Molecular Phylogenetics and Evolution 65 (2012) 317–322

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Species trees for the tree swallows (Genus *Tachycineta*): An alternative phylogenetic hypothesis to the mitochondrial gene tree

Roi Dor [a,*], Matthew D. Carling [a,b], Irby J. Lovette [a,c], Frederick H. Sheldon [d,e], David W. Winkler [a,c]

[a] Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA
[b] Berry Biodiversity Conservation Center, Department of Zoology and Physiology, University of Wyoming, Laramie, WY 82071, USA
[c] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14950, USA
[d] Museum of Natural Science, 119 Foster Hall, Louisiana State University, Baton Rouge, LA 70803, USA
[e] Department of Biological Sciences, 119 Foster Hall, Louisiana State University, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 10 October 2011
Revised 12 May 2012
Accepted 23 June 2012
Available online 30 June 2012

Keywords:
Nuclear
Multi-locus
Phylogeny
Bayesian
*BEAST
BEST

ABSTRACT

The New World swallow genus *Tachycineta* comprises nine species that collectively have a wide geographic distribution and remarkable variation both within- and among-species in ecologically important traits. Existing phylogenetic hypotheses for *Tachycineta* are based on mitochondrial DNA sequences, thus they provide estimates of a single gene tree. In this study we sequenced multiple individuals from each species at 16 nuclear intron loci. We used gene concatenated approaches (Bayesian and maximum likelihood) as well as coalescent-based species tree inference to reconstruct phylogenetic relationships of the genus. We examined the concordance and conflict between the nuclear and mitochondrial trees and between concatenated and coalescent-based inferences. Our results provide an alternative phylogenetic hypothesis to the existing mitochondrial DNA estimate of phylogeny. This new hypothesis provides a more accurate framework in which to explore trait evolution and examine the evolution of the mitochondrial genome in this group.

1. Introduction

The swallow genus *Tachycineta* comprises nine species which inhabit most of the New World from Alaska and Canada in the north to the southern tip of South America (Whittingham et al., 2002; Turner, 2004). These nine species are similar in appearance and in some aspects of their ecology. *Tachycineta* swallows adopt cavities as their nesting sites; therefore they readily accept nest boxes, a feature that facilitates comparative research on the group. Most previous studies have focused on the North American species *T. bicolor*, which is a model species for many aspects of ecology and behavior (Winkler and Allen, 1996; Ferretti and Winkler, 2009; Winkler et al., 2011). Recently, other *Tachycineta* species have been studied mainly in an effort to understand general patterns of variation in breeding biology across the distribution of *Tachycineta* (Massoni et al., 2007; Liljesthrom et al., 2009; Ferretti et al., 2011; Dor et al., 2012; see also http://golondrinas.cornell.edu/). In order to conduct comparative analyses of recent ecological and behavioral data available for *Tachycineta* species (see below) we require an accurate phylogenetic framework in which to examine the evolution of traits.

To date, existing phylogenetic reconstructions of *Tachycineta* have been based on mitochondrial DNA (mtDNA) sequences (Whittingham et al., 2002; Cerasale et al., 2012) or on mitochondrial DNA with one nuclear intron (βfib7; Sheldon et al., 2005). The Sheldon et al. (2005) study was based on Whittingham et al. (2002) and used in addition to mtDNA sequences (ND2 and cyt-b), one intron sequence (βFib7) for four out of nine *Tachycineta* species. Both studies used maximum likelihood and Bayesian analyses on the concatenated sequences. These estimates of phylogeny all divide *Tachycineta* into two clades: one consisting of the four South American (*T. albiventer*, *T. stolzmanni*, *T. leucorrhoa* and *T. meyeni*) and one Central American (*T. albilinea*) species, and the other consisting of two North American (*T. bicolor* and *T. thalassina*) and two Caribbean (*T. cyaneoviridis* and *T. euchrysea*) species. However, several nodes in the phylogenetic trees are not fully resolved, including the position of *T. stolzmanni* within the South and Central American clade and the relationship between *T. thalassina* and the two Caribbean species (*T. cyaneoviridis* and *T. euchrysea*). The main difference between Sheldon et al. (2005) and Whittingham et al. (2002) trees concerned *T. thalassina*, *T. cyaneoviridis* and *T. euchrysea*, but in any case their inter-relationships have low support values in both studies. The new phylogenetic estimate based

* Corresponding author. Present address: Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309, USA. Fax: +1 303 492 8699.
E-mail address: rd324@cornell.edu (R. Dor).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.06.020

on entire mtDNA sequences (Cerasale et al., 2012) confirmed the Sheldon et al. (2005) estimate and provided a resolved mtDNA tree.

Mitochondrial DNA is expected to experience shorter coalescence times relative to nuclear DNA (nDNA) loci and therefore has been considered a robust marker for inferring phylogeny (Moore, 1995; Zink and Barrowclough, 2008). However, nuclear DNA markers on different chromosomes represent independent gene genealogies whereas mitochondrial DNA represents only one maternally-inherited gene genealogy (Edwards and Bensch, 2009). Moreover, although mitochondrial DNA is considered a neutral locus, there is evidence that selection may affect patterns of mitochondrial DNA variation (reviewed in Hudson and Turelli, 2003), and some studies have found that the mitochondrial DNA tree does not represent the species tree (Carling and Brumfield, 2008; Leache, 2010). Thus, single-locus mitochondrial DNA trees are less likely to represent the phylogeny of a group as accurately as trees based on multiple independent loci (Edwards et al., 2005; Edwards and Bensch, 2009).

Comprehensive phylogenetic hypotheses are required to conduct comparative analyses of trait evolution and, thus, study the evolutionary history of taxonomic groups (Brooks and McLennan, 1991; Harvey and Pagel, 1991). *Tachycineta* swallows have a wide geographic distribution and inhabit diverse habitats. Accordingly, they exhibit variation both within- and among-species in ecologically important traits such as body size, clutch size, time of breeding and extra-pair paternity (Turner and Rose, 1989; Turner, 2004) making them excellent candidates for comparative analyses of trait variation. In this study, we generated a multi-locus dataset based on nuclear DNA sequences in order to reconstruct phylogenetic relationships within *Tachycineta*. This multi-locus nDNA tree provides an alternative phylogenetic hypothesis to the mtDNA trees previously published, will provide a tool for examining trait evolution in this group, and will introduce an independent framework for examining the evolution of mtDNA genes in this group.

2. Methods

2.1. Sampling and laboratory methods

We analyzed DNA sequences of 36 individuals representing all nine generally recognized species in the genus *Tachycineta* (Peters, 1960; Dickinson, 2003) (Table 1). *Progne chalybea* (Grey-breasted Martin) and *Stelgidopteryx serripennis* (Northern Rough-winged Swallow), members of the sister clade of *Tachycineta* (Sheldon et al., 2005), were employed as outgroups for phylogenetic reconstruction. Genomic DNA was obtained from pectoral muscle using DNeasy tissue extraction kits (Qiagen, Valencia, CA) and from blood using Perfect gDNA Blood Mini kits (Eppendorf, Westbury, NY).

Each individual was amplified and sequenced at 16 nuclear introns (primers are located in exons) (Table 2), using the following PCR conditions in a 10 μL amplification reaction: 1 μL undiluted DNA, 10 μM Tris–HCl , 50 μM KCl, 4 mM $MgCl_2$, 0.25 mM of each nucleotide, 0.25 mM of each primer, and 0.025 U Jumpstart Taq polymerase (Sigma). PCR amplification conditions were: initial denaturation at 95 °C for 4 min 30 s; 30 cycles of denaturating at 95 °C for 1 min, locus-specific temperature primer annealing step (Table 2) for 1 min, and extension at 72 °C for 1 min 20 s; and a final extension at 72 °C for 4 min 30 s. We checked for amplification by electrophoresing 2 μL of each PCR amplicon on a 2% agarose gel.

PCR products were purified using exonuclease and shrimp alkaline phosphatase enzymatic reactions (United States Biochemical). Purified products were cycle sequenced in both directions using amplification primers and ABI BigDye Terminator v 3.1. Sequencing products were cleaned using Sephadex columns and processed in an ABI 3730 Automated DNA Analyzer (Applied Biosystems). We aligned forward and reverse strands for each specimen using Sequencher 4.7 (Gene Codes Corp.) and confirmed alignment by eye. We used standard IUPAC codes for ambiguities due to heterozygosity (R, Y, S, W, K, M), calling overlapping peaks (more than 50%) as heterozygous. All sequence data are deposited in GenBank (Accession Nos. JX298934–JX299532, Table 2).

2.2. Concatenated sequence analysis

To estimate the phylogeny, we used maximum likelihood (ML) and Bayesian analysis methods implemented in RAxML-HPC2 v7.2.8 (Stamatakis, 2006) and MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), respectively, on the concatenated sequences. RAxML was run on the CIPRES cluster (Miller et al., 2010) and MrBayes on the Computational Biology Service Unit at Cornell University. We considered posterior probabilities values of $\geqslant 0.9$ and bootstrap values of $\geqslant 70$ to represent credible support values for nodes.

Maximum likelihood (ML) analyses, using RAxML, were conducted using the GTR + G model for each partition (by gene) with 1000 bootstrap replicates. For the Bayesian analyses, using MrBayes, we identified the most appropriate substitution model for each partition for by comparing the AIC scores of the 28 possible models in FindModel (Tao et al., 2005), and we applied the most similar model (the one more parameterized) available on MrBayes (Table 2). In each MrBayes analysis four independent runs, each with four chains, were conducted for 20 million generations (sampling every 1000 generations). Convergence was assessed by examining the cumulative posterior probabilities of clades and the correlation of split frequencies between runs (AWTY; Nylander et al., 2008). The first 2500 trees (2,500,000 generations) were discarded as burnin, and the remainder were used to estimate tree parameters and topology. We also performed constrained Bayesian analyses to compare the previously published mitochondrial trees (Whittingham et al., 2002; Sheldon et al., 2005) with the tree generated in this study. We used Bayes factors (logarithms of the harmonic means) to assess the difference between unconstrained and constrained trees (Kass and Raftery, 1995).

2.3. Species tree analysis

We used the programs *BEAST (Bayesian Evolutionary Analysis Sampling Trees; Heled and Drummond, 2010) and BEST (Bayesian Estimation of Species Trees; Liu and Pearl, 2007) to estimate the species tree from non-concatenated gene sequences of *Tachycineta*. Whereas BEST uses a two-stage algorithm to infer the species tree, *BEAST attempts to sample gene trees and species tree simultaneously, therefore is more computationally efficient than BEST. To assess convergence of *BEAST results we repeated the *BEAST analysis in 17 separate runs, each for at least 230 million iterations (Mean: 275 million iterations, range: 230–407 million iterations) for ~5 days (the maximum allowance on the Computational Biology Service Unit at Cornell University). We used similar substitution models as the ones used for the MrBayes analysis (variables estimated for each locus independently), and we used the uniform default *BEAST priors. The Yule process was used as species tree prior, and the strict clock as the molecular clock model. We assessed convergence by comparing the topologies and support values of all independent *BEAST runs, and then we combined the results after burnin from all runs using LogCombiner v1.6.1 (Drummond and Rambaut, 2007) and examined the effective sample sized (ESS) in TRACER v1.5 (Drummond and Rambaut, 2007). In each run we discarded the first half of the iterations as burnin, and every 50,000th tree was kept afterwards. *Progne chalybea* and