.

In the first two chapters of my thesis, I present a novel approach of examining the phylogenetic relationships of the main linages of cichlids in Lake Tanganyika, based on a large set of genomic markers. So far, mainly mitochondrial markers have been used to study the phylogenetic relationships in Tanganyikan cichlids. The resulting tree topologies were in part contradicting and generally neither well resolved nor supported.

In my thesis, I thus investigated the phylogenetic signal provided by a large sequence-based nuclear multi-marker dataset obtained my next-generation DNA sequencing. In the first chapter[3] I describe the starting point of this project, in which I used mainly expressed sequence tags (ESTs) from several cichlid species[4-9], the unassembled whole genome shotgun (wgs) database of Oreochromis niloticus and the Medaka fish genome to identify genes that could be suitable as phylogenetic markers.

Due to the chosen marker length between 400 and 500 bp, I could successfully sequence these loci with the traditional Sanger method and later with the 454 next-generation sequencing platform. One advantage of our primer design strategy is that the rough location of the markers within the genome was known (in the comparison to other fish genomes), which enabled us to sample many independent gene histories to infer the species tree. This is in clear contrast to previous studies based on a single marker (or very few markers), where the history of one or few genes has been used to represent species history. We argue that genome-wide multi-marker datasets should better be suited to resolve the complex phylogenetic history within an adaptive radiation, where lineage formation is generally rapid and where more than two species might emerge from a common ancestor (i.e. speciation is not necessarily bifurcating). I would like to note here that the markers were chosen before the five sequenced and assembled cichlid genomes became available by the Broad Institute, which would, of course, have facilitated primer design (e.g. using bioinformatic pipelines searching for single copy genes)[10] nowadays.

In the second chapter I applied an extended version of my newly developed multi-marker primer set to a representative set of East African cichlids. In total, I compiled a dataset based on 42 markers (17,545 bp in total) and 45 species. With this dataset and up-to-date phylogenetic analyses, we provide a new phylogenetic hypothesis for the relationships between the main cichlid lineages, so-called 'tribes', in Lake Tanganyika. In particular, our analyses revealed new insights into the phylogenetic relationships within the mouth brooding lineages. For example, the nuclear markers uncovered a close affiliation of the Eretmodini to the more derived Haplochromini. This finding is concordant to another recently published study[11] and to an older allozyme study[12], however quite distinct to mitochondrial DNA based studies, which placed the Eretmodini closer to the substrate spawning Lamprologini[13-15]. While most parts of our new phylogeny were well resolved, a few areas of uncertainty remained. A novel approach to evaluate the phylogenetic signal in our dataset, in which we combined simulations and resampling techniques, suggested that our dataset contained enough phylogenetic signal. It thus seems that the remaining polytomies at the basis of the derived mouthbrooders (the 'H-lineage') represent biological reality. In particular, there are still uncertainties with respect to the positioning of the two tribes Cyphotilapiini and Limnochromini, as well as the Perissodini and Cyprichromini. It remains to be tested if the low resolution is due to the stochastic sorting of ancestral polymorphisms, past hybridization, the short time frame, the lack of mutations and thus the lack of phylogenetic signal or if we only have chosen inconclusive and/or too few markers.

Overall, however, the amount of markers was suitable to provide new insights to the cichlid phylogeny of Lake Tanganyika, as shown in the resampling tests, where we combined and compared different sets of randomly chosen markers. Further, the high concordance of the concatenated tree estimates and the species tree provided by the Bayesian concordance analysis with Bucky, show the reliability and power of our multi-marker dataset.

The usefulness of my markers has previously been shown in other studies, for example in the study of Muschick et al. (2012)[1], and, additionally, in the study by Emilia Santos presented in the fourth chapter of my thesis, where I contributed to the phylogenetic analyses. In this project, nine of my markers were used to generate a strong phylogenetic backbone for the Haplochromini, the most species-rich tribe of cichlids. This was essential as it strengthened the assumptions about the timing of the emergence of the studied novel color trait.

In the third chapter I applied phylogenetic analysis methods to investigate the placement of a newly discovered species from Lake Tanganyika, Haplochromis sp.

'Chipwa'. This resulted in the surprising finding of a close genetic affiliation of this species to the Lake Victoria cichlid superflock. Parallel to our study other researchers postulated the non-inclusive monophyly of the three Great Lakes, as they observed high degrees of shared genetic polymorphisms in a SNP study[16]. However, we are the first ones to potentially support their hypotheses with the existence of a species, which invaded Lake Tanganyika from another drainage system. If this newly detected species has ever hybridized with other Lake Tanganyika haplochromines remains to be tested.

In the last chapter we studied the potential contribution of parasites to the Lake Tanganyika cichlid radiation investigating several loci of MHC genes and infection levels of metazoan macroparasites in combination with a trophic and morphological dataset. This analysis gave hints about the enormous diversity at the MHC class II loci for East African cichlids, as this was the first study including Lake Tanganyikan tribes (other than Oreochromini and Haplochromini). Besides resolving some influences of habitat, diet and morphological traits to the level of parasite infection, we could finally detect a correlation between infections and the MHC genes. Nevertheless, we would like to increase the intensity of sequencing to gain a better coverage for our data matrix, as it was shown that not all known loci could be amplified in all species and tribes.

Since it is known that concatenation of many markers can lead to wrong signals in phylogenetic inferences[17-19], we plan to extend our analyses to coalescent-theory based species tree estimates[20-23] in the future. We would particularly like to cross-validate our phylogenetic results presented in the second chapter. Coalescence methods account for gene tree discordance originating from stochastic processes such as incomplete lineage sorting. For this purpose we still have phased data of 42 markers for around 400 more cichlid individuals. This dataset would further allow studying the genetic diversity on an individual level within species for single genes. Thus, different selection regimes could be potentially detected, as we did, for example, for the MHC genes (fifth chapter).

Multi-marker studies are the link between single marker and whole genome and transciptome studies. Since the field of phylogenomics and its standards and "doctrines" are rapidly changing, a new range of possibilities is constantly established. On the one hand, this is due to the developments in the field of next-generation sequencing methods[24, 25]. On the other hand, also theory and analytical tools are constantly being improved [25-27]. The challenging and time-consuming steps, in the analyses of the data obtained by whole-transcriptome and genome sequencing, are the choice of the appropriate phylogenetic markers out of a huge data cloud. Such

markers should reflect appropriate evolutionary rates, should be homologous, and be expressed across a range of species (in the case of transcriptomes). An additional challenge is the computationally time expensive analysis of millions of base pairs, as well as the computational expenses of data partitioning and model choice. Most of the available phylogenetic software packages and species tree estimation algorithms were not designed for datasets of this size. Thus our strategy of multi-marker sequencing seems still very appropriate to be applied to answer phylogenetic questions, specially if we take into account our planned extension to the coalescent theory based analyses.

Still, if possible, I would add more markers and more taxa per lineage in the future, to sample a wider range of possible gene histories and to better estimate the degree of incomplete lineage sorting within an adaptive radiation. I would further focus on species with 'mixed' origins (such as *Eretmodus cyanostictus*) to gain insights in the processes of hybrid speciation or the impact of hybridization to adaptive radiations[28]. Additionally to "zooming out" and adding all Lake Tanganyika tribes, "zooming in" (more species and even individuals within specific tribes) could potentially reveal more resolution both with coalescent-based species tree estimates as well as the concatenation approach (now that we know more about potential sister tribes). In addition to the transcriptomes and genomes I would further think about new methods, such as paired end RAD sequencing, which could potentially produce assemblies of up to 500 bp[24].

In this thesis I report the successful development and application of a newly developed marker set for East African cichlid fishes. It was used for the amplification of more than 40 loci in 45 species covering a phylogenetic range of over 10 million years, revealing a new perspective to the phylogenetic history of East African cichlids.

1. Muschick, M., Indermaur, A. & Salzburger, W. Convergent Evolution within an Adaptive Radiation of Cichlid Fishes. Current Biology 22, 2362-2368 (2012).

2. Santos, M.E. & Salzburger, W. Evolution. How cichlids diversify. Science 338, 619-21 (2012).

3. Meyer, B.S. & Salzburger, W. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. Molecular Ecology Resources (2012).

4. Baldo, L., Santos, M.E. & Salzburger, W. Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. Genome Biol Evol 3, 443-55 (2011).

5. Kobayashi, N., Watanabe, M., Horiike, T., Kohara, Y. & Okada, N. Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. Gene 441, 187-91 (2009).

6. Lee, B.-Y. et al. An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. BMC genomics 11, 278 (2010).

7. Salzburger..., W., Renn, S., Steinke, D. & Braasch, I. Annotation of expressed sequence tags for the East African cichlid fish *Astatotilapia burtoni* and .... BMC genomics (2008).

8. Tine, M. et al. Transcriptional responses of the black-chinned tilapia *Sarotherodon melanotheron* to salinity extremes. Marine Genomics 1, 37-46 (2008).

9. Tsai, C.L., Wang, L.H., Shiue, Y.L. & Chao, T.Y. Influence of temperature on the ontogenetic expression of neural development-related genes from developing tilapia brain expressed sequence tags. Mar Biotechnol 9, 243-61 (2007).

10. Li, C., Riethoven, J.-J.M. & Ma, L. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. BMC Evol Biol 10, 90 (2010).

11. Friedman, M. et al. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proceedings of the Royal Society B 280, 20131733 (2013).

12. Nishida, M. Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes : inferences from allozyme data Experientia 47, 974-979 (1991).

13. Koblmüller, S., Sefc, K.M. & Sturmbauer, C. The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. Hydrobiologia 615, 5-20 (2008).

14. Clabaut, C., Salzburger, W. & Meyer, A. Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. Journal of Molecular Evolution 61, 666-81 (2005).

15. Day, J.J., Cotton, J.A. & Barraclough, T.G. Tempo and mode of diversification of lake Tanganyika cichlid fishes. PLoS ONE 3, e1730 (2008).

16. Loh, Y.-H.E. et al. Origins of Shared Genetic Variation in African Cichlids. Molecular Biology and Evolution (2012).

17. Kubatko, L. & Degnan, J. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology (2007).

18. Leaché, A.D. & Rannala, B. The Accuracy of Species Tree Estimation under Simulation: A Comparison of Methods. Syst Biol (2010).

19. Maddison, W.P. Gene Trees in Species Trees. Systematic Biology 46, 523-536 (1997).

20. Brito, P.H. & Edwards, S.V. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135, 439-55 (2009).

21. Liu, L. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542-3 (2008).

22. Heled, J. & Drummond, A.J. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27, 570-80 (2010).

23. Fan, H. & Kubatko, L.S. Estimating species trees using approximate bayesian computation. Mol Phylogenet Evol (2011).

24. McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66, 526-38 (2013).

25. Lemmon, E.M. & Lemmon, A.R. High-Throughput Genomic Data in Systematics and Phylogenetics. Annual Review of Ecology, Evolution, and Systematics (2012).

26. Bayzid, M.S. & Warnow, T. Naive binning improves phylogenomic analyses. Bioinformatics 29, 2277-84 (2013).

27. Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol 29, 1917-32 (2012).

28. Seehausen, O. Hybridization and adaptive radiation. Trends in Ecology & Evolution 19, 198-207 (2004).

Acknowledgement

First of all, I would like to express my deep gratitude to Prof. Dr. *Walter Salzburger,* because he gave me the opportunity to conduct my PhD studies in his research team. Besides a familiar atmosphere, I enjoyed the diverse journeys to conferences, field trips and our lab-retreats. It was great to have the chance to visit many workshops and even to organize an own conference. Thank you, dear Walter, for the big support, the never ending optimism and the open door.

I am thankful to Prof. Dr. *Markus Pfenninger* for his willingness to read and to supervise this PhD thesis as well as to join my defence.

I acknowledge the generous financial support by the University of Basel, the Freiwillige Akademische Gesellschaft Basel and the Burckhardt-Bürgin-Stiftung.

I want to thank all members from the *SalzburgerLab* for being an open, enthusiastic and interesting team. More particular, I want to thank *Brigitte Aeschbach* for the various chats and cheering me up with good (lab-)advices and pieces of chocolate. Thanks to *Nicolas Boileau* for the help and time together in the lab. Especially, I am extremely grateful to my office mates *-Anya Theis, Adrian Indermaur, Corina Heule and Emília Santos-* for the last years together, not only for your contributions to my work, but also for being parts of your lives. Thank you for your love, magic and sparkles, the adventures, the coffee coins, the cooking nights and just being my best friends. And of course thanks to my other friends: Thank you *Astrid Böhne, Eveline Diepeveen, Moritz Muschick* and *Yuri Kläfiger* for the great time and helping out when support was necessary! I want to thank my co-authors *Michael Matschiner, (Super-)Joost Raeymaekers* and *Bernd Egger* for helpful comments and discussions in times where I needed it. I hope that our (scientifique) lives will be further connected. I am grateful for *Zuzana Musilová, Hugo Gante* and *Sara Stieb* for fruitful ideas about my research and for their smiles when entering our office. I will miss you! I wanted to thank *Lukas Zimmermann* and *Jean-Claude Walser* for providing help with software issues. Thanks to the whole Zoological Institute, I enjoyed the great opportunities and facilities provided (e.g. all the seminars, the Guarda workshop, the Cooking Club). You all made me feel welcome!

I am happy and thankful that I had the opportunity to be part of the great team from Cesky Krumlov and to learn so much. I will never forget this intense time with *Scott Handley, Naiara Rodríguez-Ezpeleta, Michael Cummings* and of course *Ruxandra Molnar and Linn Groeneveld*.

My biggest thank I owe to *Dennis Münd*. Not only being the most important person, but also for regular helping with Perl scripts. And I want to thank *my family* for the love and interest in everything I do.

So long, and thanks for all the fish.

Appendix

Appendix 1a: Used barcodes for the fusion primers of 454 pyrosequencing. Fusionsprimers were synthesized at Microsynth (Balgach, Switzerland). The forward fusionprimers composed of the template specific forward primer, the B-Adaptor and the respective TCMID1-10 barcodes. Reverse primer are composed of the template specific reverse primer and the A-Adaptor.

MID Name	MID Sequence
TCMID1	ACGAGTGCGT
TCMID2	ACGCTCGACA
TCMID3	AGACGCACTC
TCMID4	AGCACTGTAG
TCMID5	ATCAGACACG
TCMID6	ATATCGCGAG
TCMID7	CGTGTCTCTA
TCMID8	CTCGCGTGTC
TCMID9	TAGTATCAGC
TCMID10	TCTCTATGCG
TCMID11	TGATACGTCT
TCMID12	TACTGAGCTA
TCMID13	CATAGTAGTG
TCMID14	CGAGAGATAC
TCMID15	ATACGACGTA
TCMID16	TCACGTACTA

Appendix 1b: The two used Adaptor sequences.

A Adaptor

CGTATCGCCTCCCTCGCGCCATCAG

B Adaptor

CTATGCGCCTTGCCAGCCCGCTCAG

Appendix 2 a: 5 Primermixes (A-E) per individual, primer sequences, lengths and estimated temperatures

name	sequence 5'-3'	length of primers	product length with primers	Tm (by Primer3)	Tm (by Microsynth)	insilico with 3 µM Magnesium	primer mix
rag_exon3_bm1f	TCGGCGCTTTCGGTACGATGTG	22	462	59.87	61.4	72.9 °C	A
rag_exon3_bm2r	TGCCCCTGAAGTGGAASSGA	20	?	?		71.9 °C	
b2m_bm3f	GCCACGTGAGTRATTTCCACCCC	23	509	?	?	73.7 °C	A
b2m_bm4r	ACGCTAYACRGYGGACYCTGA	21	?	?		75.8 °C	
gapdhs_bm_3f	CCCTGGCCAAAGTCATCCACGATA	24	500	58.51	61.2	72.4 °C	A
gapdhs_BM_3r	CACCACTGACACATCGGCCACT	22		59.1	61.4	72.9 °C	
ptr_bm1f	GCGGGTAGTGAATGTGAGTGCG	22	437	58.26	61.4	71.4 °C	A
ptr_bm1r	ACCCAAGACACCCAGCTCCA	20		57.21	59.5	71.7 °C	
enc1_bm_2f	CRGTTCGCCTTGCGCTRTTGC	21	417	?		74.6 °C	A
enc1_bm1_r	TGGGTGCCGCCTTTGACCAT	20		58.9	59.4	72.8 °C	
38a_F	AGCAGGGTTGACCTTCTCAA	20	487	52.91	?	67.1 °C	A
38a_R	TGGCTAAAATCCCCGATGTA	20		50.4	?	64.5 °C	
s7_bm_1f	CGTGCCATTTTACTCTGGACTKGC	24	570	?		60.4 71.6 °C	A
s7_bm_1r	AACTCGTCYGGCTTCTCGCC	20		?		60.7 72.5 °C	
MHC_TU383	CTCTTCATCAGCCTCAGCACA	21		?	?	67.7 °C	A
MHC_TU377	TGATTTAGACAGARKGKYGCTGTA	24		?	?	71.1 °C	
tbr1_bm_1f	ATCGTGCCGGGTGCGAGATA	20	592	58.02	59.5	71.6 °C	A
tbr1_bm_1r	AGGACGGCGTCTCAATCCAGCT	22		59.98	61.4	73.6 °C	
aqua_bm_1f	ATCAACCCTGCTCGCTCCTTCG	22	478	58.97	61.4	72.4 °C	A
aqua_bm_1r	TGCATCGTTGCCTCCGTTGACG	22		60.3	61.4	73.5 °C	
hprt1_bm_1f	TCAGYGATGAGGAGCAGGGTTATG	24	505	58.35	60.4	71.8 °C	B
hprt1_bm_1r	CGACCGTCATTGGGATGGAGC	21		57.62	61.5	71.3 °C	
anxa4_bm_1f	TGGACGAGGCCCAGGCTATTCAAG	24	505	60.29	62.9	73.9 °C	B
anxa4_bm_1r	ACGTCTTCCAGGCAGCCAGACA	22		60.11	61.4	73.8 °C	
pgk_bm_3f	CGGTACCTCCCTGTATGACGAGGA	24	437	58.82	?	73 °C	B
pgk_bm_3r	GCAGCCAGATTTGGTCACCTCGA	23		59.2	?	72.7 °C	
bmp4_bm_1f	GAGGACCCATGCCCATTCGTTT	22	578	57.21	59.7	71.2 °C	B
bmp4_bm_2r	GCCACTATCCAGTCATTCCAGCC	23		57.23	61.5	70.7 °C	
bmp2_bm_1f	AGGCCCTGGCCAGCCTAAAA	20	415	57.93	59.5	71.7 °C	B
bmp2_bm_1r	TCCTGCGTCTGTGGGCATCCTT	22		60.18	61.4	74 °C	
Cich2_F tmo4c4	TTATGCTGAGGTGTTTGGCCTAC	23	473	62.53	?	68.4 °C	B
Cich2_R	CCACAGCACCCTCCTCATAAAT	21		61.27	?	68.1 °C	
fgf6a_bm1_f	CGCAAAGGTGCCACTACAG	19	512	60.46	57.4	66.7 °C	B
fgf6a_bm3_r	TCGCACTGCACGGATGCAAA	20		57.84	57.5	70.9 °C	
runx_bm_1f	CGGGGTTGGTGTTTGAGGGCAA	22	411	59.79	61.4	73.8 °C	B
runx_bm_1r	GCTGACATGGTGTCACTGTGCTGA	24		59.42	61.2	72.8 °C	
furin_bm_1f	GCTGCATGGGGACAGACAGTCA	22	357	58.83	61.4	72.4 °C	B
furin_bm_1r	ATAGTCACTGGCACCCGCCACA	22		60.18	61.4	73.9 °C	
wnt7bb_bm_1f	GCGTCTCGGGATCCTGTACCACTA	24	435	59.35	62.9	73.1 °C	B
wnt7bb_bm_1r	TGCAGGTAAACACCTCCGTCCT	22		57.58	59.7	71.7 °C	
pax9_bm_2bf	TCCCACGGCTGTGTCAGYAA	20	434	?	?	72.3 °C	C
pax9_bm_1r	ACAGAGTGCGAGGAAGGCCA	20		57.7	59.5	71.5 °C	
sox10b_bm_1bf	TSCRGGGTCTGGGAAACCTCAT	22	486	?	?	73.8 °C	C
sox10b_bm_1r	TGGTGGTCGGCGTATTCTGCAA	22		58.84	59.7	72.4 °C	
otx2_bm_3f	GCAGAACAAAGTGCGACCTGCC	22	457	59.18	61.4	72 °C	C
otx2_bm_3r	GTCTGCTGTGGAGTTGAAGCCCA	23		59.07	61.5	72.6 °C	
otx1b_bm_1f	TACACCTCCTGCTGTCTCCAGCAC	24	402	59.94	62.9	73.5 °C	C
otx1b_bm_1r	ATAGATGAGGCCGTCATGGGGC	22		58.24	61.4	72 °C	
dlx2_bm_1f	ATCGCCAACTCCCGCAGACA	20	563	58.43	59.5	72.1 °C	C
dlx2_bm_1r	TCCGTTGAAGYGCAGCCAGT	20	?		58.7	72.5 °C	
dlx4b_bm_1f	GCGTGGATTTCTTCCAGGCTGTC	23	403	58.25	61.5	71.5 °C	C
dlx4b_bm_1r	CTGTGTGCTCTAATCTGCTGTGGG	24		57.55	61.2	70.9 °C	
barx1_bm_1f	TCTCGCAGAGTCTCTCGGTCTG	22	389	57.34	61.4	71 °C	C
barx1_bm_2r	TCGCTGCTGGGGATGGAGTT	20		57.75	59.5	71.7 °C	

to be continued next page...

name	sequence 5‘-3‘	length of primers	product length with primers	Tm (by Primer3)	Tm (by Microsynth)	insilico with 3 µM Magnesium	primer mix
ednrb1_bm_1f	CGTTGGCCTGCACTGCCATT	20	481	58.45	59.5	71.6 °C	C
ednrb1_bm_1r	AGGCAGCCAGCACAGAGCAAA	21		59.24	59.5	72.3 °C	
mc1r_bm1f	GACCACGGCCTCCTGGATGT	20	510	58.13	61.6	72.3 °C	C
mc1r_bm1r	GTTGCAGAAGGGGCTGGTGG	20		57.8	61.6	71.4 °C	
C-Ski_BM_4F	CGACCAGCTGGAGATCCT	18	492	58.86	57.1	66.4 °C	C
C-Ski_BM3_r	TCCTCTTGTACTTGTTGGCG	20		58.92	55.4	65.8 °C	
kita_BM1_f	CAGAGTACTGCTGTTTCGGMGAT	23	611 ?		58.7	70.5 °C	D
kita_BM1_r	GGCTAAGAACTCCATGCCTTTGGC	24		58.3	61.2	71.4 °C	
Mitfa_bm4f	CCTGGCATGAAGCARGTACTGGAC	24	456 ?		62	73.1 °C	D
Mitfa_bm4r	TTGCYAGAGCACGAACTTCRGC	22	?		59.7	73.2 °C	
Tyr_bm1f	TGGGTGGACGCAACTCCCTT	20	659	57.89	59.5	72.2 °C	D
Tyr_bm1r	TGGCAAATCGGTCCATGGGT	20		56.17	57.5	70.4 °C	
hag_bm_7f	AAACTGGTACARYGGGVTCTGC	22	470 ?	?		72.7 °C	D
hag_bm_4r	AGCGRCAGACGTCACCCTTGT	21	?	?		74.3 °C	
AIM1_AU_F1	GAGCTATGGACTGGGGTCAC	20	323	53.7	59.5	67.6 °C	D
AIM1_AU_R1	TGGCTGTTTGACACTTGAGG	20		52.53	55.4	66.4 °C	
rh1_bm_2f	TCGCCTTGGCTGCAATCTGG	20	444	57.6	59.5	70.8 °C	D
rh1_bm_2r	ACCATGCGGGTGACTTCCCT	20		57.67	59.5	72 °C	
LWS_bm_1f	ATTGCTGCTCTTTGGTCCCTGACA	24	685	58.56	59.6	72.3 °C	D
LWS_bm_1r	AGCCAGAGGGTGGAAGGCAT	20		57.35	59.5	71.5 °C	
sws1_bm_2f	TGGGTCACACGCTGTGTGCT	20	514	58.85	59.5	72.5 °C	D
sws1_bm_1r	CAGCAGCTGGGAGTAGCAGAARA	23	?		60.3	71.4 °C	
Ccng1_bm_2f	CTGCTTGCCCTGGCTCTCCT	20	707	58.4	61.6	71.9 °C	E
Ccng1_bm_2r	AGCTGACTCAGGTATGGTCGGA	22		56.6	59.7	70.7 °C	
Snx33_bm1f	TGGCTGTACAACCGCCTGCT	20	478	58.63	59.5	72 °C	E
snx33_bm2r	CCAAYRTGAATGCSTGGCTGA	21	?	?		71.1 °C	
rpl13a_bm_1f	ACCTGGCTTTCCTGCGCAAGA	21	645	58.97	59.5	72.3 °C	E
rpl13a_bm_1r	TTGCGAGAGGGCTTCAGACGCA	22		60.75	61.4	73.9 °C	
edar_bm_1f	TGAGCAGCTGTTGAGCCGCA	20	477	59.28	59.5	72.1 °C	E
edar_bm_8r	CRCATKGCARGYYCTGGCATACA	23	?		60.3	76.7 °C	
shh_bm_1f	TGGCACCAAGGAAGCCGTCA	20	512	58.55	59.5	72.4 °C	E
shh_bm_2r	CACTGCTTGGAGGCTGGGA	19	?	?		69.8 °C	
msx1_bm_1f	AAAGGGAGCGCGGATGGGTT	20	531	58.69	59.5	72.5 °C	E
msx1_bm_1r	TGCGCGCTCGGCGATAGAAA	20		59.57	59.5	72.1 °C	
csfr1_bm1f	AAGCACAGATGGGACACGCC	20	459	57.25	59.5	70.8 °C	E
csfr1_bm1r	TGTACTGGCCCTGCTCCTGT	20		57.01	59.5	71 °C	
pax3a_bm_1f	AAGAGCCCGGTGGAGGAAGCAA	22	471	60.11	61.4	74 °C	E
pax3a_bm_2r	TGACGGCGTTGGTGTGTCCT	20		58.58	59.5	72.5 °C	

Appendix 2 b: Normalized primer stock for 10 or 8 primerpairs

	Qiagen 100 µM (100 pmol/µL)	for 10 (20)	for 8 (16)
Each primer	10 µL	5 µL	5 µL
TE-Buffer	variable	150 µL	170 µL
Total	500 µL	250 µL	250 µL

Appendix 3a: Standard multiplex PCR Mastermix using Multiplex PCR Kit (Qiagen, Hombrechtikon, Switzerland)

Reagent	Qiagen reaction	final conc	with less Primer	final conc
2x Qiagen Multiplex PCR Mastermix	25 µL	1x	12.5 µL	1x
10x primer Mix of 2µM each primer	5 µL	0.2 µM	5 µL (1:4 diluted -> 100 +300)	0.1 µM
RNase free water	variable	-	5.5 µL	-
template DNA	variable	<= 1µg/50 µL	2 µL	<= 1µg/50 µL
Total Volume	50 µL		25 µL	

Appendix 3b: Standard multiplex PCR conditions

1x	95° C (HotStarTaq Polymerase)		15 min
40x	94° C denaturation		30s
	60° C annealing	(A, B, D, E mix)	1:30 min
	[62° C annealing (C mix)]		
	72° C extension		90s
1x	72° C final extension		10 min