

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).

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

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

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 (T_m) 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 manufacturer's protocol (Qiagen, Hombrechtikon, Switzerland). PCRs were performed in a final volume of 12.5 μ L containing REDTaq[®] DNA Polymerase (0.04 units/ μ L), its PCR Buffer (1 \times) (Sigma-Aldrich, Buchs, Switzerland), 200 μ M of each dNTP (Promega, Dübendorf, Switzerland), 0.2 μ M of each sense and anti-sense 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 \times 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 μ L purified PCR products, the primers specified in Table 1 (0.5 μ L/10 μ M), and 1 μ L BigDye[®] Terminator v3.1 Ready Reaction Mix (Applied Biosystems, Rotkreuz, Switzerland) in a total volume of 8 μ L 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 3130xl 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	GAGGACCCATGCCATTCGTTT/GCCACTATCCAGTCATTCCAGCC	577	ENSORLGG00000013304	22	482	0
bmp2	AGGCCCTGGCCAGCCTAAAA/TCCTGCGTCTGTGGGCATCCTT	414	ENSORLGG00000009772	24	315	0
fgf6	CGCAAAGGTGCCACTACAG/TCGCACTGCACGGATGCAAA	512	ENSORLGG00000015820	23	286	158
furina	GCTGCATGGGGACAGACAGTCA/ATAGTCACTGGCACCCGCCACA	357	ENSORLGG00000009133	3	154	94
runx2	CGGGGTTGGTGTGTTGAGGGCAA/GCTGACATGGTGTCACTGTGCTGA	411	ENSORLGG00000010169	24	95	218
shh	TGGCACCAAGGAAGCCGTCA/CACTGCTTGGAGGCTGGGA	512	ENSORLGG00000010463	20	421	0
pax9	TCCCACGGCTGTGTCAGYAA/ACAGAGTGGCAGGAAGGCCA	434	AB187122.1	?	338	0
sox10b	TSCRGGTCTGGGAAACCTCAT/TGGTGGTGGCGTATTCTGCAA	486	ENSORLGG00000014587	8	310	0
ednrbl	CGTTGGCCTGCATGCCATT/AGGCAGCCAGCACAGAGCAAA	479	ENSORLGG00000011054	17	54	320
mc1r	GACCACGGCCTCCTGGATGT/GTTGCAGAAGGGGCTGGTGG	510	ENSORLGG00000009400	3	401	0
c-ski	CGACCAGCTGGAGATCCT/TCCTCTTGTACTTGTGGCG	491	ENSORLGG00000016855	7	408	0
kita	CAGAGTACTGCTGTTTCGGMGAT/GGCTAAGAAGTCCATGCCT TTGGC	611	ENSORLGG00000000569	4	237	270
mitfa	CCTGGCATGAAGCARGTACTGGAC/TTGCYAGAGCACGAA CTTCRGC	456	ENSORLGG00000003123	5	25	373
tyr	TGGGTGGACGCAACTCCCTT/TGGCAAATCGGTCCATGGGT	659	ENSORLGG00000010905	13	155	413
csfr1	AAGCACAGATGGGACACGCC/TGTAAGTGGCCCTGCTCCTGT	459	ENSORLGG00000004849	10	25	324
pax3	AAGAGCCCGGTGGAGGAAGCAA/TGACGGCGTTGGTGTGCTCT	471	ENSORLGG00000015932	17	254	130
hag	AAACTGGTACARYGGVVTCTGC/AGCGRACAGCTCACCCCTGT	470	ENSORLGG00000000906	15	115	309
rag	TCGGCGCTTTCGGTACGATGTG/TGCCCTGAAGTGGAASSGA	461	ENSORLGG00000011969	6	373	0
b2m	GCCACGTGAGTRATTTCCACCCC/ACGCTAYACRGYGGACYCTGA	508	ENSORLGG00000012506	23	235	183
gapdhs	CCCTGGCCAAAGTCATCCACGATA/CACCACTGACACATCGGCCACT	499	ENSORLGG00000006033	16	171	258
cnng1	CTGCTTGGCCCTGGCTCTCCT/AGCTGACTCAGGTATGGTCCGA	707	ENSORLGG00000005817	10	210	444
ptr-like	GCGGGTAGTGAATGTGAGTGGC/ACCCAAGACACCCAGCTCCA	436	ENSORLGG00000015652	24	368	0
enc1	CRGTTCCCTTGGCCTRTTGC/TGGGTGCCCTTTGACCAT	417	ENSORLGG00000003288	12	329	0
s7	CGTGCCATTTACTCTGGACTKGC/AACTCGTCYGGCTTCTCGCC	569	ENSORLGG00000018123	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
<i>Bathybates graueri</i>	9528	37.2	1.5	0
<i>Cyprichromis leptosoma</i>	9616	54.7	0.5	1
<i>Ophthalmotilapia ventralis</i>	9517	56.9	1.6	7
<i>Oreochromis tanganicae</i>	9552	31.9	1.2	1
<i>Cyphotilapia frontosa</i>	9532	52.2	1.4	1
<i>Eretmodus cyanostictus</i>	9437	14.4	2.4	5
<i>Tropheus moori</i>	9482	7.3	1.9	4
<i>Ctenochromis horei</i>	9480	1.9	1.9	2
<i>Astatotilapia burtoni</i>	9504	13.1	1.7	2
<i>Lamprologus callipterus</i>	9593	18.6	0.8	5
<i>Perissodus microlepis</i>	9489	14.7	1.9	0
<i>Neolamprologus pulcher</i>	9530	31.8	1.4	1
<i>Tylochromis polylepis</i>	9633	37.9	0.4	2
<i>Limnochromis abeelei</i>	9494	13.1	1.8	5
<i>Sargochromis spec.</i>	9523	11.9	1.5	7
<i>Pseudocrenilabrus philander</i>	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

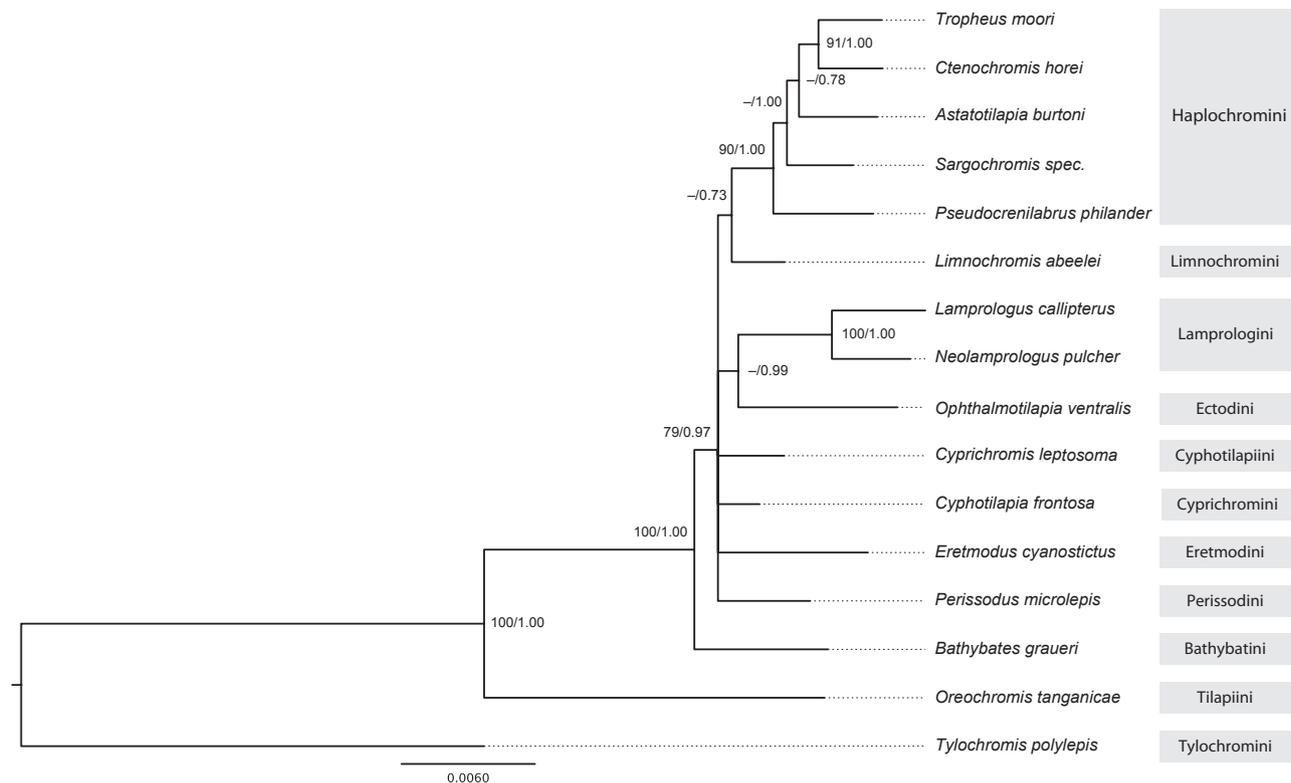


Fig. 1 Bayesian inference topology inferred with MRBAYES for the 16 species in our test data set and based on 24 genes (concatenated, 9669 bp). Bootstrap support for ML > 50 and posterior probability of the MRBAYES analysis > 0.50 are shown. Branch lengths are proportional to the number of mutations 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, Streebman 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.
- Brilo 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, Amodè 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-1 α 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, Schliwien 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 *Lepidolamprologus* (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, Schlieven 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 super-flock 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.