Right around the Amazon: the origin of the circum-Amazonian distribution in Tangara cayana

Authors: Savit, Aaron Zvi, and Bates, John Marshall

Source: Folia Zoologica, 64(3) : 273-283

Published By: Institute of Vertebrate Biology, Czech Academy of Sciences

URL: https://doi.org/10.25225/fozo.v64.i3.a8.2015

Folia Zool. – 64 (3): 273–283 (2015)

Right around the Amazon: the origin of the circum-Amazonian distribution in *Tangara cayana*

Aaron Zvi SAVIT[1,2,3]* and John Marshall BATES[2]

[1] *Instituto para el Desarrollo Sustentable de Ceja de Selva, Universidad Nacional Toribio Rodríguez de Mendoza, Chachapoyas, Amazonas, Perú; e-mail: azsavit@gmail.com*
[2] *Bird Division, Field Museum of Natural History, Chicago, Illinois, U.S.A.; e-mail: jbates@fieldmuseum.org*
[3] *Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois, U.S.A.*

Received 6 December 2014; Accepted 16 September 2015

Abstract. The effects of climate fluctuations on seasonally dry forests and their fauna are important to a holistic understanding of diversification in the South American lowlands. We document the intraspecific genetic structure of the burnished-buff tanager (*Tangara cayana*), a species common in seasonally dry tropical forests throughout South America. Using both mitochondrial sequence and nuclear microsatellite markers, we present an intraspecific phylogeny, haplotype network, and a STRUCTURE analysis. We also develop environmental niche models and project them onto two alternate paleoclimate models of the Last Glacial Maximum and mid-Holocene. Paleoclimate projections indicate a much greater extent and connectivity of suitable *T. cayana* habitat during the Last Glacial Maximum (LGM), decreasing through the mid-Holocene toward the present. Both microsatellite and mtDNA sequence data are consistent with a clockwise route of colonization for the current circum-Amazonian distribution of *T. cayana*. The species likely originated in the Cerrado of Brazil and expanded westward through Bolivia, across the seasonally dry forests at the base of the Andes, and into Guyana and northern Brazil. Northeastern populations then expanded south into coastal Pernambuco, Brazil completing the current ring-like distribution of this species.

Key words: biogeography, ecological niche modeling, paleoclimate, phylogeography, seasonally dry forest

Introduction

The role of lowland habitats in the accumulation of tropical diversity has been the subject of considerable research (Mittelbach et al. 2007, Rull 2011). A great deal of scrutiny has been directed upon Amazonia in particular and the patterns of endemism found therein (e.g. Tuomisto et al. 1995, Wesselingh et al. 2009, Hoorn et al. 2010, Cheng et al. 2013). However, studies from complementary lowland biomes are necessary to develop a complete picture of habitat change in the low-elevation tropics as a whole. The evolutionary history of lowland tropical savannas and seasonally dry forests in the Americas is essential for an integrated understanding of biome dynamics across space and time. We use both microsatellite markers and mitochondrial DNA sequence to document phylogeographic and population-level genetic structure in the burnished-buff tanager (*Tangara cayana* Linnaeus), a species that is ubiquitous and widespread throughout South American savanna and seasonally dry forests. We relate these results to inferred *T. cayana* distributions during the mid-Holocene and Last Glacial Maximum (LGM) using environmental niche modeling.

The biogeography of lowland open biomes in South America is complex, and our understanding of the evolutionary dynamics for this suite of ecosystems is far from complete (Werneck et al. 2011). The open woodlands, savannas, and tropical dry forests of South America are patchily distributed along the periphery of Amazonia (Fig. 1), and natural histories can vary from one area to another (Linares-Palomino et al. 2011). These habitats include seasonally dry forest nuclei (e.g. Caatinga), tropical savannas and woodlands (e.g. Cerrado, llanos Pampas del Heath, Pantanal), and several inter-montane valleys along the Andes in western South America (e.g. Huallaga and Mayo River valleys in Peru). The complexity of the various open lowland biomes in the Tropical Americas reflects their disparate evolutionary histories. Some authors have posited that the circum-Amazonian distribution is the result of a "dry forest refugia" scenario (Prado & Gibbs 1993, Pennington et al. 2000), in which formerly extensive tracts of dry forest have contracted

* *Corresponding Author*

toward the present. An alternate scenario is that these habitats have always been fragmentary and isolated, with floristic and faunistic similarities resulting from dispersal (Mayle 2004, 2006).

We study the intraspecific genetic structure of a species with a typical "circum-Amazonian" distribution, *T. cayana*. The patchy nature of *T. cayana's* distribution is ideal for examining the competing hypotheses of vicariant refugia versus long-distance dispersal. Shared haplotypes among populations may indicate that gene flow occurred during the recent past. Conversely, populations which have coalesced may be consistent with relatively rare colonization events via long-distance dispersal with no subsequent gene flow.

The implications of these competing hypotheses for the historical biogeography of seasonally dry tropical forests are profound. Evidence in support of vicariant refugia implies large-scale expansion of suitable *T. cayana* habitat over much of South America. Alternately, support for long-distance dispersal may indicate that tropical dry forest centers are quite stable, undergoing relatively modest episodes of expansion and contraction.

T. cayana is a member of the largest genus of passerine birds in South America. Species of the genus occur in nearly all forested biomes of Central America and northern South America from the Atlantic and Amazonian forests in the lowlands up to about 3500 m in the cloud forests of the Andes. The systematics and phylogeny of the *Tangara* have been studied in detail (Isler & Isler 1999, Burns & Naoki 2004, Sedano & Burns 2010, Burns et al. 2014). *T. cayana* is a member of a clade of seven allospecies spread throughout continental South America, Central America, and the Lesser Antilles. The monophyly of the clade has been strongly supported in molecular studies (Sedano & Burns 2010, Burns et al. 2014). Of the seven, *T. cayana* has by far the broadest geographic distribution despite being one of the younger members of the clade (Burns & Naoki 2004). Other members of this clade include rufous-winged tanager, *Tangara lavinia* (Central America and the Chocó), black-backed tanager, *Tangara peruviana* (southern Atlantic Forest), Lesser Antillean tanager, *Tangara cucullata* (Lesser Antilles), chestnut-backed tanager, *Tangara preciosa* (southern Atlantic Forest), scrub tanager *Tangara vitriolina* (intermontane valleys of Colombia) and the green-capped tanager, *Tangara meyerdeschauenseei* (range restricted in central Andes of Peru). This species group is ecologically distinct from other *Tangara* species with a preference for open, mostly drier habitats (Isler & Isler 1999). The majority of the clade tends to have relatively restricted distributions, with *T. preciosa*, *T. peruviana*, *T. cucullata*, and *T. meyerdeschauenseei* all being examples of tanagers with particularly narrow geographical ranges. Within this group, *T. cayana* is the notable exception, occurring in the tropical Savannas that are widespread both to the north ("llanos") and south ("cerrado") of the Amazon basin. In addition, *T. cayana* is patchily distributed to the west of the Amazon in drier patches at the base of the Andes such as the Mayo River valley and the Pampas del Heath as well as in isolated patches of savanna habitat within Amazonia. Together, this range forms a classic "circum-Amazonian" distribution (Remsen et al. 1991, Bates 1997).

The origin of the circum-Amazonian distribution has been the subject of some speculation. Although long-distance dispersal cannot be ruled out, the patchiness of the present *T. cayana* distribution to the west of the Amazon suggests that its range may have been more extensive in the past (Mayle 2006). Populations in areas that are now isolated by distances of thousands of kilometers were likely established at a time when suitable habitat for *T. cayana* was more contiguous.

T. cayana has seven described subspecies and considerable phenotypic variation. However, differences are frequently clinal despite *T. cayana's* disjunct distribution, with the most divergent phenotypes found in the northern and southern extreme of the species' range (Isler & Isler 1999).

Material and Methods

To model the current and historical distributions of *T. cayana*, we used the program MaxEnt (Phillips & Dudík 2008), which produces a map of continuous habitat suitability scores. To model distributions, we gathered locality data from museums, accessed via the GBIF portal. Duplicate occurrence records were trimmed using the program ENMTools (Warren et al. 2010), as were any occurrences that fell over 100 km outside of the known documented distribution of *T. cayana*. In addition, we used the ArcGIS Toolkit SDMtoolbox (v1.1c, Brown 2014) to rarify spatially clustered observations within a 10 km radius which have the potential to bias model development. Coordinates that fell within the same pixel in the Worldclim raster were likewise trimmed for a total of 434 unique localities. Localities of tissues used in subsequent genetic studies were not included, because many of them lacked the precise georeferences characteristic of the MaxEnt dataset. To represent climatic conditions, we used the Worldclim climatic

layers (Hijmans et al. 2005) at 2.5 arc-minute resolution and cropped to 15° N, 60° S, 90° W and 30° W to encompass all of South America. To avoid overfitting of the environmental niche models, highly correlated Worldclim layers were eliminated. In total, six layers were selected based on uniqueness and relative importance in model development. The aforementioned six Worldclim layers are detailed in Table 1. Environmental niche models were projected onto two alternate climate models at the LGM ~21kybp, and at the mid-Holocene, ~6kybp. These alternate paleoclimate reconstructions are based upon the general circulation models MIROC-ESM and CCSM4 (Otto-Bliesner et al. 2006). In estimating the environmental niche model, twenty replicates were developed and their performance was assessed via cross-validation. We used the overprediction correction function in SDMtoolbox (v1.1c, Brown 2014) to crop the layers to a polygon corresponding to the geographic limits of *T. cayana* occurrences used in the model development, with a 100 km buffer zone.

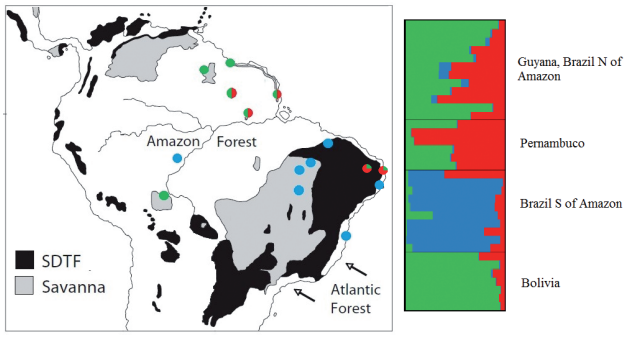

Fig. 1. Geographic localities of individuals included in this study, along with the distribution of the Seasonally Dry Tropical Forest (SDTF) and Savanna biomes (modified from Werneck et al. 2011) that largely define the distribution of *Tangara cayana*. Colors of the sampling localities correspond to the approximate microsatellite group assignment of those populations as determined by STRUCTURE analysis. To the right is the STRUCTURE analysis of *T. cayana* microsatellites indicating approximate individual genotype assignment to three microsatellite groups. Inividuals belonging to the red microsatellite group were recovered in two different localities in Pernambuco Brazil along with admixed individuals. The same microsatellite group is also found admixed in populations north of the River Amazon.

Table 1. BIOCLIM layers used in development of the ENM for *Tangara cayana,* and their respective definitions.

BIOCLIM variable	Definition
BIO 2	Mean diurnal range, mean of monthly (max temp-min temp)
BIO 4	Temperature seasonality (standard deviation *100)
BIO 5	Max temperature of warmest month
BIO 7	Temperature annual range (BIO5-BIO6)
BIO 8	Mean temperature of wettest quarter
BIO 12	Annual precipitation

Geographic sampling

We acquired tissues through grants from museum collections for a total of 37 individual *T. cayana* from across the breadth of its geographic range (see Fig. 1 for sampling locations). In assembling these tissues, we focused on not only the broadest geographic sample possible, but also requested a complete sampling of all individuals collected at any particular site (Table 2). This allows for cursory analysis of population-level demography, and better comparison of geographic variation to local heterogeneity.

Genetic data

Extraction of whole DNA from tissue samples was performed using a Qiagen DNeasy kit (Qiagen U.S.A., Valencia, CA) using the standard protocol as recommended by the manufacturer for animal tissue. For analysis of phylogeographic and demographic patterns of intraspecific divergence, we gathered data from both nuclear (microsatellite) and mitochondrial sources. The mitochondrial data gathered consisted of cytochrome *b* and the cytochrome oxidase three (*CO3*) genes sequence (maximum coverage of 1020 and 715, respectively). Amplification via PCR was performed using the following protocol for both genes: after an initial denaturation step at 96 °C for five minutes, 40 cycles were performed. Each cycle consisted of a denaturation step at 96 °C for 30 seconds, an annealing temperature of 50 °C for 30 seconds, an extension temperature of 72 °C for one minute. At the end of the 35 cycles, the PCR product was held at 4 °C indefinitely. Amplification protocol was nearly identical for microsatellite loci, with the only exception that there were 50 cycles instead of 40, and annealing temperature was gradually ramped up from 50 °C to 65 °C over the course of the 50 cycles. PCR product was cycle sequenced using a BigDye Terminator 3.1 cycle sequencing kit (Catalog #4337455 Life Technologies, Grand Island, NY), cycle sequenced for 35 repetitions at an annealing

Table 2. Sampling localities for *Tangara cayana* tissues analyzed in this study. Abbreviations for tissue sources are as follows: FMNH = Field Museum of Natural History, MPEG = Museu Paraense Emílio Goeldi, LSUMNS = Louisiana State University Museum of Natural Science, ANSP = Philadelphia Academy of Natural Science.

Country	Locality	Province	Source	ID Number
Guyana	10 km W Georgetown		ANSP	21006
Brazil	Manicoré 08o28′20.8″ S, 61o23′35.7″ W	Amazonas	MPEG	57797
Brazil	Óbidos, ESEC Grão-Pará (00o37′50″ N, 55o43′40″ W)	Pará	MPEG	66771
Brazil	Óbidos, ESEC Grão-Pará (00o37′50″ N, 55o43′40″ W)	Pará	MPEG	66772
Brazil	Óbidos, ESEC Grão-Pará (00o37′50″ N, 55o43′40″ W)	Pará	MPEG	66773
Brazil	Óbidos, ESEC Grão-Pará (00o37′50″ N, 55o43v40″ W)	Pará	MPEG	66774
Brazil	Curimatá, Serra Vermelha (9o47′272″ S, 44o28′856″ W)	Piauí	MPEG	68767
Brazil	Guadalupe, Fazenda Maharishi, Praia Indianos (6o44′47.2″ S, 43o50′21.8″ W)	Piauí	MPEG	68769
Brazil	Barreiros, Engenho Cachoeira Linda	Pernambuco	MPEG	70537
Brazil	Ilhéus, Ecoparque de UNA	Bahia	MPEG	70806
Brazil	Ilhéus, Ecoparque de UNA	Bahia	MPEG	70807
Brazil	Tartarugalzinho, Fazenda Casimiro	Amapa	FMNH	391627
Brazil	Tartarugalzinho, Fazenda Casimiro	Amapa	FMNH	391629
Brazil	Tartarugalzinho, Fazenda Casimiro	Amapa	FMNH	391630
Brazil	Tartarugalzinho, Fazenda Casimiro	Amapa	FMNH	391631
Brazil	Tartarugalzinho, Fazenda Casimiro	Amapa	FMNH	391632
Brazil	Timbauba	Pernambuco	FMNH	392387
Brazil	Timbauba	Pernambuco	FMNH	392388
Brazil	Monte Alegre, PA-423, 4 km	Pará	FMNH	392632
Brazil	Monte Alegre, PA-423, 4 km	Pará	FMNH	392633
Brazil	Monte Alegre, Campo do Desterro	Pará	FMNH	392634
Brazil	Taquaritinga	Pernambuco	FMNH	427283
Brazil	Taquaritinga	Pernambuco	FMNH	427284
Brazil	Taquaritinga	Pernambuco	FMNH	427285
Brazil	Taquaritinga	Pernambuco	FMNH	427286
Bolivia	Serrania de Huanchaca, 45 km E Florida	Santa Cruz	LSUMNS	B-13907
Bolivia	Serrania de Huanchaca, 21 km SE Catarata Arco Iris	Santa Cruz	LSUMNS	B-14436
Bolivia	Serrania de Huanchaca, 25 km SE Catarata Arco Iris	Santa Cruz	LSUMNS	B-14840
Bolivia	Serrania de Huanchaca, 21 km SE Catarata Arco Iris	Santa Cruz	LSUMNS	B-14853
Bolivia	Velasco; Parque Nacional Noel Kempff Mercado, 30 km E Aserradero Moira	Santa Cruz	LSUMNS	B-15319
Bolivia	Serrania de Huanchaca, 45 km E Florida	Santa Cruz	LSUMNS	B-15357
Bolivia	Serrania de Huanchaca, 45 km E Florida	Santa Cruz	LSUMNS	B-15414
Guyana	West Demerara District, Polder ca 4 km W the River Demerara on Canal #2 Road		LSUMNS	B-48280
Guyana	West Demerara District, Polder ca 4 km W the River Demerara on Canal #2 Road		LSUMNS	B-48281
Guyana	Region 9; the River Ireng, Karasabai; 03°53′ N, 59°35′ W		LSUMNS	B-48582
Brazil	PI; Mun. José de Freitas, Eco Resort Nazareth	Piauí	MPEG	LZUFPI0050
Brazil	PI; Uruçuí, Vale do Rio Pratinha	Piauí	MPEG	LZUFPI0972

temperature of 50 °C. Samples were sequenced on a 3730xl DNA Analyzer (Applied Biosystems, Carlsbad, CA). Sequences were corrected and edited manually in Sequencher (version 5.1, Gene Codes Corporation, Ann Arbor, MI), and Mega v5.1 (Tamura et al. 2011).

The microsatellite data set consists of 10 loci, the primers for which were developed by Corrêa et al.

(2010) for the thraupid species *Neothraupis fasciata*, and which were shown to successfully amplify across a wide suite of thraupid taxa. All of these loci were polymorphic within and/or between individuals. The use of microsatellite primers designed for an outgroup avoids any potential for development bias toward a particular *Tangara* species. This same lack of development bias may also result in the use of microsatellite markers that are less variable than those that could be developed for a particular species. However, given the time scale that separates many populations, highly variable markers would be of limited utility and potentially confounding. Microsatellites were amplified via PCR using fluorescent-tagged nucleic acids (Schuelke 2000) using the protocol described above. Fragment length was measured using a 3730xl DNA Analyzer (Applied Biosystems, Carlsbad, CA), and results were processed using Genepop software (version 1.2, Raymond & Rousset 1995).

Microsatellite analysis

We performed a STRUCTURE analysis on the microsatellite data set (version 2.3, Pritchard et al. 2000, Hubisz et al. 2009). This algorithm has the advantage that population membership is not specified *a priori*. Rather, probability of group membership is assigned *a posteriori*. The first 100000 iterations of each MCMC chain were discarded as burn-in, and an additional 1000000 steps were logged thereafter. For each potential number of microsatellite groups (k) from one through ten, we performed ten replicates. Output from STRUCTURE analyses was processed using STRUCTURE HARVESTER (version 0.6.93, Earl & vonHoldt 2012), and further analyzed using CLUMPP (version 1.1.2, Jakobsson & Rosenberg 2007). We utilized the Evanno method as implemented in STRUCTURE HARVESTER to identify the number of microsatellite groups with the highest likelihood. Graphs of the results of the STRUCTURE analysis were produced using the program Distruct (Rosenberg 2004).

Sequence analyses

For analysis, sequences from the two mitochondrial genes were concatenated, since they are non-recombinant and evolve as a unit. We estimated ultrametric phylogenies in BEAST using the HKY substitution model a relaxed clock (Drummond et al. 2006), with 10 million iterations of the MCMC chain and the first 10 % of trees discarded as burn-in. Results of the BEAST analyses were processed in TreeAnnotator (version 1.7, Drummond et al. 2012) to obtain a consensus tree and posterior probabilities for the nodes. To root the tree, we used data from *Tangara chilensis*, as sequences for the *CO3* gene were not available from closer relatives to *T. cayana*. To estimate divergence times we used the ultrametric phylogeny and assumed a substitution rate of 2 % per million years (Lovette 2004a, Päckert et al. 2007). We used the program Tracer (version 1.7, Rambaut et al. 2014) for parameter analysis to ensure an effective sample size of > 200 for all parameters.

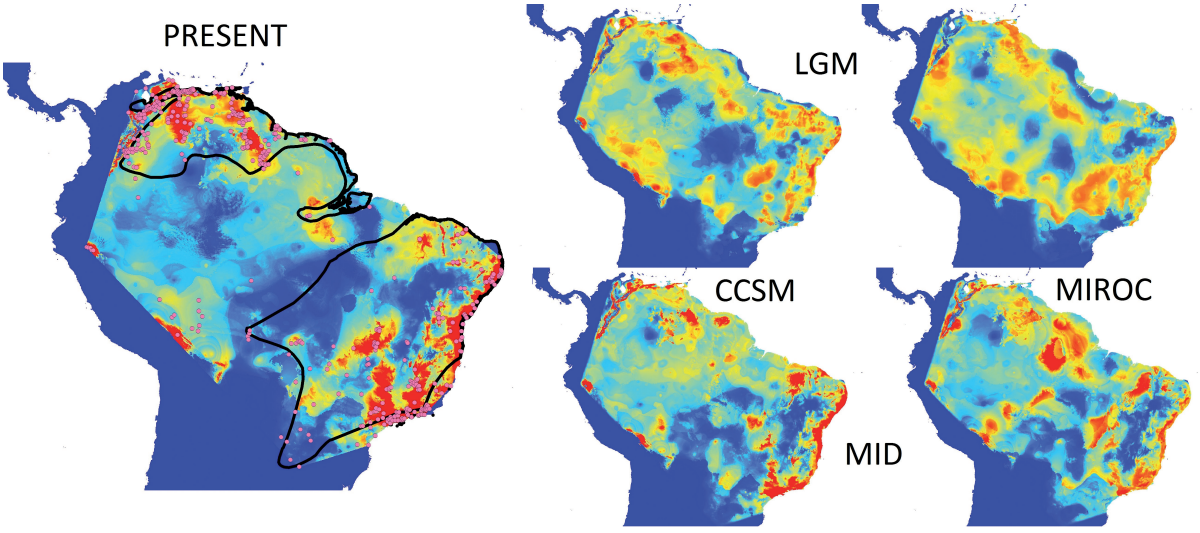

Fig. 2. Environmental niche model for *Tangara cayana* projected onto present day bioclimatic variables (left), as well as two alternate models of paleoclimate at the Last Glacial Maximum (CCSM, center and MIROC, right). Niche models are projected onto paleoclimate layers at both Last Glacial Maximum (~22 kybp, top) and mid-Holocene (~6 kybp, bottom). Warmer colors indicate a greater degree of habitat suitability. Black outline on the present-day ENM is a published range map for this species (Birdlife International & Nature serve 2012). Note that this range map omits populations to the west of the Amazon in Peru and Bolivia which are documented elsewhere (e.g. Isler & Isler 1999). Pink dots represent *T. cayana* localities used in model development.

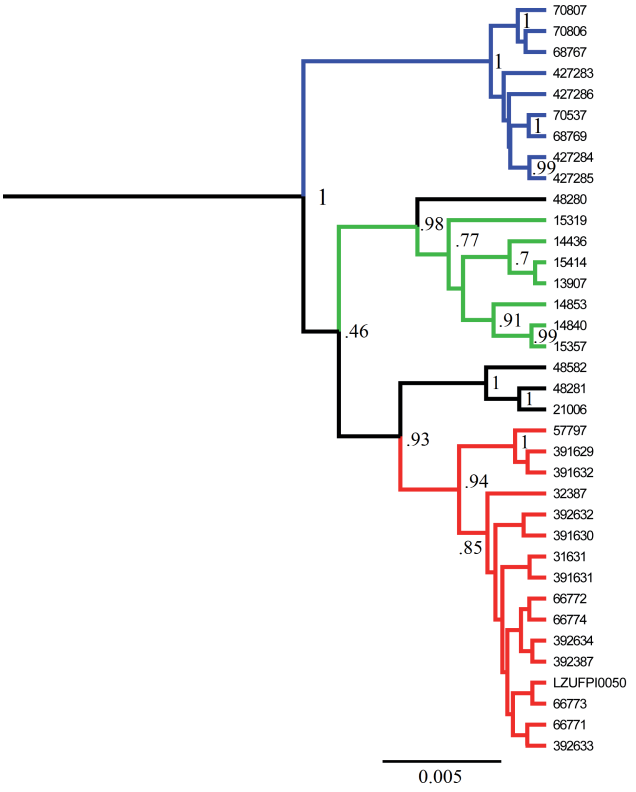

Fig. 3. Intraspecific phylogeny for *Tangara cayana*. Asterisks (*) indicate nodes with posterior probability > 0.98. Clade colors refer to geographic origin as follows: blue = southeast Brazil, green = Bolivia, black = Guyana and red = northeast Brazil. Scale bar indicates 0.5 % uncorrected sequence divergence.

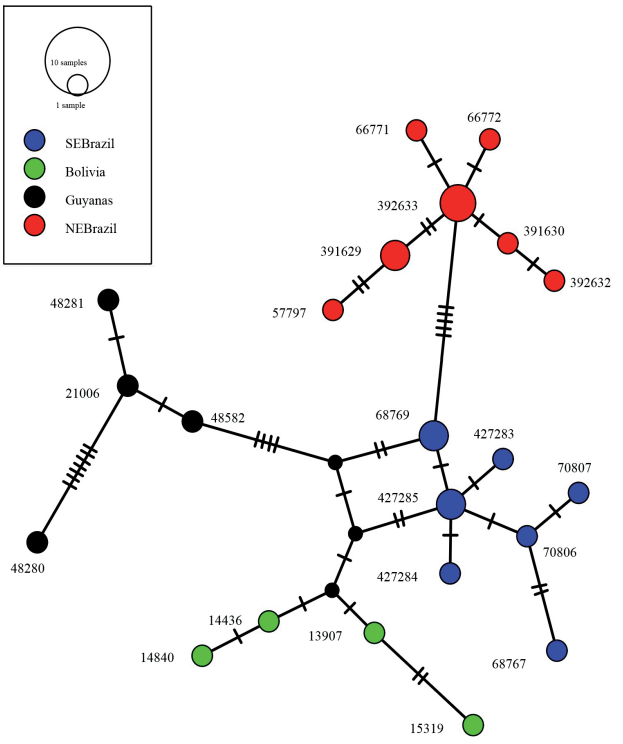

Fig. 4. Median Joining haplotype network for *Tangara cayana* (Bandelt et al. 1999). Size of circles indicates the number of individuals found with that haplotype. Small black circles represent hypothetical haplotypes. Hash marks are equivalent to one base mutation. Coors correspond to sampling location and geographic group as in Fig. 3. Numbers refer to accession IDs listed in Table 2.

We constructed a Median-Joining haplotype network (Bandelt et al. 1999) using the program PopART (Leigh 2015). The same program was also used to produce the graphical representation of the haplotype network.

Results

Environmental niche modelling and paleoclimate projection

The Area Under the Curve (AUC, a measure of model performance) of the replicate MaxEnt runs was 0.868 with a standard deviation of 0.018. The species distribution model for the present is consistent with published accounts of the species range (Fig. 2). Both paleoclimate models indicate more suitable habitats for *T. cayana* at mid-Holocene and LGM, and agree broadly on several points. Both MIROC and CCSM show increased suitable habitats along the base of the Andes particularly in southwest Amazonia. This habitat suitability was highest in both models at LGM and in both cases decreased toward the present, though CCSM model shows a pattern much closer to present-day values by the mid-Holocene. Both models also indicate more suitable *T. cayana* habitat during LGM across much of the Guyanas, Rondonia, and northeast Amazonia. Both paleoclimate models indicate large, contiguous areas of suitable habitat in western Amazonia toward the LGM. While both models show good habitat suitability throughout the Atlantic forest region, there are also large swathes of adjacent habitat in southeastern Amazonia that is unsuitable for *T. cayana* from the present through the LGM. The distribution of suitable habitat in these areas varies between the two paleoclimate reconstructions in these areas. Both paleoclimate models also indicate varying degrees of low habitat suitability in southeastern Amazonia during the LGM.

Microsatellite analysis

Analysis of STRUCTURE output using Evanno et al.'s (2005) K statistic found strong support for three microsatellite groups (K = 3, Fig. 2). These groups include; one covering most of central Amazonia and Brazil south of the Caatinga; another somewhat admixed and centered around the Caatinga and associated Atlantic forest; and a third extending from Bolivia all the way north to the Guyanas and N Brazil. The microsatellite group centered in the

eastern Atlantic forest is admixed to a greater degree with populations to the north on the other side of the Amazon delta. This admixture is in contrast to the population in Bolivia which does not appear to have mixed at all with neighboring Amazonian populations, despite having a sample size (n = 7) comparable to that of the eastern Atlantic forest (n = 6) instead, the southwestern (Bolivian) populations appear to belong to the microsatellite groups found admixed in the northern Brazilian/Guyanan populations ("green" microsatellite group, Fig. 1).

Intraspecific mtDNA phylogeny

Several nodes on the resulting phylogeny have high posterior probability. These include a basal split between southeastern Brazilian individuals and the rest of the birds sampled. The clade comprising all Bolivian individuals plus one Guyanan sample also has strong support, as does the clade comprising the remaining Guyana samples (Fig. 3). Considerable levels of diversity and divergence are observed within each of these three clades. Geographic structure was apparent in all three groups, with the only notable exception being the samples from Guyana. Three of the four mitochondrial haplotypes from Guyana were basal to the geographically proximate samples from Northern Brazil. The fourth Guyanan sample grouped basally with the Bolivian samples (Fig. 3).

Haplotype network

The median joining network (Bandelt et al. 1999) for *T. cayana* shows a close affinity to the Bolivian populations, and is more distant from the Guyanan individuals and similarly divergent from populations north of the Amazon (Fig. 4). Perhaps the most interesting result from haplotype network analyses was the positioning of the "Bolivian" haplotype from Guyana as a highly divergent Guyanan form. This is despite the fact that it is assigned to the Bolivian group in Bayesian analysis with a posterior probability of 0.89 (Fig. 3).

Discussion

T. cayana is a comparatively young species for the genus, having diverged from the Lesser Antillean tanager (*T. cucullata)* within the last 800000 years according to a previously published phylogeny (Sedano & Burns 2010). Levels of uncorrected sequence divergence in the intraspecific *T. cayana* phylogeny in this study are as high as 1 %, or 500000 years assuming a rate of 2 % sequence divergence per million years (Lovette 2004a, Päckert et al. 2007). This is in concordance with previously published genus-level phylogenies (Burns & Naoki 2004, Sedano & Burns 2010, Burns et al. 2014).

The most closely related taxa to *T. cayana* are *T. cucullata* and *T. vitriolina*, which are found in the Lesser Antilles and dry intermontane valleys of Colombia, respectively (Burns et al. 2014). However, the basal divergence in the intraspecific *T. cayana* phylogeny presented here subtends the populations from southeastern Brazil including the states of Piauí, Pernambuco, and Bahia (Fig. 3), though it should be noted that this node has very weak support in the Bayesian phylogeny. The next most divergent clade within *T. cayana* corresponds to the populations in Bolivia. The remaining populations, from north of the River Amazon, fall into two groups. Individuals from Guyana are a strongly supported clade which is basal to the rest of the northern populations. The remaining populations east of the Guyanas in northeastern Brazil show little discernable structure. The phylogeographic structure recovered in *T. cayana* suggests an origin in the south. Subsequently there was expansion from Bolivia probably along the base of the Andes to reach Guyana and finally northeastern Brazil. The potential relationship between Guyanan and northern Amazonian populations is observed on the Bayesian phylogeny, but not in the haplotype network analysis. The recovery of a highly divergent "Bolivian" haplotype in the Guyanan population is consistent with a historical connection across western Amazonia, as predicted by ENM projections onto both paleoclimate models. Environmental niche models also support the hypothesis of a colonization route through western Amazonia along the base of the Andes. Both CCSM and MIROC paleoclimate models show greater degrees of habitat suitability and connectivity during the LGM and decreasing through the mid-Holocene toward the present, particularly in western Amazonia (Fig. 2).

Mitochondrial sequence and microsatellite data show generally congruent patterns, although there is evidence for some additional diversity in the microsatellite data. The mitochondrial phylogeny groups the populations from Taquaritinga in Pernambuco state, Brazil, with the basal clade including populations from Bahia and Piauí (corresponding to the blue microsatellite group in Fig. 3). Other populations in Pernambuco state are grouped with a clade that includes populations to the north of the Amazon. In the microsatellite data, these Pernambuco populations are assigned to a different microsatellite group (red in Fig. 1) which predominates in Pernambuco populations, but which

is also found in *T. cayana* north of the Amazon. Thus, both microsatellite and mtDNA sequence data support a relationship between eastern Pernambuco populations and those north of the River Amazon. The microsatellite data further suggest recent or ongoing admixture between northern populations and those in eastern Pernambuco (Fig. 2). While the mitochondrial phylogeny has a strongly supported node that separates Bolivian *T. cayana* from those north of the Amazon, the microsatellite data largely assigns these populations to the same microsatellite group (green in Fig. 2). The Bolivian populations show no evidence of admixture; however, populations east of the Guyanas in northern Brazil show increasing evidence of admixture with eastern Pernambuco populations. Relative to the northern Brazilian birds, the Guyanan samples have a lower degree of admixture with the Caatinga microsatellite group, and their microsatellite profile is comparable to Bolivian *T. cayana* (Fig. 1). An affinity between the Guyanan samples and those from Bolivia is further evidenced by the mitochondrial data, in which a Bolivian haplotype clusters with that population (see Fig. 3). The co-occurrence of two otherwise geographically segregated haplotypes in Guyana suggests it as an area that has had some degree of gene flow from different biogeographic regions during the last several hundred thousand years. However, the Bolivian haplotype recovered from the Guyanas is consistent with introgression from the Bolivian population into the Guyanas and/or possible retention of ancestral polymorphism.

Microsatellite data indicate two distinct groups centered around the Caatinga and other seasonally dry forests in southeastern Brazil (Fig. 2). The geographic ranges of these microsatellite groups closely coincide with areas of endemism that have been documented across a variety of vertebrate taxa (Carnaval et al. 2009). This includes anurans (Carnaval & Bates 2007, Carnaval & Moritz 2008) and birds (Cracraft 1985, defined therein as the "Parana Center"). While there has historically been a paucity of documented endemism among mammals (Mares et al. 1985), recent work has documented considerable differentiation in the tigrina (*Leopardus tigrinus*) between Caatinga populations and those further to the south (Trigo et al. 2013).

Many avian species and species groups have a "circum-Amazonian" distribution (sensu Remsen et al. 1991). This distribution is characterized by populations distributed along the base of the Andes, as well as north and south of the Amazon. The distributions of dispersal-prone volant species suggest at least two different evolutionary scenarios for seasonally dry forest species. Often, these species lack a clear eastern connection on the Atlantic coast, particularly around the Amazon delta. However, there are many species which show a conflicting distribution, with largely continuous ranges to the east and a lack of connectivity along the base of the Andes (e.g. da Silva & Bates 2002). Previous studies examining the evolution of the circum-Amazonian distribution have shown ambiguous and sometimes conflicting patterns. For example, Lovette (2004b) finds a strongly supported relationship between Guyanan and Atlantic forest populations in the *Phaeothlypis* wood-warbler complex, while a similar study of the *Synallaxis ruficapilla* species group showed evidence for an affinity between Caatinga populations and those in the N Andean foothills (Batalha-Filho et al. 2013). The possibility of a connection between Guyanan and Bolivian populations via the "dry forest arc" along the base of the Andes is a pattern that has also been observed in other taxa (e.g. rattlesnakes, Wüster et al. 2005).

Studies of understory taxa with circum-Amazonian distributions have recovered considerable levels of divergence on the order of 1-2 million years (Lovette 2004a, Batalha-Filho et al. 2013), which far exceeds the divergence in *T. cayana* documented here. While many species with circum-Amazonian distributions also show considerable morphological variation (Bates 1997), this variation may not necessarily coincide with genetic breaks between populations (Irwin 2002, Lovette 2004b). Previous studies have established that ecology is directly related to dispersal ability (e.g. Terborgh 1974, Faaborg 1979, Lewis et al. 2004, Moore et al. 2008), and that this in turn is a major predictor for the amount of differentiation between populations or within a taxon (Burney 2009, Smith et al. 2014).

Foraging stratum has also been shown to have an important effect on lineage diversification (Smith et al. 2014). *Tangara* in general are canopy species, and canopy dwellers tend to have lower levels of intraspecific differentiation relative to understory species (Burney & Brumfield 2009). *T. cayana* inhabits fairly open habitats with patchy tree cover, and thus may be even more prone than many other *Tangara* species to disperse over fairly long distances. Bonaccorso et al. (2006) presented evidence from environmental niche modeling which suggests that species inhabiting drier habitats around the periphery of Amazonia may have had varied responses to climate fluctuations. As such, the timing and route

of *T. cayana's* spread across its circum-Amazonian distribution may not be applicable to many other circum-Amazonian taxa. These disparities among particular evolutionary histories suggests that current circum-Amazonian distributions may have come to be established several different ways.

Conclusions

The evolution and diversification of *T. cayana* across the breadth of its geographic range informs our understanding of seasonally dry forest and forested savannas across time and space. The mtDNA data for *T. cayana* illustrate a pattern of dispersal and diversification involving a likely origin in southeastern South America, with subsequent colonization of the circum-Amazonian dry forests in a clockwise fashion; beginning in the south, expanding west and proceeding northward along the base of the Andes toward the Guyanas, finally crossing into northeast Brazil. The presence of a southern (Bolivian) haplotype in the Guyanan population may be indicative of historical connectivity across the dry forest arc distributed along the base of the Andes. The position of the Guyanan samples at the base of the Bolivian clade indicates that this haplotype may reflect retention of ancestral diversity and is not necessarily a result of recent introgression or gene flow. The Guyanan populations of *T. cayana* are crucial to understanding the evolutionary history of the species and warrant further study. There is also evidence in the microsatellite data for a relatively recent introgression of northern Brazilian populations into eastern Pernambuco to the south of the Amazon. Data from mtDNA sequence and microsatellites are in broad agreement, both supporting discrete groups found in southeast Brazil and an affinity between Bolivian populations and those to the north of the Amazon. Microsatellite data indicate additional diversity in the Caatinga populations in Pernambuco. This diversity is congruent with an area of endemism based on differentiation within a wide variety of vertebrate taxa. Microsatellites do not show a clear distinction between Bolivian populations and those to the north of Amazonia. However, populations to the north are more highly admixed with the "Caatinga" microsatellite group moving further east.

While the available data suggest that this species successfully colonized patches of suitable habitat via a western route, better sampling of the highly isolated *T. cayana* populations in the Andean dry forest arc is needed. In particular, the inclusion of samples from the two isolated populations in Peru (Pampas del Heath and the Mayo Valley) would help to clarify the timing and precise route by which the Andean dry forest arc was colonized by *T. cayana*.

Acknowledgements

We would like to thank the following people and institutions for providing tissue grants for use in this study: D. Willard and S. Hackett, Field Museum of Natural History; A. Aleixo, Museu Paraense Emílio Goeldi; D. Dittman and R. Brumfeld, Louisiana State University Museum of Natural Science; P. Sweet, Philadelphia Academy of Natural Science. Three anonymous reviewers provided helpful comments on an earlier version of this manuscript. This research was partially funded through a collaborative grant, Dimensions US-Biota-São Paulo (#1241066) supported by the US National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Literature

Bandelt H., Forster P. & Röhl A. 1999: Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol. 16: 37–48.*

Batalha-Filho H., Irestedt M., Fjeldså J., Ericson P.G.P., Silveira L.F. & Miyaki C.Y. 2013: Molecular systematics and evolution of the *Synallaxis ruficapilla* complex (Aves: Furnariidae) in the Atlantic forest. *Mol. Phylogenet. Evol. 67: 86–94.*

Bates J.M. 1997: Distribution and geographic variation in three South American grassquits (Emberizinae, *Tiaris*). *Ornithol. Monogr. 48: 91–110.*

BirdLife International & Nature Serve 2012: Bird species distribution maps of the world, version 2.0. *BirdLife International, Cambridge, U.K. and Nature Serve, Arlington, U.S.A.*

Bonaccorso E., Koch I. & Peterson A.T. 2006: Pleistocene fragmentation of Amazon species' ranges. *Divers. Distrib. 12: 157–164.*

Brown J.L. 2014: SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol. Evol. 5: 694–700.*

Burney C.W. 2009: Comparative phylogeography of Neotropical birds. *PhD Thesis, Louisiana State University, Baton Rouge, Louisiana.*

Burney C.W. & Brumfield R.T. 2009: Ecology predicts levels of genetic differentiation in Neotropical birds. *Am. Nat. 174: 358–368.*

Burns K.J. & Naoki K. 2004: Molecular phylogenetics and biogeography of Neotropical tanagers in the genus *Tangara*. *Mol. Phylogenet. Evol. 32: 838–854.*

Burns K.J., Shultz A.J., Title P.O., Mason N.A., Barker F.K., Klicka J., Lanyon S.M. & Lovette I.J. 2014: Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. *Mol. Phylogenet. Evol. 75: 41–77.*

Carnaval A.C. & Bates J.M. 2007: Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in Northeastern Brazil. *Evolution 61: 2942–2957.*

Carnaval A.C., Hickerson M.J., Haddad C.F.B., Rodrigues M.T. & Moritz C. 2009: Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science 323: 785–789.*

Carnaval A.C. & Moritz C. 2008: Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *J. Biogeogr. 35: 1187–1201*.
Cheng H., Sinha A., Cruz F.W., Wang X., Edwards R.L., d'Horta F.M., Ribas C.C., Vuille M., Stott L.D. & Auler A.S. 2013: Climate change patterns in Amazonia and biodiversity. *Nat. Commun. 4: 1411*.
Corrêa C.L., Collevatti R.G. & Caparroz R. 2010: Isolation and characterization of microsatellite loci for *Neothraupis fasciata*, (Emberizidae, Passeriformes) with widely cross amplification in Neotropical passerines. *J. Hered. 101: 385–389.*
Cracraft J. 1985: Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. *Ornithol. Monogr. 36 (Suppl.): 49–84.*
da Silva J.M.C. & Bates J.M. 2002: Biogeographic patterns and conservation in the South American Cerrado: a tropical savanna hotspot. *BioScience 52: 225–234.*
Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. 2006: Relaxed phylogenetics and dating with confidence. *PLoS Biol. 4: e88.*
Drummond A.J., Suchard M.A., Xie D. & Rambaut A. 2012: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol. 29: 1969–1973.*
Earl D.A. & vonHoldt B.M. 2012: STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour. 4: 359–361*.
Evanno G., Regnaut S. & Goudet J. 2005: Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol. 14: 2611–2620.*
Faaborg J. 1979: Qualitative patterns of avian extinction on Neotropical land-bridge islands: lessons for conservation. *J. Appl. Ecol. 16: 99–107.*
Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A. 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol. 25: 1965–1978.*
Hoorn C., Wesselingh F.P., ter Steege H., Bermudez M.A., Mora A., Sevink J. et al. 2010: Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science 330: 927–931.*
Hubisz M.J., Falush D., Stephens M. & Pritchard J.K. 2009: Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour. 9: 1322–1332.*
Irwin D.E. 2002: Phylogeographic breaks without geographic barriers to gene flow. *Evolution 56: 2383–2394.*
Isler M.L. & Isler P.R. 1999: Tanagers. *Smithsonian Institution Press, Washington D.C., U.S.A.*
Jakobsson M. & Rosenberg N.A. 2007: CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics 23: 1801–1806.*
Leigh J.W. 2015: PopART: Population Analysis with Reticulate Trees. *Accessed July 4, 2015. http://popart.otago.ac.nz*
Lewis S.L., Phillips O.L., Baker T.R., Lloyd J., Malhi Y., Almeida S. et al. 2004: Concerted changes in tropical forest structure and dynamics: evidence from 50 South American long-term plots. *Philos. Trans. R. Soc. Lond. B 359: 421–436.*
Linares-Palomino R., Oliveira-Filho A.T. & Pennington R.T. 2011: Neotropical seasonally dry forests: diversity, endemism, and biogeography of woody plants. In: Dirzo R., Young H.S., Mooney H.A. & Ceballos G. (eds.), Seasonally dry tropical forests. *Island Press/Center for Resource Economics, Washington D.C., U.S.A.*
Lovette I.J. 2004a: Mitochondrial dating and mixed support for the "2% rule" in birds. *Auk 121: 1–6.*
Lovette I.J. 2004b: Molecular phylogeny and plumage signal evolution in a trans Andean and circum Amazonian avian species complex. *Mol. Phylogenet. Evol. 32: 512–523.*
Mares M.A., Willig M.R. & Lacher T.E. 1985: The Brazilian Caatinga in South American zoogeography: tropical mammals in a dry region. *J. Biogeogr. 12: 57–69.*
Mayle F.E. 2004: Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. *J. Quaternary Sci. 19: 713–720.*
Mayle F.E. 2006: The Late Quaternary biogeographic history of South America seasonally dry tropical forests: insights from palaeoecological data. In: Pennington R.T., Lewis G.P. & Ratter J.A. (eds.), Neotropical savannas and dry forests: plant diversity, biogeography and conservation. *CRC Press, Taylor & Francis, Systematics Association special volume, no. 69: 395–416.*
Mittelbach G.G., Schemske D.W., Cornell H.V., Allen A.P., Brown J.M., Bush M.B. et al. 2007: Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett. 10: 315–331.*
Moore R.P., Robinson W.D., Lovette I.J. & Robinson T.R. 2008: Experimental evidence for extreme dispersal limitation in tropical forest birds. *Ecol. Lett. 11: 960–968.*
Otto-Bliesner B.L., Brady E.C., Clauzet G., Tomas R., Levis S. & Kothavala Z. 2006: Last Glacial Maximum and Holocene climate in CCSM3. *J. Climate 19: 2526–2544.*
Päckert M., Martens J., Tietze D.T., Dietzen C., Wink M. & Kvist L. 2007: Calibration of a molecular clock in tits (Paridae). Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? *Mol. Phylogenet. Evol. 44: 1–14.*
Pennington R.T., Prado D.E. & Pendry C.A. 2000: Neotropical seasonally dry forests and quaternary vegetation changes. *J. Biogeogr. 27: 261–273.*
Phillips S.J. & Dudík M. 2008: Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography 31: 161–175.*
Prado D.E. & Gibbs P.E. 1993: Patterns of species distributions in the dry seasonal forests of South America. *Ann. Mo. Bot. Gard. 80: 902–927.*
Pritchard J.K., Stephens M. & Donnelly P. 2000: Inference of population structure using multilocus genotype data. *Genetics 155: 945–959.*
Rambaut A., Suchard M.A., Xie D. & Drummond A.J. 2014: Computer program and documentation distributed by the author, tracer v1.6. *Accessed 14 August 2014. http://beast.bio.ed.ac.uk/Tracer*

Raymond M. & Rousset F. 1995: GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered. 86: 248–249.*

Remsen J.V., Rocha O., Schmitt C.G. & Schmitt D.C. 1991: Zoogeography and geographic variation of *Platyrinchus mystaceus* in Bolivia and Peru, and the circum-Amazonian distribution pattern. *Ornitol. Neotrop. 2: 77–83.*

Rosenberg N.A. 2004: Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes 4: 137–138.*

Rull V. 2011: Neotropical biodiversity: timing and potential drivers. *Trends Ecol. Evol. 26: 508–513.*

Schuelke M. 2000: An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol. 18: 233–234.*

Sedano R.E. & Burns K.J. 2010: Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). *J. Biogeogr. 37: 325–343.*

Smith B.T., McCormack J.E., Cuervo A.M., Hickerson M.J., Aleixo A., Cadena C.D. et al. 2014: The drivers of tropical speciation. *Nature 515: 406–409.*

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. 2011: MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum Parsimony methods. *Mol. Biol. Evol. 28: 2731–2739.*

Terborgh J. 1974: Preservation of natural diversity: the problem of extinction prone species. *BioScience 24: 715–722.*

Trigo T.C., Schneider A., de Oliveira T.G., Lehugeur L.M., Silveira L., Freitas T.R.O. & Eizirik E. 2013: Molecular data reveal complex hybridization and a cryptic species of Neotropical wild cat. *Curr. Biol. 23: 2528–2533.*

Tuomisto H., Ruokolainen K., Kalliola R., Linna A., Danjoy W. & Rodriguez Z. 1995: Dissecting Amazonian biodiversity. *Science 269: 63–66.*

Warren D.L., Glor R.E. & Turelli M. 2010: ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography 33: 607–611.*

Werneck F.P., Costa G.C., Colli G.R., Prado D.E. & Sites Jr., J.W. 2011: Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. *Glob. Ecol. Biogeogr. 20: 272–288.*

Wesselingh F.P., Hoorn C., Kroonenberg S.B., Antonelli A., Lundberg J.G., Vonhof H.B. & Hooghiemstra H. 2009: On the origin of Amazonian landscapes and biodiversity: a synthesis. In: Hoorn C. & Wesselingh F.P. (eds.), Amazonia: landscape and species evolution. *Wiley-Blackwell Publishing Ltd.*

Wüster W., Ferguson J.E., Quijada-Mascareñas J.A., Pook C.E., Da Graça Salomão M. & Thorpe R.S. 2005: Tracing an invasion: landbridges, refugia, and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: *Crotalus durissus*). *Mol. Ecol. 14: 1095–1108.*