Short Communication

Outgroup effects on root position and tree topology in the AFLP phylogeny of a rapidly radiating lineage of cichlid fish

Paul C. Kirchberger [1], Kristina M. Sefc, Christian Sturmbauer, Stephan Koblmüller *

Department of Zoology, Karl-Franzens-University Graz, Universitätsplatz 2, A-8010 Graz, Austria

A B S T R A C T

Phylogenetic analyses of rapid radiations are particularly challenging as short basal branches and incomplete lineage sorting complicate phylogenetic inference. Multilocus data of presence-absence polymorphisms such as obtained by AFLP genotyping overcome some of the difficulties, but also present their own intricacies. Here we analyze >1000 AFLP markers to address the evolutionary history of the Limnochromini, a cichlid fish lineage endemic to Lake Tanganyika, and to test for potential effects of outgroup composition on tree topology. The data support previous mitochondrial evidence on the tribe's taxonomy by confirming the polyphyly of the genus *Limnochromis* and – in contradiction to a recent taxonomic revision – nesting the genus *Greenwoodochromis* within the Limnochromini. Species relationships suggest that ecological segregation occurred during the rapid basal radiation of the Limnochromini. The large phylogenetic distance between candidate outgroup taxa and the Limnochromini radiation caused random outgroup effects. Bootstrap support for ingroup nodes was lower in outgroup-rooted than in midpoint-rooted trees, and root positions and ingroup tree topologies varied in response to the composition of the outgroup. These observations suggest that the predisposition for homoplastic evolution makes AFLP-based phylogenetic analyses particularly susceptible to random biases introduced by too-distant outgroup taxa.

© 2013 The Authors. Published by Elsevier Inc. Open access under CC BY license.

1. Introduction

Adaptive radiations such as the Darwin's finches on the Galapagos archipelago, the Hawaiian silver swords, the Caribbean anoles lizards, and the cichlid fish of the East African Great Lakes provide opportunities for studying the processes underlying rapid speciation (Schluter, 2000; Brakefield, 2006; Salzburger, 2009). Yet, resolving the phylogeny of adaptive radiations still remains a challenge. Due to the extreme rapidity of the radiations, lineage sorting often lags behind cladogenesis, complicating phylogenetic inference based on single or a few molecular markers (Pamilo and Nei, 1988; Maddison and Knowles, 2006). Furthermore, the short basal branches typical for rapid radiations can make the phylogenetic reconstructions sensitive to the choice of outgroup taxa. In particular if the phylogenetic distance between outgroup and ingroup taxa is large, there is a high likelihood of homoplasy

between ingroup and outgroup taxa. Consequently, the outgroup may attach randomly to the ingroup and bias the inferred ingroup tree topology (Wheeler, 1990; Huelsenbeck et al., 2002; Rota-Stabelli and Telford, 2008; Rosenfeld et al., 2012). The outgroup effect on tree topology is well established for sequence data, but has so far received no attention in connection with multilocus binary presence–absence data such as those obtained by AFLP genotyping. However, the elevated potential for homoplasy in this type of data (Althoff et al., 2007; García-Pereira et al., 2010) implies that random outgroup effects on AFLP tree topologies are at least as likely as in analyses of DNA sequences. Despite certain other AFLP-specific shortcomings, which are associated with low information content, dominance and phenetics (e.g. García-Pereira et al., 2010), AFLP data nonetheless harbor potential for improved species tree approximation as the phylogenetic signal is integrated from numerous loci distributed throughout the genome, thus reducing the confounding effects of incomplete lineage sorting and hybridization on phylogenetic inference (Koopman, 2005). Consequently, AFLP markers are frequently being applied to taxa originating from rapid radiations (Ogden and Thorpe, 2002; Schliewen and Klee, 2004; Herder et al., 2006; Fink et al., 2010; Geiger et al., 2010; Koblmüller et al., 2010; Sturmbauer et al., 2010; Joyce et al., 2011).

The present study addresses the phylogeny of the Limnochromini (Cichlidae, Perciformes), a tribe of cichlid fish endemic to Lake Tanganyika. The deep-water habitat of this East African rift lake formed 5–6 million years ago (Tiercelin and Mondeguer, 1991) and provided a prolific environment for intralacustrine speciation (Koblmüller et al., 2008a). The Limnochromini originated in the course of a rapid radiation into several mouthbrooding cichlid tribes (the C-lineage, Clabaut et al., 2005), and subsequently diversified into 10 benthic, mainly deepwater-dwelling (up to depths >100 m), species (Coulter, 1991; Konings, 1998; Duftner et al., 2005). Previous work on the Limnochromini entailed repeated taxonomic revisions. Initially, the tribe included 13 species in 8 genera, *Benthochromis*, *Baileychromis*, *Gnathochromis*, *Greenwoodochromis*, *Limnochromis*, *Reganochromis*, *Tangachromis* and *Triglachromis* (Poll, 1986). Extending the morphological data, Takahashi (2003) erected two new tribes, the Benthochromini and the Greenwoodochromini, for the corresponding genera, and excluded *Gnathochromis pfefferi* from the Limnochromini. The establishment of the Benthochromini and the exclusion of *G. pfefferi* were later corroborated by mtDNA data, whereas the validity of Greenwoodochromini as a separate tribe was rejected (Duftner et al., 2005; Fig. 1A). Since questions on the taxonomy and evolutionary history of the Limnochromini have not been fully settled by the existing data, additional information was sought by employing AFLP markers in the present study. The phylogenetically unresolved radiation of the C-lineage, from which the Limnochromini originate (Clabaut et al., 2005), made it difficult to choose a suitable set of outgroup taxa from among the C-lineage tribes. Because of the large phylogenetic distance between the Limnochromini and the candidate outgroup taxa (Duftner et al., 2005), particular attention was paid to the effect of outgroup composition on ingroup topology.

2. Materials and methods

2.1. Ingroup and outgroup taxa

The Limnochromini were represented by 31 specimens representing all described species of the tribe, except for the small deep-water-dwelling *Tangachromis dhanisi*, which is seldom caught and rarely found in the aquarium trade. Outgroup effects were examined by comparing a midpoint-rooted tree with trees rooted with different sets of outgroup taxa, which were assembled from a pool of 17 species in six tribes representing the entire C-lineage (Takahashi and Koblmüller, 2011; Muschick et al., 2012; see "Phylogenetic analyses").

Fish were either caught on site using gill nets, acquired from aquarium trade or bought from local fishermen (Supplementary table 1).

2.2. DNA extraction and AFLP data collection

DNA extraction and AFLP genotyping (ten primer combinations for selective amplification: *Eco*RI-ACA/*Mse*I-CAA, *Eco*RI-ACA/*Mse*I-CAG, *Eco*RI-ACA/*Mse*I-CAC, *Eco*RI-ACA/*Mse*I-CAT, *Eco*RI-ACT/*Mse*I-CAA, *Eco*RI-ACT/*Mse*I-CAG, *Eco*RI-ACT/*Mse*I-CAC, *Eco*RI-ACT/*Mse*I-CAT, *Eco*RI-ACC/*Mse*I-CAA, *Eco*RI-ACT/*Mse*I-CAC) were performed as described in Egger et al. (2007). Initial scoring of electropherograms (size range of 50–300 bps for each primer combination; bin width was adjusted manually for each peak) was done using GeneMapper 3.7 (Applied Biosystems). AFLPScore 1.4a (Whitlock et al., 2008) was then used to convert the un-normalized peak-height data into a presence/absence matrix. 20 replicate samples were used to estimate the mismatch error rate, and a rate of <2% was set as threshold for the inclusion of data in the final matrix.

2.3. Phylogenetic analyses

PAUP 4.05b (Swofford, 2000) was used to construct a neighbor joining (NJ) tree based on Nei-Li distances (Nei and Li, 1979) and estimate statistical support from 1000 bootstrap replicates. To test for effects of outgroup composition, analyses were conducted with only ingroup taxa (the "ingroup dataset") and with eight different sets of outgroup taxa, i.e. the "full dataset" using the complete outgroup (17 taxa from six tribes, encompassing the diversity of the entire C-lineage), and seven datasets with smaller outgroups (Ectodini; a monophyletic clade of Perissodini + Benthochromini + Cyprichromini; Cyprichromini; Benthochromini; Perissodini; Cyphotilapiini; Tropheini). For the ingroup dataset, a Neighbor-Net analysis based on a Nei and Li (1979) distance matrix was conducted in SplitsTree v.4.10 (Huson and Bryant, 2006) as an explorative approach for identifying conflicting phylogenetic signal in the data.

To test for consistency between mtDNA- and AFLP-based tree topologies, we evaluated the fit of the mtDNA sequence data (Duftner et al., 2005) to the unrooted NJ tree topology derived from ingroup AFLP data, by constraining a maximum likelihood (ML) tree search with the substitution model GTR + I + G (Duftner et al., 2005) to the AFLP tree. The Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999; full optimization, 10,000 bootstrap replicates; implemented in PAUP) was then used to test whether the likelihood of the mtDNA data was significantly lower under the AFLP-NJ topology than under an unconstrained topology optimized from the mtDNA data.

3. Results

3.1. Phylogenetic reconstruction of the ingroup dataset

Phylogenetic analyses were based on 1128 AFLP marker loci. Bootstrap support for the monophyly of all species and for the reconstructed species clades was relatively good and ranged from 71% to 100% (Fig. 1B). Polyphyly of the genus *Limnochromis* was confirmed, as *L. auritus* was most closely related to *T. otostigma*, and *L. abeelei* and *L. staneri* clustered with the genus *Greenwoodochromis*. *Gnathochromis permaxillaris* resulted as sister group to the *L. abeelei*/*L. staneri*/*Greenwoodochromis* clade and *B. centropomoides* plus *R. calliurus* constituted a well supported monophylum. Mid-point rooting placed *T. otostigma* + *L. auritus* as sister to the remaining limochromines. The SH-test ($p = 0.059$) revealed no significant differences between the AFLP topology of the ingroup dataset and the phylogenetic tree reconstructed from mitochondrial DNA sequence data by Duftner et al. (2005).

The phylogenetic relationships inferred by the NeighborNet analysis (Fig. 1C; least squares fit value = 95.633) were consistent with the tree-based inferences. Most of the internal splits in the network are represented by boxes, which is indicative of conflicting signals in the data. Incomplete or differential lineage sorting of marker loci as well as reticulation due to hybridization can underlie these ambiguities.

3.2. Outgroup effects

The NJ analysis of the full dataset (Fig. 2) confirmed the monophyly of the tribe Limnochromini (including the Greenwoodochromini). In contrast to the analysis of the ingroup dataset, statistical support for most of the basal nodes within the Limnochromini was low. The eight tested outgroup sets gave rise to five different tree topologies, with variability in root position, ingroup topology and bootstrap support (Fig. 2).