Molecular Phylogenetics and Evolution 82 (2015) 95–110

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Cryptic speciation in the white-shouldered antshrike (*Thamnophilus aethiops*, Aves – Thamnophilidae): The tale of a transcontinental radiation across rivers in lowland Amazonia and the northeastern Atlantic Forest

Gregory Thom [a], Alexandre Aleixo [b,*]

[a] Curso de Pós-graduação em Zoologia, Universidade Federal do Pará/Museu Paraense Emílio Goeldi, Caixa Postal 399, CEP 66040-170 Belém, PA, Brazil
[b] Coordenação de Zoologia, Museu Paraense Emílio Goeldi, Caixa Postal 399, CEP 66040-170 Belém, PA, Brazil

ARTICLE INFO

Article history:
Received 23 June 2014
Revised 20 September 2014
Accepted 26 September 2014
Available online 5 October 2014

Keywords:
Diversification hypotheses
Historical biogeography
Phylogeography
Population genetics
Refuges
Taxonomy

ABSTRACT

The growing knowledge on paleogeography and the recent applications of molecular biology and phylogeography to the study of the Amazonian biota have provided a framework for testing competing hypotheses of biotic diversification in this region. Here, we reconstruct the spatio-temporal context of diversification of a widespread understory polytypic Amazonian bird species (*Thamnophilus aethiops*) and contrast it with different hypotheses of diversification and the taxonomy currently practiced in the group. Sequences of mtDNA (cytochrome *b* and ND2) and nuclear (β-fibrinogen introns 5 and 7 and the Z-liked Musk4) genes, adding up to 4093 bp of 89 individuals covering the Amazonian, Andean, and Atlantic Forest populations of *T. aethiops* were analyzed. Phylogenetic and population genetics analyses revealed ten reciprocally monophyletic and genetically isolated or nearly-isolated lineages in *T. aethiops*, highlighting several inconsistencies between taxonomy and evolutionary history in this group. Our data suggest that the diversification of *T. aethiops* started in the Andean highlands, and then proceeded into the Amazonian lowlands probably after the consolidation of the modern Amazonian drainage. The main cladogenetic events in *T. aethiops* may be related to the formation and structuring of large Amazonian rivers during the Late Miocene–Early Pleistocene, coinciding with the dates proposed for other lineages of Amazonian organisms. Population genetics data do not support climatic fluctuations as a major source of diversification in *T. aethiops*. Even though not entirely concordant with paleobiogeographic models derived from phylogenies of other vertebrate lineages, our results support a prominent role for rivers as major drivers of diversification in Amazonia, while underscoring that different diversification scenarios are probably related to the distinct evolutionary origins of groups being compared.

1. Introduction

Since the 19th century, the high species diversity of the Amazon allied to multifaceted patterns of geographical distributions instigated naturalists and researchers to explain such complexity (Wallace, 1852; Haffer, 1969; Bates, 2001).

Despite the large number of distinct hypotheses to explain Amazonian biodiversity (e.g., refuge hypothesis, Haffer, 1969; riverine hypothesis, Wallace, 1852; Ayres and Clutton-Brock, 1992; riverine-refuge hypothesis, Ayres and Clutton-Brock, 1992; Haffer, 1993, 2001; ecological gradients hypothesis, Endler, 1977; "museum" hypothesis, Roy et al., 1997; and marine incursions, Bates, 2001), and the common sense that many causations have operated for the formation of such diversity (Bush, 1994; Haffer, 2001; Miller et al., 2008), recent phylogeographic and paleobiogeographic reconstructions (Aleixo and Rossetti 2007; Patel et al., 2011; Weir and Price, 2011; Ribas et al., 2012) have postulated the formation of the current Amazonian physical landscape as the main source of cladogenetic events among the studied lineages.

The formation of the Amazon basin was ultimately driven by the Andean uplift and related arches, which resulted in the formation of a large fluvio-lacustrine system in western Amazonia, during the early Miocene (Espurt et al., 2010; Mora et al., 2010). With the continuous rise of the Andes in the middle Miocene, this system probably expanded to the Purus Arch, a tectonic structure located 300 km west of Manaus, and which separates the Solimões and Amazonas sedimentary basins (Figueiredo et al., 2009; Hoorn

* Corresponding author.
E-mail address: aleixo@museu-goeldi.br (A. Aleixo).

http://dx.doi.org/10.1016/j.ympev.2014.09.023
1055-7903/

et al., 2010). The period when the Purus Arch was breached probably marked the onset of the development of the transcontinental Amazon River, draining eastward the western fluvio-lacustrine system, and enabling the formation of upland *terra-firme* forests in the western Amazonia; however, the timing of this event is still a matter of controversy, with two main proposed periods based on different sources of evidence: late Miocene, between 6.5 and 10 Ma (Hoorn et al., 2010; Latrubesse et al., 2010) and Pliocene (approximately 2.5 Ma; Campbell et al., 2006). Despite the fact that major interfluves of the Amazonian region delineate the geographical distribution of most endemic vertebrate taxa and that this has provided the basis for the recognition of the so called areas of endemism (Cracraft, 1985; Silva et al., 2005), the process of river dynamics driving population divergence is still poorly understood, with multiple area-relationships patterns documented so far (Aleixo, 2004; Fernandes et al., 2012; Ribas et al., 2012; d'Horta et al., 2013). This suggests that although rivers have been barriers to gene flow, other historical events probably contributed to the overall diversification pattern. However, this knowledge is still meager when compared with the rich and complex biodiversity of Amazonia, preventing the recognition of a general model or sets of heuristic models of diversification, corroborated by several lineages.

The polytypic species *Thamnophilus aethiops* (Aves: Thamnophilidae) is a good model to study the paleobiogeography of the Amazon basin due the following reasons: (1) it is widely distributed in Amazonia and neighboring areas such as the Andes and the Atlantic Forest in eastern Brazil, two areas known to be historically connected to the development of modern Amazonia; (2) it is restricted to the understory of upland *terra-firme* forest, rapidly responding to environmental change (Stotz et al., 1996); and (3) subspecies distributions are mainly bounded by the major tributaries of the Amazon River, with some important exceptions that can provide insights into the circumstances whereby a single lineage does and does not respond to rivers as barriers. Thus, we estimated the spatio-temporal scenario of diversification of the polytypic *T. aethiops* to address the following questions: (1) what are the evolutionary relationships among the populations/subspecies of *T. aethiops*? (2) what are the relationships between the events that led to diversification in *T. aethiops* and the paleogeographic models proposed for the formation of Amazonia? (3) do the demographic history of the studied populations support any particular diversification hypothesis?

2. Material and methods

2.1. Samples, laboratory procedures, and data analyses

A total of 89 individuals from 55 localities were sampled throughout the distribution of *T. aethiops* in Amazonia, northeastern Atlantic Forest and the Andean foothills, covering nine of ten described subspecies (Zimmer and Isler, 2003; Fig. 1b, Table 1). Only *T. a. wetmorei*, purportedly endemic to the Andean foothills of Colombia, was not sampled. Nonetheless, we assume here that that this taxon is closely related to the *polionotus* subspecies from northwestern Amazonia, from which it is hardly differentiated based on plumage characters (Zimmer and Isler, 2003). Subspecific identification of our samples were made through the inspection of voucher specimens and followed the most recent taxonomy proposed for *T. aethiops* (Zimmer and Isler 2003). We used *Thamnophilus aroyae*, *T. unicolor*, and *T. caerulescens* as outgroups due their close relationship to *T. aethiops* (Brumfield and Edwards, 2007).

Total DNA was extracted from approximately 20 mg of muscle tissue following a standard phenol/chloroform protocol (Sambrook et al., 1989) or with the aid of a DNeasy Quiagen (Hilden, Germany) extraction kit. We amplified the mitochondrial genes cytochrome *b* (cyt *b*, 992 bp, $n = 88$) and NADH dehydrogenase subunit 2 (ND2, 1041 bp, $n = 79$), as well as the nuclear genes β-fibrinogen intron 7 (Bf7, 963 bp, $n = 51$), β-fibrinogen intron 5 (Bf5, 532 bp, $n = 63$), and the intron 4 of the skeletal muscle receptor of tyrosine kinase (Musk4, 565 bp, $n = 64$) linked to the Z chromosome (see Table 2 for the primers used). Polymerase chain reaction (PCR) was performed with 25 μl of final volume, and approximately 50 ng of genomic DNA, 1.5–2.5 mM of $MgCl_2$, 200 mM of dNTPs and 0.1 U of Taq DNA polymerase Promega (Madison, WI, USA). Reactions started with a denaturation step at 94 °C for 5 min, followed by 35 cycles of three steps: (1) 94 °C for 1 min; (2) annealing temperatures ranging from 50 °C (cyt *b* and Bf5) and 52 °C (Musk4) to 59 °C (ND2, Bf7) for 1 min; and (3) 72 °C for 1 min. The last step, for the extension, was at 72 °C for 5 min. PCR products were purified with the Polyethylene glycol protocol (PEG). Sequencing was carried out on an ABI 3130 automated capillary sequencer (Applied Biosystems, Foster City, California, USA) with the ABI Prism Big Dye terminator Kit. To confirm observed mutations, both strands of each sample were sequenced. The DNA strands were edited and aligned manually in BIOEDIT 7.0.5 (Hall, 1999). Saturation of the nucleotide substitutions in the mitochondrial DNA were evaluated using the software DAMBE (Xia and Xie, 2001). For nuclear loci, heterozygous nucleotide positions were inferred by the presence of double peaks of the same size in the electropherogram. To obtain the gametic phase of haplotypes of the heterozygous individuals, we used a Bayesian approach as implemented in PHASE 2.1 1 (Stephens et al., 2001; Stephens and Donnelly, 2003; Stephens and Scheet, 2005). The threshold of 70% of posterior probability was assumed as a confidence value for the analyzed haplotypes (see Harrigan et al., 2008). Lower values were regarded as ambiguous. To obtain the gametic phase for individuals heterozygous in size (presence of indels in one of the strands), we used the program CHAMPURU v1.0 (Flot et al., 2006; Flot, 2007; available online at http://www.mnhn.fr/jfflot/champuru/). Females did not have the gametic phase estimated for Musk4, since this gene is linked to de Z chromosome. To test the hypothesis of neutral evolution among independent loci we performed the Tajima's D test (Tajima, 1989) using DNASP v.4.10.9 (Rozas et al., 2003). The significance for these tests was assessed through 10.000 coalescence simulations. To check for possible recombination among the nuclear loci we used the *phi* test as implemented in the SPLITSTREE 4.10 software (Huson and Bryant, 2006).

2.2. Phylogenetic analyses

2.2.1. Concatenated genes and gene trees

The phylogenetic analyses were performed using Bayesian inference (BI) in MrBAYES 3.1.2 (Ronquist and Huelsenbeck, 2003) and Maximum likelihood (ML) in RaxML 7.0.3 (Stamatakis, 2006). The evolutionary models which best explain the evolution of each dataset were selected in JMODELTEST 0.1.1 (Posada, 2008), using the Bayesian information criterion (BIC) for BI and the Akaike information criterion (AIC) for ML. For the BI analyses with the concatenated dataset, a Bayes factor analysis (Brandley et al., 2005) selected four partitions (mtDNA + one separate partition for each nuclear gene) as the best partitioning scheme. The evolutionary models for each partition were: (1) mtDNA (GTR + G); (2) Bf5 (GTR + I + G); (3) Bf7 (GTR + I); and (4) Musk4 (HKY + G). The BI estimated based on mtDNA only used one model for the two genes (GTR + G). BI analyses were generated through two independent runs of 1×10^7 generations, each with four Markov chains. The parameters of the chains were sampled every 1000 generations and the first 1000 trees were discarded as burn-in. The posterior probabilities for each estimated node were obtained