Molecular Phylogenetics and Evolution xxx (2012) xxx–xxx

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 xxxx

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

Please cite this article in press as: Dor, R., et al. Species trees for the tree swallows (Genus *Tachycineta*): An alternative phylogenetic hypothesis to the mitochondrial gene tree. Mol. Phylogenet. Evol. (2012), http://dx.doi.org/10.1016/j.ympev.2012.06.020

Table 1
Specimens and localities included in this study.

Species	Common name	Type[a]	Collection locality	Museum[b]	Sample #
Tachycineta bicolor	Tree swallow	T	USA, NY	CUMV	50502
T. bicolor	Tree swallow	T	USA, NY	CUMV	50315
T. bicolor	Tree swallow	T	USA, NY	CUMV	51647
T. bicolor	Tree swallow	T	USA, NY	CUMV	51457
T. thalassina	Violet-green swallow	T	USA, CA	LSUMNS	B19363
T. thalassina	Violet-green swallow	T	USA, AK	CUMV	50936
T. thalassina	Violet-green swallow	T	USA, NM	LSUMNS	B36165
T. thalassina	Violet-green swallow	T	USA, TX	LSUMNS	B38776
T. thalassina	Violet-green swallow	T	USA, WA	LSUMNS	B43265
T. euchrysea	Golden swallow	T	Dominican rep.	LSUMNS	B22018
T. euchrysea	Golden swallow	T	Dominican rep.	LSUMNS	B22021
T. euchrysea	Golden swallow	B	Dominican rep.		ATOB535
T. cyaneoviridis	Bahama swallow	T	Bahamas	LSUMNS	B48916
T. cyaneoviridis	Bahama wallow	T	Bahamas	LSUMNS	B59081
T. cyaneoviridis	Bahama swallow	T	Bahamas	LSUMNS	B59082
T. cyaneoviridis	Bahama swallow	T	Bahamas	LSUMNS	B59203
T. leucorrhoa	White-rumped swallow	T	Bolivia	LSUMNS	B38190
T. leucorrhoa	White-rumped swallow	T	Bolivia	LSUMNS	B38192
T. leucorrhoa	White-rumped swallow	T	Paraguay	LSUMNS	B25995
T. leucorrhoa	White-rumped swallow	T	Bolivia	LSUMNS	B38028
T. leucorrhoa	White-rumped swallow	T	Bolivia	LSUMNS	B38194
T. meyeni	Chilean swallow	T	Chile	LSUMNS	B14014
T. meyeni	Chilean swallow	T	Chile	AMNH	12228
T. meyeni	Chilean swallow	T	Chile	AMNH	12229
T. meyeni	Chilean swallow	T	Argentina	AMNH	13475
T. stolzmanni	Tumbes swallow	T	Ecuador	LSUMNS	B25372
T. stolzmanni	Tumbes swallow	T	Ecuador	LSUMNS	B25373
T. albilinea	Mangrove swallow	T	Belize	CUMV	50162
T. albilinea	Mangrove swallow	T	Costa Rica	LSUMNS	B27292
T. albilinea	Mangrove swallow	T	Costa Rica	LSUMNS	B27312
T. albilinea	Mangrove swallow	T	Panama	LSUMNS	B28416
T. albilinea	Mangrove swallow	T	Panama	LSUMNS	B28417
T. albiventer	White-winged swallow	T	Bolivia	LSUMNS	B12853
T. albiventer	White-winged swallow	T	Peru	LSUMNS	B28089
T. albiventer	White-winged swallow	T	Bolivia	LSUMNS	B12860
T. albiventer	White-winged swallow	T	Venezuela	AMNH	2982
Progne chalybea	Grey-breasted martin	T	Uruguay	CUMV	50672
Stelgidopteryx serripenis	Northern rough-winged swallow	T	USA, NY	CUMV	51710

[a] Specimen type: B – blood; T – frozen or buffered–preserved tissue.
[b] Institutional Source of samples: AMNH: American Museum of Natural History, New York, NY, USA; CUMV: Cornell University Museum of Vertebrates, Ithaca, NY, USA; LSUMNS: Louisiana State University Museum of Natural science, Baton Rouge, LA, USA.

Table 2
Locus type, sequence location in chicken genome, sequence length, number of variable sites, PCR conditions, nucleotide substitution model and GenBank Accession numbers for 16 intron loci used in this study.

Locus	Chromosome[a]	Aligned length (bp)	Variable sites (bp)	Annealing temp. (°C)	Nucleotide substitution model[b]	Accession Nos.
12884[c]	1	463	33	56	HKY	JX299495–532
11074[c]	20	480	20	56	HKY	JX299457–494
16214[c]	4	431	12	64	F81	JX299419–456
08352[c]	5	455	23	56	JC	JX299382–418
27356[c]	1	385	23	56	TrN	JX299346–381
βact3[d]	NA	463	31	64	TrN	JX299309–345
00504[c]	23	465	27	56	TVM	JX299272–308
tropo6[e]	NA	495	8	64	F81	JX298934–970
00811[c]	27	619	30	56	K81uf	JX299234–271
02154[c]	18	859	67	56	TVM	JX299198–233
04080[c]	28	472	43	56	K80 + Γ	JX299160–197
15463[c]	2	698	68	56	TrN	JX299122–159
09092[c]	12	771	38	64	F81	JX299084–121
21281[c]	1	535	31	64	HKY + Γ	JX299046–083
27623[c]	1	720	64	60	HKY + Γ	JX299008–045
ODC[e]	NA	715	45	60	TrN	JX298971–007

[a] Chromosome of best BLAST hit in the chicken genome (see Backström et al., 2008).
[b] Nucleotide substitution models selected (out of 28 possible models on FindModel) based on AIC scores criterion.
[c] Backström et al. (2008).
[d] Carling and Brumfield (2008).
[e] Primmer et al. (2002).

Stelgidopteryx serripennis, members of the sister clade, were analyzed together with ingroup members in *BEAST.

To estimate the *Tachycineta* species tree using BEST, we ran four independent analyses, each with a single chain, for 300 million

generations and sampled trees every 50,000 generations. We used a GTR + gamma substitution model, in which the variables for each locus were estimated independently, and used the default BEST settings for thetapr (invgamma (3,0.03)) and genemupr (uniform (0.5,1.5)). Because only one outgroup can be used in BEST, we included only *Stelgidopteryx serripennis* in this analysis. We assessed convergence by comparing visually the topologies of the four independent BEST analyses (Linnen and Farrell, 2008). Given the lack of convergence among the BEST runs (see below), we only briefly discuss the BEST results.

3. Results

3.1. Data partitioning and model selection

The combined 16-locus data matrix consisted of 9026 bp and 563 variable characters (Table 2). The tropo6 locus (495 bp) contributed the fewest number of variable sites (8), and the 15463 locus (698 bp) contributed the highest number of variable sites (68). The nucleotide substitution models selected varied among nuclear loci (Table 2).

3.2. Gene trees and concatenated sequence analysis

We generated Bayesian individual gene trees for each of the 16 nDNA loci using MrBayes (Supplementary material). Individual gene trees provided little resolution at most of the basal nodes or between species pairs. Thus the combined result of all loci was not driven by one specific locus, and there were no significant incongruences among the gene trees.

The partitioned Bayesian analysis of the concatenated matrix of 16 nuclear loci supported the monophyly of the South and Central American clade (Fig. 1). Strong support was provided also for the clade that consisted of *T. thalassina* and the two island species, *T. euchrysea* and *T. cyaneoviridis* (posterior probability = 1.0). *T. bicolor* was positioned as sister to all other *Tachycineta* taxa (posterior probability = 0.93). The concatenated nuclear data did not support the reciprocal monophyly of two pairs of species from the South and Central American clade: *T. leucorrhoa*/*T. meyeni* and *T. albiventer*/*T. albilinea*. We used the constraint option in MrBayes to compare the previously published phylogenies (Whittingham et al., 2002; Sheldon et al., 2005) with the phylogeny generated in this study. The newer phylogeny (lnL = −17030.8) was a significantly better fit to the data than trees constrained to the mitochondrial phylogeny of Whittingham et al. (2002) or the mitochondrial and βfib7 phylogeny of Sheldon et al. (2005) (lnL = −17092.4 and −17086.2, respectively; difference in Bayes Factors >50, representing a very strong result (Kass and Raftery, 1995)).

Maximum likelihood analysis provided support for the South and Central American clade. However, support for the clade that consisted of *T. thalassina* and the two island species was lower (bootstrap = 74; Fig. 2). In this analysis, *T. bicolor* was placed as sister to the South and Central American clades, although support for this relationship was not high either (bootstrap = 74). As in the Bayesian analysis, *T. leucorrhoa* and *T. albiventer* were paraphyletic in the maximum likelihood analysis.

3.3. Species tree analysis

We combined results from 17 independent *BEAST runs to reconstruct the phylogeny of *Tachycineta* via gene tree comparisons. The topology and support values generated by each of the 17 independent analyses were similar. However, the effective sample sizes (ESSs) of parameters of interest (gene trees and species-tree root-age; plotted on TRACER v1.5) were above 200 only after we combined the post-burnin trees from all independent runs. This analysis supports the South and Central American clade, but the position of *T. stolzmanni* within this clade is not resolved (Fig. 3). As in the Bayesian concatenated analysis, *T. bicolor* was placed as sister to all other *Tachycineta*. However, *T. cyaneoviridis* appeared

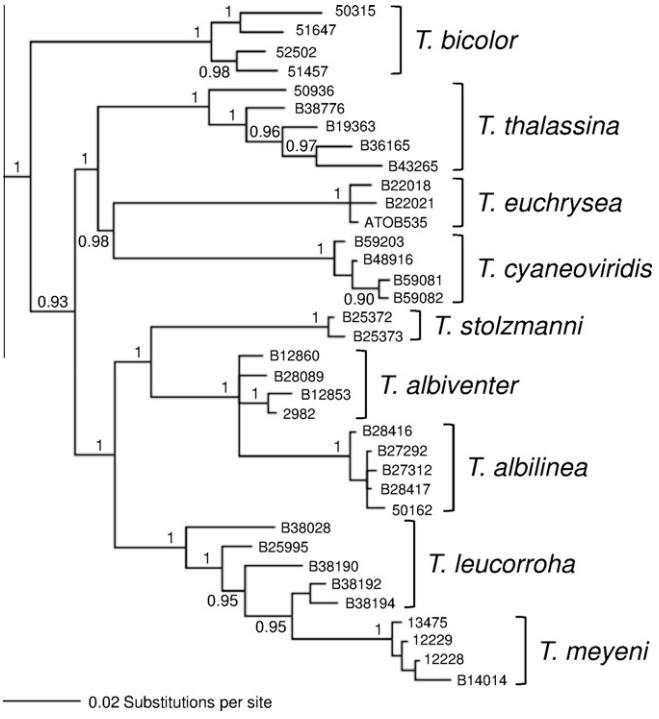

Fig. 1. Phylogenetic relationship of the nine species of *Tachycineta* based on concatenated sequences of 16 nDNA loci analyzed using MrBayes. Outgroups for these analysis were *Progne chalybea* and *Stelgidopteryx serripennis* (not shown). Numbers indicate the posterior probability (⩾0.90).

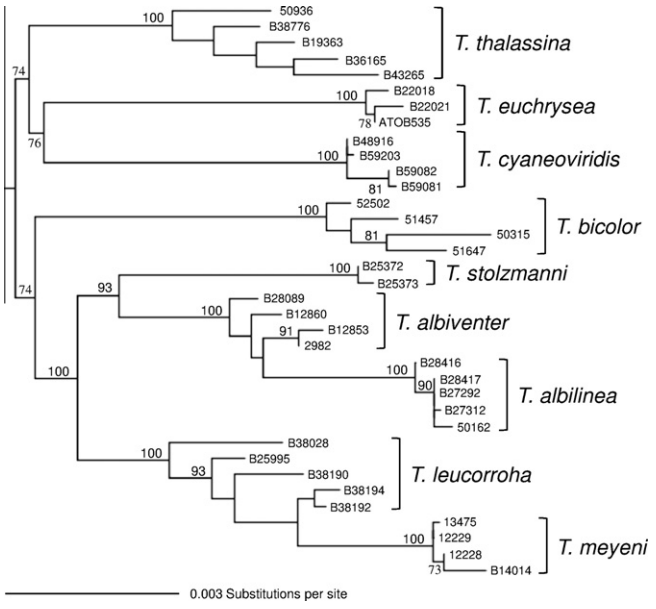

Fig. 2. Phylogenetic relationship of the nine species of *Tachycineta* based on concatenated sequences of 16 nDNA loci using maximum likelihood implemented in RAxML. Outgroups were *Progne chalybea* and *Stelgidopteryx serripenis* (not shown). Numbers indicate ML bootstrap values (⩾70).

Please cite this article in press as: Dor, R., et al. Species trees for the tree swallows (Genus *Tachycineta*): An alternative phylogenetic hypothesis to the mitochondrial gene tree. Mol. Phylogenet. Evol. (2012), http://dx.doi.org/10.1016/j.ympev.2012.06.020

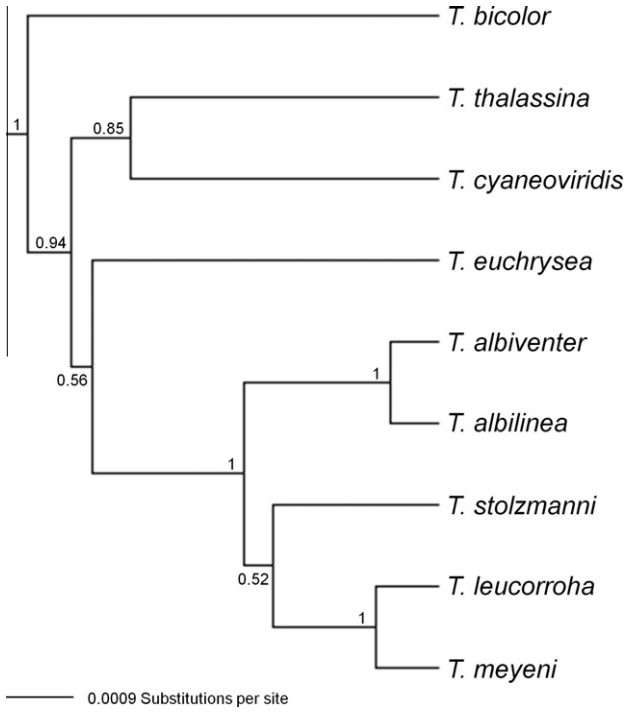

Fig. 3. Consensus species tree of *Tachycineta* generated using *BEAST with 16 locus nDNA matrix. The outgroup included *Progne chalybea* and *Stelgidopteryx serripenis* (not shown). Numbers indicate the posterior probability from the Bayesian analysis.

as sister to *T. thalassina*, and the location of *T. euchrysea* was not resolved.

Visual assessment of the BEST analyses indicated a lack of convergence as the four topologies were quite different (data not shown). The only relationships supported by a posterior probability of 0.4 or greater in the consensus tree (built from the four independent runs after discarding the first 4000 of the 6000 saved trees as burnin) were the sister relationship of *T. leucorrhoa* and *T. meyeni* (posterior probability = 0.45) and the sister relationship of *T. albiventer* and *T. albilinea* (posterior probability = 0.53). Both of these pairs were present in all the other analyses. Given this lack of convergence among the BEST runs, we only focus on the results from the *BEAST analysis below.

4. Discussion

Previous phylogenetic inferences of species relationships within *Tachycineta* relied primarily on mitochondrial sequences and therefore provide only a mitochondrial gene tree for the group (Whittingham et al., 2002; Cerasale et al., 2012), or they used a single intron in addition to mtDNA sequences (βfib7; Sheldon et al., 2005). Here we provide an alternative phylogenetic hypothesis, based on a multi-locus nDNA matrix. We analyzed this matrix using traditional concatenated gene approaches (Bayesian and maximum likelihood) as well as more recently developed coalescent-based species tree inference methods. Phylogenetic reconstructions based on concatenated sequences generated very similar genealogies. Both Bayesian and maximum likelihood methods using the concatenated sequence supported a South and Central American clade and yielded the same relationships within this clade (Figs. 1 and 2). Both concatenation methods also supported a clade consisting of *T. thalassina* and the two Caribbean species (*T. euchrysea* and *T. cyaneoviridis*). However, the ML bootstrap support value for this relationship was lower (bootstrap = 74; Fig. 2). The position of *T. stolzmanii*, as sister to the *T. albiventer* and *T. albilinea* clade, was well supported in the concatenated analyses, but was not in the *BEAST analysis. Similarly, the location of *T. thalassina* as sister to *T. euchrysea* and *T. cyaneoviridis* was not well supported in the *BEAST analysis or the ML analysis, but only in the MrBayes analysis. One potential point of conflict between the Bayesian and ML methods was the location of *T. bicolor*, which was sister to all other *Tachycineta* in the Bayesian analysis and sister to the South and Central American clade in the maximum likelihood analysis. However, the location of *T. bicolor* had lower support in the ML analysis (bootstrap value = 74).

Our tree based on the concatenated sequence of 16 nDNA loci and constructed using Bayesian analysis is significantly different from previously published mtDNA trees (Whittingham et al., 2002; Sheldon et al., 2005). Although both nDNA and mtDNA support a South and Central American clade, they conflict regarding the structure and positions of the North American and Caribbean clades. The nDNA phylogeny strongly supports a clade comprising the two Caribbean species that is sister to *T. thalassina* (Fig. 1). The earlier mtDNA study, however, although unable to resolve these relationships, favored *T. thalassina* and *T. euchrysea* as sister species (Whittingham et al., 2002). Interestingly, a newer tree based on the entire mtDNA genome sequence (Cerasale et al., 2012) concurs with our nDNA tree regarding the relationships among *T. thalassina*, *T. euchrysea* and *T. cyaneoviridis*, with *T. thalassina* as sister to *T. euchrysea*/*T. cyaneoviridis*. A more important difference between the mtDNA- and nDNA-based phylogenies concerns the location of *T. bicolor*. In mtDNA trees, *T. bicolor* is always situated in a clade comprising North American and Caribbean species, as sister to *T. thalassina* and the two Caribbean species; in the Bayesian nDNA reconstructions, *T. bicolor* is positioned as sister to all other *Tachycineta*. The nDNA phylogeny probably provides a more reliable estimate of phylogeny than the mtDNA trees because it is built from data representing 16 independent genes.

Previous studies of *Tachycineta* sampled only one or two individuals per species. However, in this study multiple individuals were sampled per species (Table 1), enabling us to investigate more rigorously the relationships between closely allied pairs of species. Among these are the well known Neotropical pairs – *T. leucorrhoa*/*T. meyeni* and *T. albiventer*/*T. albilinea* – which were found not to be reciprocally monophyletic, suggesting incipient speciation and/or gene flow between these taxa. Large sizes of extant and ancestral populations, together with short time intervals between speciation events, increase the susceptibility of lineages to deep coalescence (Pamilo and Nei, 1988; Maddison, 1997). However, gene flow between those populations after divergence can also contribute to this pattern.

Coalescent inference procedures extract information about the species phylogeny from the variability in coalescence times between independent gene genealogies (Edwards, 2009; Liu and Pearl, 2007). Therefore, they have the potential to provide more accurate species tree estimations than traditional concatenation methods (Edwards et al., 2007). We had mixed results using coalescence methods in this study. We were not able to obtain convergence using the coalescent-based BEST model despite extensive trials. BEST uses a two-stage algorithm to infer the species tree (Liu and Pearl, 2007), therefore, it is more computationally intensive and less efficient than *BEAST. However, it is possible that BEST could have reached convergence with better computational abilities and longer runs. On the other hand, the *BEAST analysis (Fig. 3) generated a very similar topology to that of the Bayesian analysis of concatenated sequences (Fig. 1), except that it did not provide strong support for the location of *T. stolzmanni* within the South and Central American clade or for *T. euchrysea* within the genus. This finding is consistent with Leache and Rannala (2011), which reports that under a range of conditions (longer tree length and smaller population sizes) both Bayesian analyses of

Please cite this article in press as: Dor, R., et al. Species trees for the tree swallows (Genus *Tachycineta*): An alternative phylogenetic hypothesis to the mitochondrial gene tree. Mol. Phylogenet. Evol. (2012), http://dx.doi.org/10.1016/j.ympev.2012.06.020

concatenated sequences and Bayesian species tree inference using the multi-species coalescent (BEST) generate similar results. The generally lower support and resolution of the coalescent-based tree compared to the concatenated sequence tree most likely reflects the additional uncertainty from the multi-species coalescent that is not accounted for by the concatenation method (Liu and Pearl, 2007; Edwards et al., 2007). In addition, under more challenging speciation histories (reduced tree length and increasing the probability of deep coalescence), Bayesian species tree inference outperforms other analyses methods (Leache and Rannala, 2011). Because of this, the coalescent tree probably provides a more conservative, and thus reliable, estimate of the species tree.

The Bayesian nDNA phylogenetic hypothesis provides an alternative to the previously published mtDNA trees. Despite a certain level of uncertainty that is reflected by the coalescent-based approach, this multi-locus tree is the broadest estimate so far for the species tree of the genus *Tachycineta*, providing an historical framework in which to explore trait evolution and comparative analyses in this group. Moreover, it provides an independent framework to examine the evolution of the mitochondrial genome, such as selection operating on metabolism genes among species that inhabit different altitudes or latitudes. Extensive genome sampling is still needed to resolve the phylogenetic relationships in this group fully.

Acknowledgments

Genetic samples were kindly provided by: American Museum of Natural History, Cornell University Museum of Vertebrates, Louisiana State University Museum of Natural Science, and through a gift from A.K. Townsend (Cornell University, USA). For laboratory assistance we thank Amanda Talaba and Laura Stenzler and two anonymous reviewers for their comments. Part of this work was carried out using the resources of the Computational Biology Service Unit at Cornell University that is partially funded by the Microsoft Corporation. This research was supported by the Cornell Lab of Ornithology and NSF Grant OISE-0730180 to DWW.

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.2012.06.020.

References

Backström, N., Fagerberg, S., Ellegren, H., 2008. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. Mol. Ecol. 17, 964–980.

Brooks, D.R., McLennan, D.A., 1991. Phylogeny, Ecology, and Behavior. Univ. Chicago Press, Chicago.

Carling, M.D., Brumfield, R.T., 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. Genetics 178, 363–377.

Cerasale, D.J., Cerasale Dor, R., Winkler, D.W., Lovette, I.J., 2012. Phylogeny of the *Tachycineta* genus of new world swallows: insights from complete mitochondrial genomes. Mol. Phylogenet. Evol. 63, 64–71.

Dickinson, E.C., 2003. The Howard and Moore Complete Checklist of the Birds of the World, third ed. Christopher Helm, London.

Dor, R., Cooper, C.B., Lovette, I.J., Massoni, V., Bulit, F., Liljesthrom, M., Winkler, D.W., 2012. *Clock* gene variation in *Tachycineta* swallows. Ecol. Evol. 2, 95–105.

Drummond, A.J., Rambaut, A., 2007. Beast: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Edwards, S.V., 2009. Is a new and general theory of molecular systematic emerging? Evolution 63, 1–19.

Edwards, S.V., Bensch, S., 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2009. Mol. Ecol. 18, 2930–2933.

Edwards, S.V., Kingan, S.B., Calkins, J.D., Balakrishnan, C.N., Jennings, W.B., Swanson, W.J., Sorenson, M.D., 2005. Speciation in birds: genes, geography, and sexual selection. Proc. Natl. Acad. Sci. USA 102, 6550–6557.

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.

Ferretti, V., Massoni, V., Bulit, F., Lovette, I.J., Winkler, D.W., 2011. Heterozygosity and fitness benefits of extrapair mate choice in white-rumped swallows (*Tachycineta leucorrhoa*). Behav. Ecol.. http://dx.doi.org/10.1093/beheco/arr103.

Ferretti, V., Winkler, D.W., 2009. Polygyny in the tree swallow *Tachycineta bicolor*: a result of the cost of searching for an unmated male. J. Avian Biol. 40, 289–295.

Harvey, P.H., Pagel, M.D., 1991. The Comparative Method in Evolutionary Biology. Oxford University Press, New York.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Hudson, R.R., Turelli, M., 2003. Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution 57, 182–190.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Leache, A.D., 2010. Species trees for spiny lizards (Genus *Sceloporus*): identifying points of concordance and conflict between nuclear and mitochondrial data. Mol. Phylogenet. Evol. 54, 162–171.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126–137.

Liljesthrom, M., Schiavini, A., Reboreda, J.C., 2009. Chilean swallows (*Tachycineta Meyeni*) adjust the number of feathers added to the nest with time of breeding. Wilson J. Ornithol. 121, 783–788.

Linnen, C.R., Farrell, B.D., 2008. Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hympenoptera: Diprionidae). Syst. Biol. 57, 876–890.

Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56, 504–514.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010, New Orleans, LA, pp. 1–8.

Massoni, V., Bulit, F., Reboreda, J.C., 2007. Breeding biology of the white-rumped swallow *Tachycineta leucorrhoa* in Buenos Aires province, Argentina. Ibis 149, 10–17.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718–726.

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 Batesian phylogenetics. Bioinformatics 24, 581–583.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Peters, J.L., 1960. Family Hirundinidae. In: Mayr, E., Greenway, J.C. (Eds.), Check-list of the Birds of the World, Museum of Comparative Zoology, vol. IX. Cambridge, MA, pp. 80–129.

Primmer, C.R., Borge, T., Lindell, J., Saetre, G.P., 2002. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Mol. Ecol. 11, 603–612.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sheldon, F.H., Whittingham, L.A., Moyle, R.G., Slikas, B., Winkler, D.W., 2005. Phylogeny of swallows (Aves: Hirundinidae) estimated from nuclear and mitochondrial DNA sequences. Mol. Phylogenet. Evol. 35, 254–270.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Tao, N., Bruno, W.J., Abfalterer, W., Moret, B.M.E., Leitner, T., Kuiken, C., 2005. FINDMODEL: A Tool to Select the Best-fit Model of Nucleotide Substitution. <http://hcv.lanl.gov/content/sequence/findmodel/findmodel.html> (accessed 2009 November).

Turner, A., 2004. Family Hirundinidae (swallows and martins). In: Del Hoyo, J., Elliott, A., Christie, D. (Eds.), The Birds of the World, vol. 9. Lynx Edicions, Barcelona, Spain, pp. 602–685.

Turner, A.T., Rose, C., 1989. Swallows and Martins. An Identification Guide and Handbook. Houghton Mifflin, Boston, MA.

Whittingham, L.A., Slikas, B., Winkler, D.W., Sheldon, F.H., 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves : Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. Mol. Phylogenet. Evol. 22, 430–441.

Winkler, D.W., Allen, P.E., 1996. The seasonal decline in tree swallow clutch size: physiological constraint or strategic adjustment? Ecology 77, 922–932.

Winkler, D.W., Hallinger, K.K., Ardia, D.R., Robertson, R.J., Stutchbury, B.J., Cohen, R.R., 2011. Tree swallow (*Tachycineta bicolor*). In: Poole, A. (Ed.), The Birds of North America Online. Cornell Lab of Ornithology, Ithaca.

Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. 17, 2107–2121.