## Title

The Use of Genetics to reconstruct the biogeography of the Callitrichidae, to examine the evolution of the major histocompatibility complex DRB region in Cebus capacious imitator and to phylogenetically place two extinct duck species within Anatidae.



## Author

Buckner, Janet Charray

## Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# The Use of Genetics to reconstruct the biogeography of the Callitrichidae, to examine the evolution of the major histocompatibility complex DRB region in *Cebus capucinus imitator* and to phylogenetically place two extinct duck species within Anatidae

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

Janet Charray Buckner

2017

ABSTRACT OF THE DISSERTATION

The Use of Genetics to

reconstruct the biogeography of the Callitrichidae,

to examine the evolution of the major histocompatibility complex DRB region in *Cebus capucinus imitator*

and to phylogenetically place two extinct duck species within Anatidae

by

Janet Charray Buckner

Doctor of Philosophy in Biology

University of California, Los Angeles, 2017

Professor Michael Edward Alfaro, Co-Chair

Professor David K. Jacobs, Co-Chair

Genetic information is indispensable in understanding several aspects of biology. In the first chapter of this dissertation, I reconstructed a phylogenetic tree from published gene fragments for the marmosets and tamarins (Callitrichidae) that I then used to reconstruct the biogeographic history of the group. Based on the results of this and previous studies, I recommended that the genus *Saguinus* be formally split into two genera, *Saguinus* and *Leontocebus*. In chapter two, I used high-throughput sequencing to genotype two exons of the MHC-DRB region in the sub-species *Cebus capucinus imitator* from Costa Rica. As in other reports on primate taxa, I found this gene region to be highly polymorphic and there is evidence of multiple copies of DRB gene. I also found evidence for positively selected sites in both exons. In chapter three, I use ancient

DNA extraction techniques combined with high-throughput sequencing to recover complete mitochondrial genomes for two extinct species of duck and place them into a phylogeny of living Anatidae. The first species, *Chendytes lawi*, is found to be the sister taxon to the tribe Anatini, despite all previous studies linking it to the tribe Mergini based on morphology. The second species, *Camptorhynchus labradorius*, is confirmed as a member of the tribe Mergini and is further shown to be the sister taxon to *Polysticta stelleri*, Steller's Eider.

The dissertation of Janet Charray Buckner is approved.

Jessica W. Lynch Alfaro

Thomas Bates Smith

Thomas Welch Gillespie

Michael Edward Alfaro, Committee Co-Chair

David K. Jacobs, Committee Co-Chair

University of California, Los Angeles

2017

# DEDICATION

I dedicate my dissertation to my family. To my parents, Ramona Jamel Orange and Ted Arnold Wilson who were always supportive and proud of my achievements. To my two younger sisters, Te'Audrea Roberta Wilson and Alexandrea Joi Wilson, who have always been incredibly enthusiastic about my research and work. To my Nana, Roberta Wilson, who never stopped checking in and always supported my pursuit of education. To my grandparents, Raymond James Buckner and Irma Buckner, who are not here to celebrate this achievement, but always loved and encouraged "Zoo-zoo Chutty." To my partner, Daniel Pineda, who kept me cognizant of the world outside academia through the last four years of my doctoral program. To my friends, especially those who went through the doctoral program with me. Finally, to my pets, who were my loyal companions through it all.

Thank You.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter One is a version of:

Buckner, J. C., Alfaro, J. W. L., Rylands, A. B., & Alfaro, M. E. (2015). Biogeography of the marmosets and tamarins (Callitrichidae). *Molecular phylogenetics and evolution*, *82*, 413-425.

JCB and JWLA conceived and developed the project. JCB performed the data collection, analyses and designed the figures. JCB, JWLA and ABR wrote the manuscript. MEA reviewed the study design and revised the manuscript.

Chapter three is in press for *Molecular Phylogenetics and Evolution*:

Buckner, J. C., Ellingson, R., Gold, D., Jones, T., Jacobs, D. (in press). Mitogenomics supports an unexpected taxonomic relationship for the extinct diving duck *Chendytes lawi* and definitively places the extinct Labrador Duck. *Molecular phylogenetics and evolution*

JCB and DKJ conceived and designed the study. TJ provided samples. JCB collected the data. JCB, RE and DG performed analyses and designed the figures. JCB and DKJ wrote the manuscript. RE, DG, TJ and DJ revised the manuscript.

# CURRICULUM VITAE

## DEGREES

2011 **Bachelor of Science**, Zoology, Cum Laude; **Minor**, Chemistry, SUNY Oswego.

## ACADEMIC & RESEARCH FELLOWSHIPS / GRANTS

**2018-2019 NSF Postdoctoral Fellowship in Biology ($138,000.00)**
**2017 Dissertation Year Fellowship ($40,000.00), UCLA**
**2016 Lida Scott Brown Fellowship ($6500), UCLA**
**2015-2016 Eugene V. Cota-Robles Graduate Fellowship Year 2 ($40,825.71), UCLA**
**2014-2015 Fulbright U.S. Student Program, Brazil ($19,700.00)**
**2014-2015 Genomic Analysis Training Grant, UCLA, (declined)**
**2013-2014 GAANN Graduate Support Fellowship ($20,116.43), UCLA**
**2011-2012 Eugene V. Cota-Robles Graduate Fellowship ($48,681.83), UCLA**

## HONORS AND AWARDS

2017 Society for the Study of Evolution Travel Stipend ($500)
2012, 2017 EEB Departmental Travel Award ($800)
2015, 2016 EEB Departmental Research Awards ($2000)
2014 American Society of Mammalogists Grants in Aid of Research ($1000)
2013 Sigma Xi Grants in Aid of Research ($500)
2013 Charlotte Mangum Student Support Program Travel Award (SICB) ($500)
2012, 2013 NSF GRFP Honorable Mention
2011 Martin V. Torok Award, SUNY Oswego

## PUBLICATIONS

**Buckner JC**, Ellingson RE, Gold DA, Jones TL, Jacobs DK (in press) Mitogenomics supports a novel taxonomic relationship for the extinct diving duck *Chendytes lawi*. *Molecular Phylogenetics and Evolution*

Lima MGM, **Buckner JC**, Sousa e Silva Júnior J, Aleixo A, Martins A, Boubli JP, Link A, Farias IP, Silva MN, Röhe F, Queiroz H, Chiou KL, Di Fiore A, Alfaro ME, Lynch Alfaro JW (2017) Capuchin monkey biogeography: Understanding sympatry between *Cebus* and *Sapajus*. *Journal of Biogeography* doi: 10.1111/jbi.12945

Rylands AB, Heymann EW, Lynch Alfaro J, **Buckner JC**, Roos C, Matauschek C, Boubli JP, Sampaio R, Mittermeier RA. (2016) Taxonomic review of the New World tamarins (Primates: Callitrichidae). *Zoological Journal of the Linnean Society* 117 (4), 1003-1028.

Blumstein D, **Buckner, J.**, Shah, S., Patel, S., Alfaro, M. E., Natterson-Horowitz, B. (2015). A clinical research pathway towards developing new insights into cardiomyopathy. *Evolution, Medicine, and Public Health* 2015, 280.

Blumstein D, **Buckner, J.**, Shah, S., Patel, S., Alfaro, M. E., Natterson-Horowitz, B. (2015). The evolution of capture myopathy in hooved mammals: a model for human stress cardiomyopathy? *Evolution, Medicine, and Public Health* 2015 (1), 195-203.

**Buckner JC**, Lynch Alfaro J, Rylands A, Alfaro ME. (2015) Biogeography of the Marmosets and

Tamarins. *Molecular Phylogenetics and Evolution* 82, 413-425.

**Buckner JC**, Welsh AB, Sime KR. (2014) Evidence for Population Differentiation in the Bog Buckmoth of New York State. *Northeastern Naturalist* 21, 506-514.

## PUBLISHED ABSTRACTS

**Buckner J**, Ellingson R, Gold DA, Jones TL, Jacobs DK. Ancient DNA analysis of the extinct diving ducks *Chendytes lawi* and the Labrador duck reveal convergent evolution of diving in boreal habitats. Evolution Meeting (2017) Oral.

**Buckner J**, Ellingson R, Gold DA, Jacobs DK. Sequencing Dead Ducks - The Labrador Duck is Sister to Steller's Eider and the Subfossil *Chendytes* lawi roots the dabbling duck clade. Society for Integrative and Comparative Biology Meeting (2017) Oral.

**Buckner J**, Ellingson R, Sim J, Jacobs D. Hydrologic mediation of dispersal in Tidewater Gobies in the mid-coast. American Fisheries Society Western Division Meetings, Reno, Nevada. (2016) Oral.

**Buckner JC,** Lynch Alfaro J, Rylands AB, Alfaro M. Biogeography of the Marmosets and Tamarins. Society for Integrative and Comparative Biology Meetings, San Francisco. (2013) Poster.

**J. C. Buckner**, J. Lynch Alfaro, A. Rylands, M. Alfaro. Statistical Phylogeography of the Marmosets and Tamarins. International Primatological Society Meetings, Cancun, Mexico. (2012) Oral.

## TEACHING

UCLA EEB 187 - Variable Topics in Biology - Teaching Assistant
UCLA EEB 110 - Vertebrate Morphology - Teaching Assistant
UCLA EEB C173 - Earth Process and Evolutionary History - Teaching Assistant
UCLA EEB 111 - Biology of Vertebrates - Teaching Associate
UCLA EEB 115 - Mammalogy - Teaching Associate

## PROFESSIONAL DEVELOPMENT

| | |
|---|---|
| 2014 | Phylogenetic Analysis Using RevBayes, NESCent Course (Participant) |
| 2014 | Central African Biodiversity Alliance Professional Development Workshop (Participant) |
| 2012 | Study and Conservation of Primates in Transformed Habitats: Neotropical Region Course (Participant) |
| 2011 | State University of New York Global Laboratories in Brazil: Ecology (Grantee) |
| 2008 -11 | Ronald E. McNair Post-Baccalaureate Achievement Program, 2008-11 (Alumnus) |

## SERVICE

| | |
|---|---|
| 2016-17 | Howard Hughes Medical Institute Mentor, UCLA |
| 2015-16 | EEB Graduate Research Awards Committee |
| 2012-16 | Undergraduate Research Mentor, UCLA |
| 2015 | Latin American Macroevolutionary Analysis Workshop, Teaching Assistant |
| 2014 | UCLA Graduate Representative to SACNAS conference |
| 2013-14 | Life Sciences Diversity Advisory Committee, Graduate Student Rep |
| 2013-14 | Graduate Recruitment Diversity Sub-Committee, Graduate Student Rep |
| 2012-14 | Ecology and Evolutionary Biology Teaching Assistant |

# Chapter 1

# Biogeography of the marmosets and tamarins (Callitrichidae)

**Janet C. Buckner, Jessica W. Lynch Alfaro, Anthony B. Rylands, Michael E. Alfaro**

Molecular Phylogenetics and Evolution 82 (2015) 413–425

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

# Biogeography of the marmosets and tamarins (Callitrichidae)

Janet C. Buckner [a,*], Jessica W. Lynch Alfaro [b,c], Anthony B. Rylands [d], Michael E. Alfaro [a]

[a] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA
[b] Institute for Society and Genetics, University of California, Los Angeles, CA 90095, USA
[c] Department of Anthropology, University of California, Los Angeles, CA 90095, USA
[d] Conservation International, Arlington, VA 22202, USA

ARTICLE INFO

*Article history:*
Received 1 October 2013
Revised 9 April 2014
Accepted 15 April 2014
Available online 20 May 2014

*Keywords:*
Biogeography
Phylogenetics
Neotropics
Platyrrhini
*Leontocebus*
Saguinus

ABSTRACT

The marmosets and tamarins, Family Callitrichidae, are Neotropical primates with over 60 species and subspecies that inhabit much of South America. Although callitrichids exhibit a remarkable widespread distribution, attempts to unravel their biogeographic history have been limited by taxonomic confusion and the lack of an appropriate statistical biogeographic framework. Here, we construct a time-calibrated multi-locus phylogeny from GenBank data and the callitrichid literature for 38 taxa. We use this framework to conduct statistical biogeographic analyses of callitrichids using BioGeoBEARS. The DIVAj model is the best supported reconstruction of biogeographic history among our analyses and suggests that the most recent common ancestor to the callitrichids was widespread across forested regions c. 14 Ma. There is also support for multiple colonizations of the Atlantic forest region from the Amazon basin, first by *Leontopithecus* c. 11 Ma and later by *Callithrix* c. 5 Ma. Our results show support for a 9 million year old split between a small-bodied group and large-bodied group of tamarins. These phylogenetic data, in concert with the consistent difference in body size between the two groups and geographical patterns (small-bodied tamarins and large-bodied tamarins have an unusually high degree of geographic overlap for congeners) lend support to our suggestion to split *Saguinus* into two genera, and we propose the use of distinct generic names; *Leontocebus* and *Saguinus*, respectively.



## 1. Introduction

Marmosets and tamarins (family Callitrichidae[1]), the smallest of the Neotropical primates, are found throughout much of South America. Although callitrichids constitute a widely-distributed family of Neotropical primates, attempts to unravel their biogeographic history have been limited by taxonomic confusion and the lack of an appropriate statistical biogeographic framework. Six callitrichid genera are currently recognized, although debates regarding taxonomic delineations are ongoing (see Van Roosmalen and Van Roosmalen, 2003; Cortés-Ortiz, 2009; Schneider et al., 2012): these genera include marmosets (*Cebuella* – two subspecies of pygmy marmosets, *Mico* – Amazonian marmosets, and *Callithrix* – Atlantic Forest marmosets), lion tamarins (*Leontopithecus*), tamarins (*Saguinus*), and Goeldi's monkey of the monotypic genus *Callimico*. Here, we largely follow the taxonomy of the Callitrichidae as given in Rylands and Mittermeier (2009), but take into account some modifications resulting from more recent publications (Röhe et al., 2009; Ferrari et al., 2010; Matauschek et al., 2011; Schneider et al., 2012; Schneider and Sampaio, 2014).

### 1.1. The diversification of callitrichids

The earliest hypotheses for callitrichid biogeography focused on the fragmentation of an ancestrally widespread species. Hershkovitz (1977) hypothesized that the ancestral taxon differentiated due to isolation caused by forest contraction as a result of flood or drought. Rylands et al. (1996) suggested that the most basal splits occurred between the northwest and southeast forest regions, resulting in a tamarin ancestor in the Amazon and a marmoset/lion tamarin ancestor in the Atlantic Forest with subsequent diversification in the respective areas. A pattern of dispersal by more locally restricted species across the Amazon has been found in other primates (Lynch Alfaro et al., 2012), but this alternative has not been evaluated yet for callitrichids.

* Corresponding author. Address: 612 Charles E. Young Drive East, P.O. Box 957246, Los Angeles, CA 90095-7246, USA.
*E-mail address:* jcharrayb@ucla.edu (J.C. Buckner).

[1] While taxonomists traditionally placed the marmosets and tamarins together as the family Callitrichidae as we do for this paper (see Napier and Napier, 1967; Hershkovitz, 1977; Rylands et al., 2000; Cortés-Ortiz, 2009) others (for example Groves, 2001, 2005) prefer to classify them as the subfamily Callitrichinae within the Cebidae, as proposed by Rosenberger (1981, 2011).

http://dx.doi.org/10.1016/j.ympev.2014.04.031
1055-7903/

### 1.2. Biogeography of Leontopithecus, Callimico and the marmosets

Hershkovitz (1977) suggested that the first split in the callitrichid lineage was of an ancestral *Cebuella* in the western Amazon and an ancestor to marmosets and lion tamarins in the Atlantic forest in eastern Brazil. A population of marmosets later dispersed into the eastern Amazon (to become modern day *Mico*) and eventually gave rise to *Saguinus* following continued dispersal of populations further west into the Amazon (Hershkovitz, 1977). Ferrari (1993a, 1993b) argued for an extra-Amazonian origin of marmosets, with the Amazonian forms invading the Amazon from the south and speciating as they moved north on either side of the Rio Tapajós. *Cebuella*, he suggested, arose from a population that managed to cross the Rio Madeira, east to west, invading forests dominated by tamarins.

The evaluation of biogeographic scenarios for the marmosets has been complicated by uncertainty in the systematic placement of the lion tamarins and Goeldi's monkey. However, recent molecular phylogenetic studies have provided new insights by clarifying the relationships among callitrichid lineages. In placing *Leontopithecus* and then *Callimico* as successive sister groups to the marmosets (reviewed in Schneider and Sampaio, 2015), modern phylogenetic studies reveal strongly discordant distributional patterns across this clade. The disjunct distribution between *Callimico* and *Cebuella* in the western Amazon (west of the Rio Madeira) and the range of *Mico*, *Callithrix* and *Leontopithecus*, in the eastern Amazon (east of the Rio Madeira), the Cerrado, Caatinga and Atlantic forest, is one of the most striking biogeographic patterns within the clade (Fig. 1; see Pastorini et al., 1998). It is difficult to interpret the occurrence of *Callimico* and *Cebuella* in the upper Amazon, otherwise the domain of tamarins. Either the ancestor to marmosets and *Callimico* originated in the western Amazon, diversified eastward, followed by re-invasion by *Cebuella*, or there were two independent colonizations of the western Amazon from the east by *Callimico* and *Cebuella*.

The larger question remains as to whether the clade consisting of *Leontopithecus*, *Callimico*, and the marmosets originated (1) west of the Madeira in the Amazon and dispersed eastward and southward, (2) east of the Madeira and dispersed both east and west, or (3) further south and east in the Atlantic Forest and dispersed northward and westward. Rylands et al. (1996) suggested that parapatric speciation of submontane/montane forest populations and lowland forest populations of an ancestral callitrichid in southeast Brazil gave rise to the marmoset clade (*Callithrix*, *Cebuella*, and *Mico*) and the lion tamarins (*Leontopithecus*), respectively. This parapatric origin of *Leontopithecus* in southeast Brazil relies on the assumption that the ancestor in question originated there and does not account for the relationships between these two clades and *Callimico*. However, if *Callimico* represents a relict species following divergence from an ancestral southeast Brazilian callitrichid then the entire diversification of the *Leontopithecus*–*Callimico*-marmoset clade may have been driven by some combination of geologic vicariances and/or parapatric speciation due to ecological specializations dividing an ancestrally widespread species.

### 1.3. Biogeography of Saguinus

This genus is the most species-rich of the callitrichids with 33 recognized taxa. It is widespread in several subregions of the Amazon, with a disjunct distribution of three species in northern Colombia and Panama, that makes its biogeographic history of particular interest. Tamarins fall into two main morphological groups based primarily on body size: the smaller saddleback and black-mantle tamarins (the *nigricollis* or white-mouthed group of Hershkovitz, 1977) and the larger species, which include Hershkovitz's (1977) moustached tamarin or *mystax* group, the mottled-face tamarin group (one species: *S. inustus*), the *midas* and *bicolor* groups, and the *oedipus* group (Hershkovitz, 1977; Cropp et al., 1999; Ackermann and Cheverud, 2002). The moustached tamarins (*mystax* group) are broadly sympatric with the saddleback tamarins of the *nigricollis* group, even forming mixed-species groups (Garber, 1992, 1993; Heymann and Buchanan-Smith, 2000; Rylands et al., 2009; see Fig. 2).

Many of the early explanations of tamarin biogeography have been linked to hypothesized evolutionary trends of either increasing or decreasing body size. Hershkovitz (1977; pp. 46–47) believed that small size was primitive for the group and interpreted *Saguinus* diversity as reflecting a trend towards increasing size that originated in the northern Atlantic forest and spread west and north through the Andes and into Colombia and Central America as well as northeast to the Guiana Shield.

Ferrari (1993b), invoking a phylogeny based on phyletic dwarfism (see Martin, 1992), proposed that the large size of the central Amazonian *midas* and *bicolor* groups indicates that these lineages represent the earliest divergences within the group, which later dispersed north and west [giving rise to the mottled-face group (*inustus*), the *oedipus* group, and the *midas* and *bicolor* groups] and, southwest (giving rise to the *mystax* group). The present day sympatry between the *mystax* and *nigricollis* groups is explained by a recent re-invasion of the western Amazon by an ancestral *midas* which gave rise to the *nigricollis* group ancestor that had differentiated enough to allow for sympatry with moustached tamarins (Ferrari, 1993b).

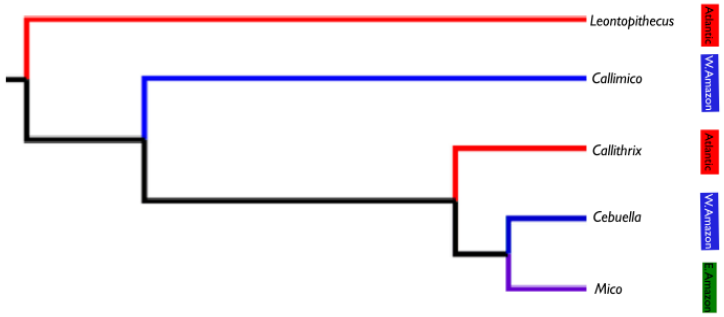


**Fig. 1.** The generic relationships between lion tamarins, marmosets and *Callimico*. Colors correspond to each genus's distributional area: red – Atlantic Forest; blue – Western Amazon (west of the Rio Madeira); Green – Eastern Amazon (east of the Rio Madeira); black branches indicate unknown historical ranges. This striking pattern of alternating regions suggests multiple invasions of the Amazon or Atlantic Forest or both. Figure modified from Schneider and Sampaio, 2015.

**Fig. 2.** Distribution map illustrating the degree of sympatry between small-bodied and large-bodied tamarins in the Amazon. Large tamarins' range indicated in light green, small tamarins' range in blue, and overlap between the two ranges in dark green. Nearly the entire range of small-bodied tamarins is shared with large-bodied tamarins *S. mystax*, *S. imperator*, and *S. labiatus*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The first phylogeny-based biogeographic study (Cropp et al., 1999) revealed patterns that were largely congruent with Ferrari (1993b) (Fig. 3), except that the ancestral *Saguinus* was inferred to range west of the Rio Madeira and south of the Rio Amazonas-Solimões, as suggested by Hershkovitz (1977), with dispersal northeast across the Solimões into the Guiana Shield, and from there northwest to northern Colombia and Panama. A crossing of the Solimões further west gave rise to *Saguinus inustus*, and dispersal west and north gave rise to the *nigricollis* group in the upper Amazon, which subsequently moved back, south and east, to the Madeira into the ranges of the mustached tamarin group that had effectively evolved *in situ*. Although Ackermann and Cheverud (2002) provided a scenario for the dispersal and evolution of tamarins based on craniofacial variation that was very

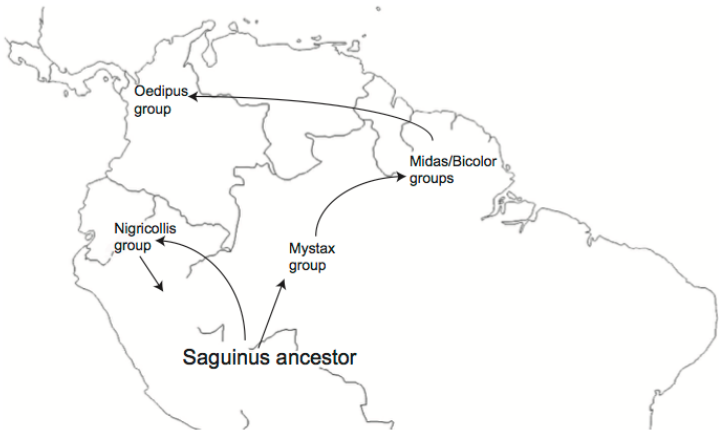


**Fig. 3.** Previous hypotheses for *Saguinus* biogeography adapted and summarized from Cropp et al. (1999) and Ackermann and Cheverud (2002).

similar to that of Cropp et al. (1999), there are presently no rigorous comparative tests of alternative biogeographic hypotheses for tamarins.

We adopt a statistical biogeographic perspective to increase our understanding of evolutionary dispersal and vicariant events in South American mammals and to better understand the processes which have allowed callitrichids to radiate in distinct biomes (the Amazon, the Atlantic Forest, the Cerrado and Caatinga) since the Middle to Late Miocene without substantial shifts in body size and ecology. We construct a time-calibrated phylogenetic framework based on sequence data and apply modern methods based on maximum likelihood in order to infer the origin of callitrichids, the origin and diversification of the marmosets and tamarins, and the history of their colonization of the major biomes.

We find that although a combination of processes including dispersal, range expansion, regional extirpation and vicariance have all played a role in shaping present day distributions, many recent divergences reflect dispersal and counter-invasion of areas occupied by sister species. We also identify several areas for future research to more fully understand the evolutionary history of the callitrichids.

## 2. Materials and methods

### 2.1. Genetic sequences

Sequences were downloaded from GenBank in November 2011 with a taxon-gene matrix in Phylota Browser using the search criterion "Callitrichinae." From this matrix we chose ten genes that allowed for the greatest possible taxonomic coverage, 38 callitrichid species and subspecies in all (see S1). These genes included four mitochondrial genes (CytB, Dloop, 16SrRNA and COII) and six nuclear genes: (von Willebrand Factor, ABCA1, ADORA3, AFF2, DMRT1 and FBN1). The sequences for each species or subspecies in this analysis were concatenated from different individuals for different genes. The sequences were aligned using the MUSCLE algorithm in Geneious Pro 5.5.3 and subsequently edited manually for accuracy.

### 2.2. Model parameters

Gene alignments were entered as partitions into BEAUti 1.7.5. The substitution model used for all genes was HKY, with Gamma as the site heterogeneity model. These models were chosen to allow for independent modeling of transition and transversion rates as well as independent rates among sites while avoiding over-parameterization of the model. The substitution and clock models for the coding mitochondrial genes, Cytochrome b and COII, were linked and the genes further partitioned by codon (1 + 2, 3) with unlinked substitution rates and base frequencies. The substitution and clock models for all included nuclear genes were linked to reduce the number of parameters, as the analysis had difficulty converging when they were unlinked. The tree was modeled under a Yule process with a lognormal relaxed clock model (uncorrelated rates). The phylogeny was time-calibrated with the fossils *Patasola* and *Lagonomico* at the root node for callitrichids (Kay and Fleagle, 2010). We used a lognormal prior with a lower bound of 13.4 my, a median age of 14.4 my and an 95% upper bound of 18.6 my (Opazo et al., 2006). We constrained the ancestral split to be between the *Saguinus* clade and the remainder of the callitrichids on the basis of recent phylogenetic studies (Cortés-Ortiz, 2009; Perelman et al., 2011; Springer et al., 2012).

### 2.3. Biogeographic analysis

Each taxon was assigned to a set of ten geographic subregions defined by existing areas of vertebrate endemism in the Amazon (see Table 1; Cracraft, 1985; Ribas et al., 2012) and breaks between major biomes. These subregions included: Chocó; Marañón; northwestern Amazon (Imeri, Napo and Negro centers of endemism); Inambari; Guianas; southeastern Amazon (Rondônia, Xingu, and Belém centers of endemism); northern Atlantic Forest; southern Atlantic Forest; Cerrado; and Caatinga (see Fig. 4). This division between north and south Atlantic Forest has been shown to be important for many primates (Kinzey, 1982; Torres, 1988). The Cerrado and the Caatinga areas were assigned as subregions due to their unique soils, vegetation and climate as compared to the tropical forests that largely characterize the remainder of the distribution of the callitrichids.

We modeled callitrichid biogeography under six different models (DEC, DECj, DIVA, DIVAj, BayArea, BayAreaj) and compared them for statistical fit using AIC implemented in the R package BioGeoBEARS (Matzke, 2014). Each of the different models allows for a subset of biogeographic possibilities, allowing us to explore the role of different phenomena such as dispersal, vicariance and extinction (Fig. 5). These biogeographic processes are implemented in a maximum likelihood framework as free parameters that are estimated from the data. Note that as these are maximum likelihood implementations of the original parsimony based DIVA method and Bayesian BayArea method, they are really DIVA-like and BayArea-like (however we continue to refer to them as DIVA and BayArea for simplicity). Additionally, BioGeoBEARS can be used with user-defined biogeographic regions rather than georeferenced locality data. We set the maximum number of areas at a given node to five in order to avoid intractability of the analysis. The reconstructions were run under a distance dependent model, where the probability of dispersal is dependent on the distance between the centers of areas (see S2).

We tested whether the ancestral population to all modern callitrichids was (1) widespread across South America or (2) had a

**Table 1**
Biogeographic regions used to reconstruct the biogeographic history of Callitrichidae. The single letter codes correspond to those in Figs. 4 and 7 and are composed of one or more areas of endemism or major biome.

| Biogeographic region | Corresponding areas of endemism/biome | Species occurrences |
|---|---|---|
| A | Chocó | *S. leucopus; S. oedipus; S. geoffroyi* |
| B | Marañon | *S. f. leucogenys; C. goeldii; C. pygmaea; S. mystax* |
| C | Imeri, Napo, Negro | *S. n. graellsi; S. n. nigricollis; S. f. lagonotus; S. tripartitus; S. f. fuscus; C. pygmaea; C. goeldii* |
| D | Inambari | *S. w. melanoleucus; S. w. wedelli; S. f. nigrifrons; S. f. fuscicollis; S. inustus; S. imperator; S. labiatus; C. pygmaea* |
| E | Guianas | *S. midas; S. martinsi; S. bicolor* |
| F | Rondônia, Xingu, Belém | *M. emiliae; M. saterei; M. argentatus; M. humeralifer; M. mauesi; S. niger* |
| G | Cerrado | *C. penicillata* |
| H | Caatinga | *C. jacchus* |
| I | Atlantic Forest (North of 20°) | *L. chrysomelas; C. kuhlii; C. geoffroyi* |
| J | Atlantic Forest (South of 20°) | *L. crysopygus; L. caissara; L. rosalia; C. aurita* |

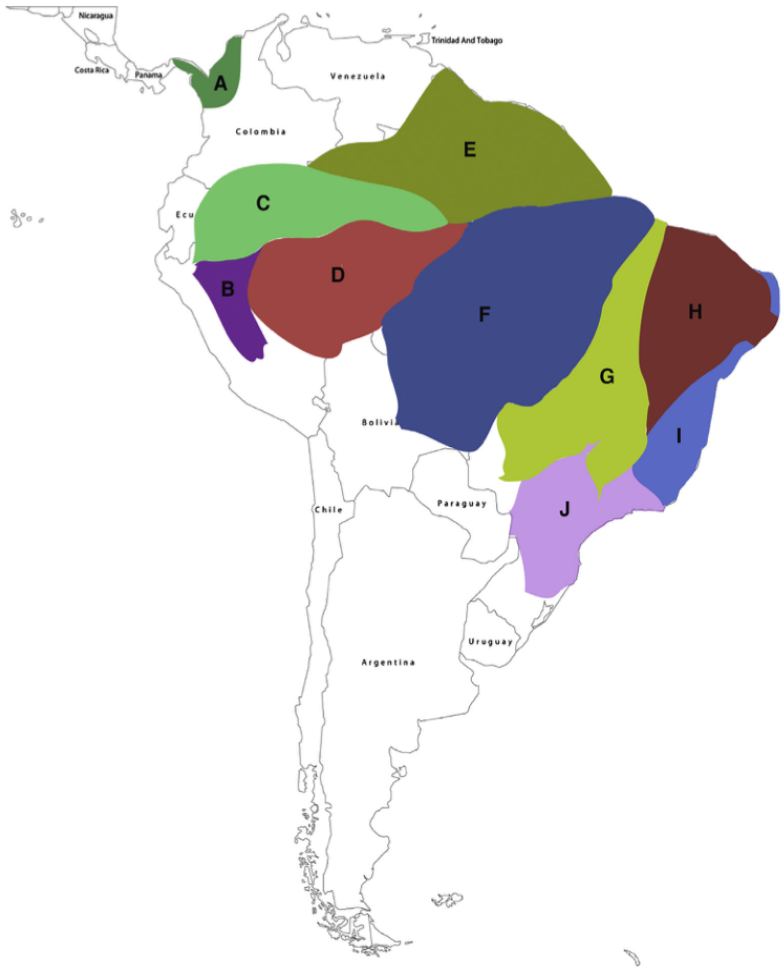


**Fig. 4.** Biogeographic subregions used for the biogeographic analyses. See Section 2.3 and Table 1 for a more detailed description of subregions. (A) Chocó, (B) Marañon, (C) Northwest Amazon, (D) Inambari, (E) Guianas, (F) Southeast Amazon, (G) Cerrado, (H) Caatinga, (I) North Atlantic Forest, and (J) South Atlantic Forest.

more regionally restricted ancestral range. If callitrichid diversification represents the fragmentation of an ancestral range, we would expect biogeographic reconstructions that emphasize vicariance (e.g. DIVA and DEC) to be preferred and the root node should be reconstructed as wide-ranging. In the case of a more restricted ancestral range we expect reconstructions to emphasize dispersal (e.g. BayArea) and the root node should be reconstructed as restricted to one, or two adjacent, regions.

We also tested whether the clade consisting of *Leontopithecus*, *Callimico*, and the marmosets originated (1) west of the Rio Madeira in the Amazon and dispersed eastward, (2) east of the Madeira and dispersed both east and west, or (3) further east in the Atlantic Forest and dispersed westward, or if instead (4) a series of geologic or ecological vicariances subdivided a widespread *Leontopithecus–Callimico*-marmoset ancestor ranging from the Atlantic forest through to the western Amazon. A reconstruction of ancestral areas in regions corresponding to one of the first three scenarios was taken as support for one alternative over the others. In the case of the fourth alternative, we would expect the ancestral range for this clade to be reconstructed in areas spanning from the Amazon to the Atlantic forest (all regions included in the first three alternatives) and support for vicariance leading to diversification and current distributions.

Additionally, we tested whether (1) the *Saguinus* ancestor originated in the southwestern Amazon and dispersal led to current distributions or (2) a widespread *Saguinus* ancestor underwent diversification by vicariance and body size evolution. The analyses will reconstruct the *Saguinus* ancestral range in the southwestern Amazon and support subsequent dispersal into neighboring areas if the ancestor was regionally restricted rather than widespread. Alternatively, models will reconstruct the ancestral range as widespread and later diversification will be associated with vicariance.

Finally, we were interested in the number of movements between the major biome regions (i.e., Amazon, Atlantic Forest, Cerrado and Caatinga). Our analyses allow us to evaluate how many times invasions of the various biomes is supported.

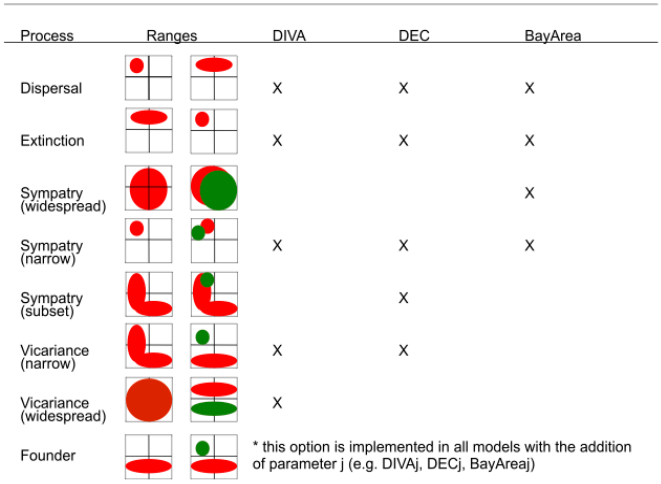


**Fig. 5.** The three main statistical biogeographic models used in our study. The figure indicates which processes can be implemented in each of the biogeographic methods of the BioGeoBEARS package (adapted from Matzke, 2014). Each process is a parameter of the model that is estimated from the data. Each of the three main models has a second version which includes the founder parameter, thus totaling six models.

## 3. Results

### 3.1. Phylogenetics

Our phylogenetic results recovered monophyly for all callitrichid genera and largely corroborate previously reported species relationships based on molecular techniques (Fig. 6). The ancestral node is dated at c.14 Ma and at this time there was a basal split between *Saguinus* and the remainder of the Callitrichidae as expected due to the constrained tree. *Callimico* is placed as sister to the marmosets with 100% node support. We recovered a split between the Atlantic forest and Amazonian marmosets at 5.4 Ma. There is strong evidence for a sister group relationship between *Cebuella* and other Amazonian marmosets (pp = 0.99). For *Saguinus*, we found a well-supported split between the large-bodied and small-bodied tamarin clades at roughly 9.1 Ma (pp = 1.0). Estimated divergence dates for the generic lineages are as follows: *Saguinus* c. 14.0 Ma; *Leontopithecus* c. 13.4 Ma; *Callimico* c. 11.1 Ma; *Callithrix* c. 5.4 Ma; and *Cebuella* split from *Mico* about c. 4.4 Ma (Tables 2 and 3).

### 3.2. Biogeographic reconstruction

Of the six biogeographic models we evaluated, the DIVAj (ΔAIC = 0) and DECj (ΔAIC = 2.9) models produced the best

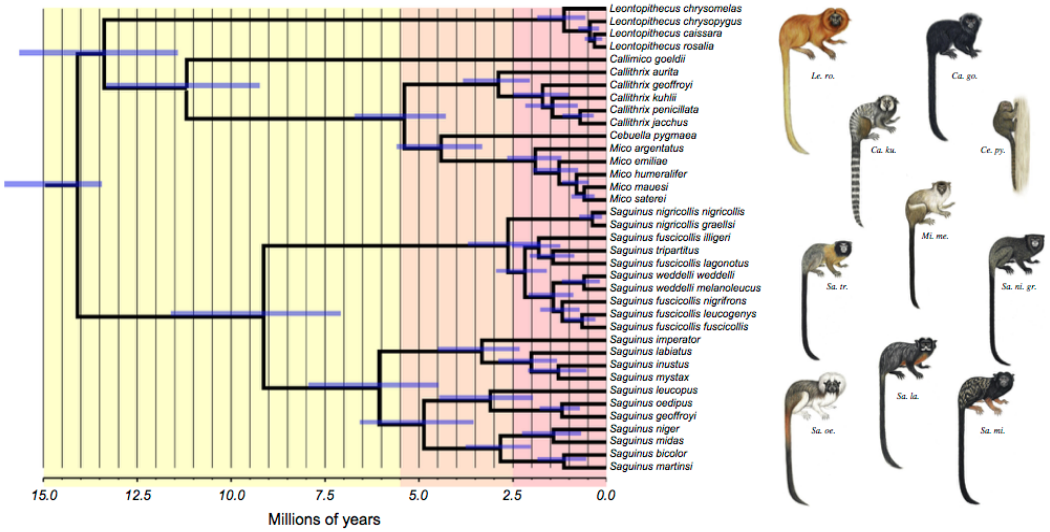


**Fig. 6.** A time-calibrated phylogeny of the Callitrichidae. Shading refers to major geologic time periods: yellow – Miocene, orange – Pliocene, pink – Pleistocene. Node bars indicate the 95% highest posterior density. Illustrations by Stephen Nash. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
Divergence times for major clades in the Callitrichidae in South America.

| Split or clade | Mean age (Ma) | Lower 95% HPD | Upper 95% HPD |
|---|---|---|---|
| *Saguinus* vs. remaining Callitrichidae | 14.1 | 13.4 | 16.0 |
| *Leontopithecus* vs. marmosets + *Callimico* | 13.4 | 11.4 | 15.6 |
| *Callimico* vs. marmosets | 11.2 | 9.2 | 13.3 |
| *Callithrix* vs. *Cebuella* + *Mico* | 5.4 | 4.3 | 6.7 |
| *Cebuella* vs. *Mico* | 4.4 | 3.3 | 5.6 |
| Large-bodied vs. small-bodied tamarins | 9.1 | 7.1 | 11.6 |
| *mystax* group vs. *oedipus* group + *midas*/*bicolor* groups | 6.1 | 4.5 | 7.9 |
| *oedipus* group vs. *midas*/*bicolor* groups | 4.9 | 3.6 | 6.6 |

HPD = Highest Posterior Density.

**Table 3**
Comparison with divergence times from other studies.

| Split or clade | Publication | Mean age (Mya) |
|---|---|---|
| *Saguinus* off-shoot | Matauschek et al. (2011) | 14.23 |
| | Perelman et al. (2011) | 14.89 |
| | Opazo et al. (2006) | 16.00 |
| | Cropp et al. (1999) | 13.5 (based on fossils only) |
| | Buckner et al. (this study) | 14.1 |
| *Leontopithecus* vs. Marmosets + *Callimico* | Perelman et al. (2011) | 13.55 |
| | Opazo et al. (2006) | 14.19 |
| | Buckner et al. (this study) | 13.4 |
| *Callimico* vs. Marmosets | Perelman et al. (2011) | 10.68 |
| | Opazo et al. (2006) | 12.11 |
| | Buckner et al. (this study) | 11.2 |
| *Callithrix* vs. *Cebuella* + *Mico* | Schneider et al. (2012) | 5.25 |
| | Perelman et al. (2011) | 5.96 |
| | Buckner et al. (this study) | 5.4 |
| *Cebuella* vs. *Mico* | Schneider et al. (2012) | 4.00 |
| | Perelman et al. (2011) | 4.82 |
| | Buckner et al. (this study) | 4.4 |
| Large tamarins vs. small tamarins | Matauschek et al. (2011) | 10.07 |
| | Perelman et al. (2011) | 8.42 |
| | Araripe et al. (2008) | 11.9–18.7 |
| | Buckner et al. (this study) | 9.1 |
| *mystax* group vs. *oedipus* group + *midas*/*bicolor* groups | Perelman et al. (2011) | 6.96 |
| | Araripe et al. (2008) | 7.40 |
| | Buckner et al. (this study) | 6.1 |
| *oedipus* group vs. *midas*/*bicolor* groups | Perelman et al. (2011) | 5.34 |
| | Araripe et al. (2008) | 4.30 |
| | Buckner et al. (this study) | 4.9 |

**Table 4**
Comparisons of the Akaike Information Criterion (AIC) score from each of the analyses in BioGeoBEARS.

| Model | AIC (*w*/*j* parameter) | AIC (*j* = 0) |
|---|---|---|
| DEC | 161.6 | 176.7 |
| DIVA | 160.5 | 170.3 |
| BayArea | 176.3 | 219.4 |

statistical fit to the data (Table 4; Fig. 7; see S1 and S2 for results from other models). As shown in Fig. S1, there is uncertainty in the reconstructions for the early history of the group. Fig. 7, however, depicts the reconstructions with the greatest weight at each node.

### 3.2.1. Origin of the callitrichid ancestor

We tested whether the ancestral population to all modern callitrichids was (1) widespread across South America or (2) had a more regionally restricted ancestral range. Both the DIVAj model and DECj model support an ancestrally widespread callitrichid in the forested regions of the Marañón, northwestern Amazon, Inambari and Southern Atlantic Forest at 14 Ma. The southeast Amazon as the fifth region contributing to the ancestral distribution in the DECj model is supported over the northern Atlantic Forest in the DIVAj model (S2). The initial split between *Saguinus* and the remainder of the Callitrichidae was due to range contraction leading to partially vicariant distributions with overlap only in the northwestern Amazon region, as confirmed by both models. A reconstruction of dispersal from a more restricted ancestral area was recovered in a third model, BayAreaj, but was not supported by the data (ΔAIC = 16.2; see S1).

### 3.2.2. Biogeography of Leontopithecus, Callimico, and the marmosets

We used biogeographic modeling to assess whether the clade consisting of *Leontopithecus*, *Callimico*, and the marmosets originated (1) west of the Rio Madeira in the Amazon and dispersed eastward, (2) east of the Madeira and dispersed both east and west, or (3) further east in the Atlantic Forest and dispersed westward, or if instead (4) a series of geologic or ecological vicariances subdivided a widespread *Leontopithecus*–*Callimico*-marmoset ancestor ranging from the Atlantic forest through to the western Amazon. In the DIVAj model, the ancestor to *Leontopithecus*, *Callimico* and the marmosets was isolated in two relict populations shortly after divergence from *Saguinus*. The population that would lead to *Leontopithecus* was isolated in the Atlantic Forest and the other population, which would give rise to *Callimico* and all marmosets, was in the northwestern Amazon around 13 Ma. While the ancestral population leading to *Leontopithecus* remained widespread in the Atlantic Forest until the recent diversification of modern lion tamarins less than 1 Ma, the marmoset ancestor dispersed into the southeastern Amazon approximately 11 Ma, while a population to become *Callimico* dispersed into Marañón, later to expand into the northeastern Amazon. The marmosets would later undergo significant range expansion c. 6 Ma to occupy much of the Amazon and Atlantic Forest, subsequently diversifying into the currently recognized genera. A vicariant event would lead to *Callithrix* and *Cebuella*/*Mico* in the Atlantic Forest and the Amazon, respectively. A *Cebuella*/*Mico* ancestral clade widespread in Marañón, the northwestern Amazon, Inambari and the southeastern Amazon would undergo a second vicariant event about 4.5 Ma leading to *Cebuella* in Marañón, the northwestern Amazon and Inambari, and *Mico* in the southeastern Amazon. Finally, a *C. jacchus*/*C. penicillata* ancestor disperses into the Cerrado and *C. jacchus* disperses to occupy the northern Atlantic Forest, the Cerrado and Caatinga c. 1 Ma. A later vicariant event leaves *C. penicillata* isolated in the Cerrado.

Under the DECj model, the initial *Leontopithecus*/*Callimico*/marmoset ancestor is reconstructed as widespread across Marañón, the northwestern Amazon, the southeastern Amazon and the southern Atlantic Forest upon diverging from *Saguinus* c. 14 Ma. A vicariant event approximately 13 Ma then leads to an isolated population in the Atlantic Forest that will become *Leontopithecus* and a widespread marmoset/*Callimico* ancestor in Marañón, the northwestern Amazon and Inambari. While *Callimico* remains distributed throughout Marañón and the northwestern Amazon, the marmoset ancestor disperses into the northern Atlantic Forest (as the reconstruction of this node in the DECj model has weak support, we consider the DIVAj scenario more likely for the reconstruction of the ancestral geography of this clade, see Fig. S1). Subsequently, this ancestor expands its range to be widespread

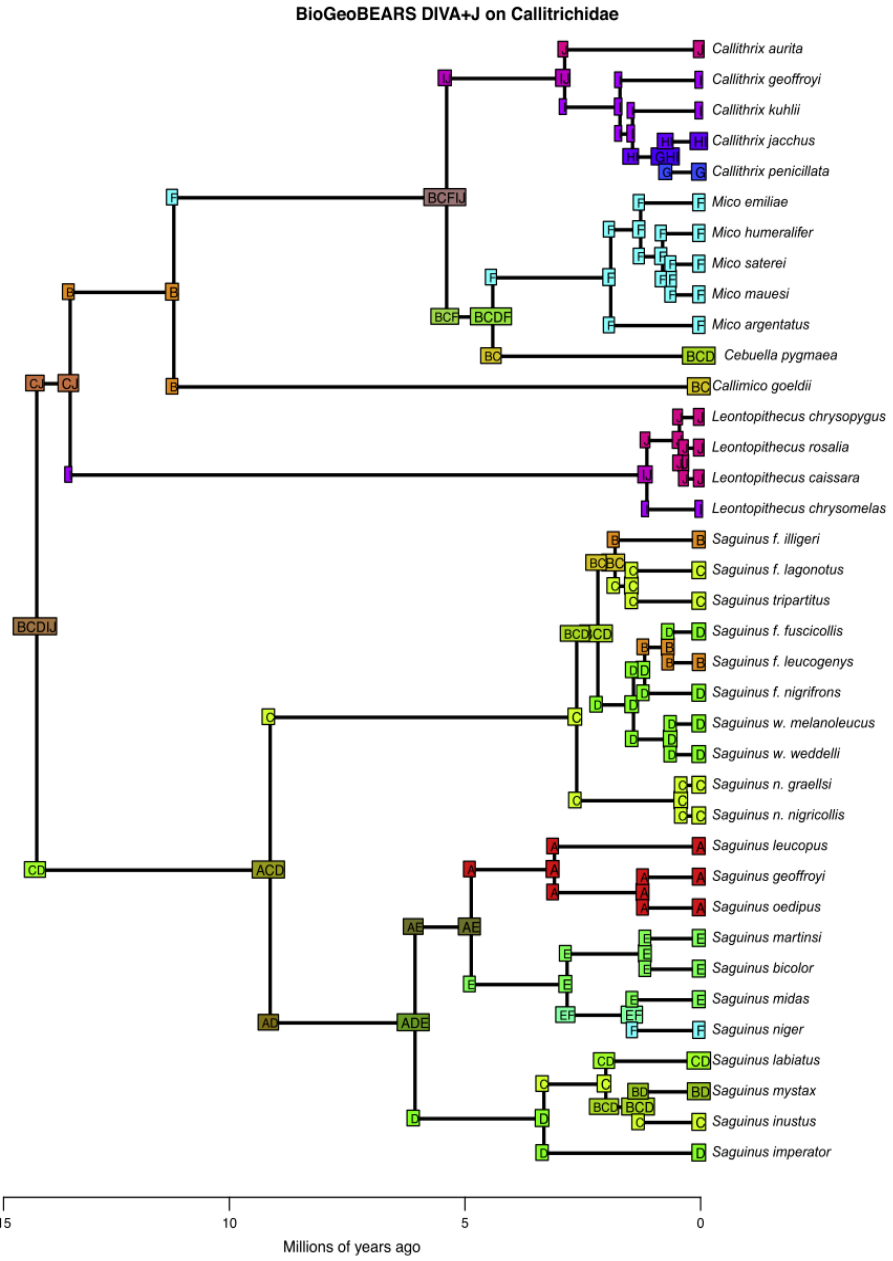


**Fig. 7.** (a) DIVAj and (b) DECj reconstructions of the historical biogeography of Callitrichidae. Letters at each node are the reconstructed ancestral area (as described in Table 1 and Fig. 4). Letters correspond to the regions in Fig. 4.

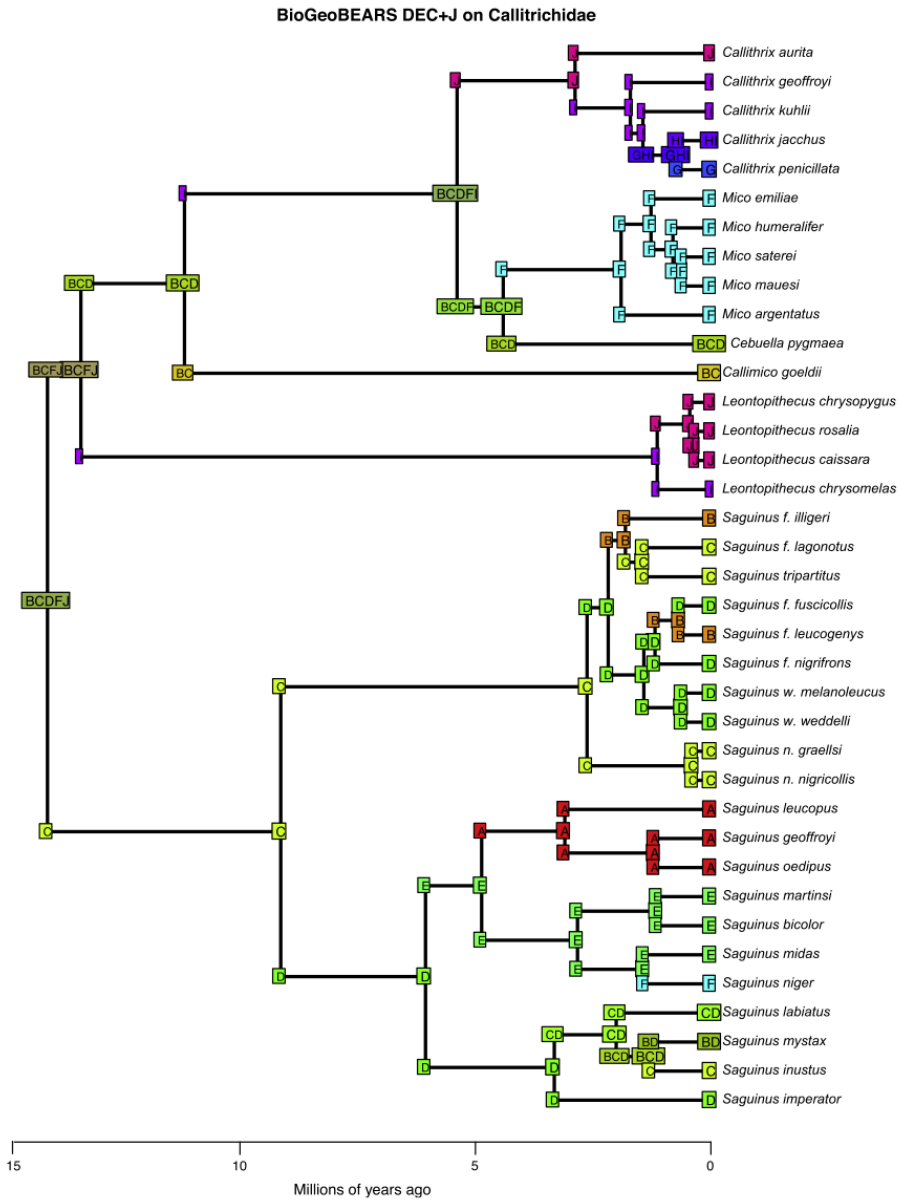


**Fig. 7** (*continued*)

in Marañon, the northwestern Amazon, Inambari, the southeastern Amazon and northern Atlantic Forest. Vicariance between the Atlantic Forest and Amazon regions leads to *Callithrix* in the east and *Mico*/*Cebuella* in the west roughly 6 Ma. Subsequent vicariance leads to *Mico* isolated in the southeastern Amazon and *Cebuella* in Marañón, the northwestern Amazon and Inambari. In agreement with the DIVAj model, a *C. jacchus*/*C. penicillata* ancestor disperses into the Cerrado and *C. jacchus* expands to occupy the northern Atlantic Forest, the Cerrado and Caatinga c. 1 Ma. A later vicariant event leaves *C. penicillata* isolated in the Caatinga.

### 3.2.3. Biogeography of Saguinus

We used modeling to determine if (1) the *Saguinus* ancestor originated in the southwestern Amazon and dispersal led to current distributions or (2) a widespread *Saguinus* ancestor underwent diversification by vicariance and body size evolution. The DIVAj

model indicates that the *Saguinus* radiation expanded to occupy the Chocó, the northwestern Amazon and Inambari with subsequent vicariance leading to the isolation of the small-bodied tamarin ancestor in the northwestern Amazon and the large-bodied ancestor in Chocó and Inambari by c. 9 Ma. The small-bodied tamarin ancestor first diversified in the northwestern Amazon and then underwent range expansion to occupy Marañón and Inambari. Subsequent vicariance and one instance of re-invasion explain current distributions of this group. A large-bodied tamarin ancestor was initially widespread across Chocó, Inambari and the Guianas and vicariance early on followed by dispersal, range expansion and further vicariance led to the current species distributions in all regions of the Amazon (see Section 4.2.1).

The DECj model supports a more restricted *Saguinus* ancestor in the northwestern Amazon around 14 Ma. The initial divergence between the small-bodied and large-bodied ancestral tamarins is explained by the dispersal of the larger ancestor into Inambari by c. 9 Ma. Multiple independent invasions and re-invasions of Marañón, the northwestern Amazon and Inambari by dispersal led to current distributions of small-bodied tamarins. Expansion of the *mystax* group of large-bodied tamarins to occupy the western Amazon and Inambari and later Marañón followed by species specific range contraction determine extant species distributions. Serial dispersal into the Guianas, Chocó and finally the southeastern Amazon explain the distributions of taxa of the *oedipus*, *midas* and *bicolor* groups.

## 4. Discussion

### 4.1. Phylogenetics

Our estimates of callitrichid relationships and divergence times are consistent with recent studies (see Fig. 6; Tables 2, 3). Interestingly, the divergence of the Atlantic forest marmosets and the Amazon marmosets coincides temporally (c. 6 Ma) and geographically with a divergence event in capuchin monkeys (Lynch Alfaro et al., 2012). Like the marmosets, there are two major clades of capuchin monkeys: the gracile or untufted group (*Cebus*) in the Amazon that show a geographic and phylogenetic pattern similar to the tamarin radiation; and the robust or tufted group (*Sapajus*) in the Atlantic Forest, Caatinga and Cerrado, with a pattern similar to the marmoset radiation (*Callithrix*). This may indicate a major environmental event, geological or climatic, that influenced multiple clades in similar ways.

Our analysis supports two major clades in *Saguinus* that correspond to body size in accordance with previous molecular phylogenetic studies (Canavez et al., 1999; Cropp et al., 1999; Tagliaro et al., 2005; Matauschek et al., 2010, 2011). The first clade, which we refer to as the large tamarins, has three subgroupings. The first is the *mystax* group, largely in agreement with the results of Cropp et al. (1999) and Araripe et al. (2008) except that we found *S. imperator* to be the outgroup to *S. labiatus* and *S. mystax*/*S. inustus*, with high node support (pp = 0.99). The second is the *oedipus* group, which consists of the northern Colombian tamarins with 100% node support for *S. leucopus* as the outgroup to *S. oedipus* and *S. geoffroyi*, consistent with Cropp et al. (1999). We did not recover *S. leucopus* as the sister to all other large-bodied tamarins as did Araripe et al. (2008). Finally, the *midas* and *bicolor* groups consist of the tamarins in the Guiana Shield (along with *S. niger*, south of the lower Rio Amazonas on the Brazilian Shield). *S. niger*/*S. midas* are sister to *S. martinsi*/*S. bicolor* in agreement with Tagliaro et al. (2005), Araripe et al. (2008), and Cropp et al. (1999). The *mystax* group is the outgroup to the *oedipus*, *midas* and *bicolor* groups with 100% support and in accordance with previous studies (Tagliaro et al., 2005; Araripe et al., 2008). Notable is the finding that *S. inustus*, the mottled-face tamarin, with no moustache, is nested in the *mystax* group, sister to *S. mystax*.

The second major tamarin clade comprises all the species in Hershkovitz's (1977) *nigricollis* groups—the saddleback and black-mantle tamarins. Our sample of *S. leucogenys* grouped with *S. weddelli*, *S. w. melanoleucus*, *S. f. fuscicollis*, and *S. nigrifrons* suggesting that it was from the south of the Río Pachitea. Matauschek et al. (2011) found that their southern individuals gave this same result, but those sampled from locations north of the Río Pachitea grouped with *S. illigeri*. Uniquely, in our study *S. illigeri* is outgroup to *S lagonotus*/*S. tripartitus*, which makes sense biogeographically as the distributions of these three taxa are adjacent to each other.

The first split in the tamarin clade was c. 9 Ma, preceding that of the marmosets by c. 4 Ma. The two subclades in *Saguinus* resulting from this 9 Ma split are not only genetically divergent but also are differentiated by geographical, morphological and behavioral ecological characteristics, reflecting the depth of their divergence from one another. For example, small-bodied tamarins are found throughout almost all of the *mystax* group's range south of the Amazon and west of the Rio Madeira; this extent of congeneric sympatry is uncommon among platyrrhines. The two subclades differ significantly in size and in foraging strategy; most large-bodied tamarins forage for insects by foliage gleaning in the middle strata of the forest and only occasionally forage on tree trunks, whereas small-bodied tamarins forage in the understory at 5–10 m above the ground searching for hidden insects and small vertebrates in specific sites: in crevices; leaf litter, and on the surface of bark (Garber, 1992, 1993; Heymann and Buchanan-Smith, 2000). The small-bodied tamarins have been somewhat understudied as an evolutionary radiation, probably because, following Hershkovitz (1977) they have all long been attributed to subspecies of just *S. fuscicollis* or *S. nigricollis*. Only recently have some taxa (Peruvian and Ecuadorian) been raised to full species (Matauschek, 2010; Matauschek et al., 2011), and genetic research on the saddleback tamarins in the Brazilian Amazon may well also result in them being given species' status.

In our phylogenetic analysis, the currently recognized *fuscicollis* subspecies are paraphyletic, as was found by Cropp et al. (1999) and Matauschek et al. (2011). Extant *fuscicollis* diversity is the result of a relatively recent radiation (c. 2.5 Ma), possibly explaining the greater extent of diversity in large-bodied tamarins. To resolve the taxonomic nomenclature so that it reflects monophyletic lineages within this group, either all small-bodied tamarins should be considered *fuscicollis* subspecies or, as has occurred in the revision of marmoset taxonomy, should be raised to full species status (Matauschek et al., 2011). The ages of many of the taxa in the small-bodied tamarin clade are older or of comparable age to species of *Callithrix* and *Mico* (Fig. 6). Finally, as discussed in the results, the current widespread sympatry among *Saguinus* clades is shown to result from diversification of the two types in isolation, and their subsequent occupation of the central Amazon. Taken together, we believe that these findings support a split of the genus *Saguinus* into two (see Section 5).

### 4.2. Biogeography

Although our data support the DIVAj ($\Delta$AIC = 0) reconstruction over DECj ($\Delta$AIC = 2.9), we cannot disregard the results from the latter model as the difference between $\Delta$AIC scores is less than 4. Both models support a widespread callitrichid ancestor in South America at 14 Ma and roles for dispersal, range contraction and vicariance in determining extant species distributions. There are some differences in the reconstructions in regard to the processes involved (e.g., dispersal, vicariance) at various times but there is general agreement about the relationship and direction of movement between regions. Common among biogeographic methods

is increasing uncertainty of the reconstructions of ancestral areas in the early history of clades (S1) due to a lack of direct information about historical distributions, extinct clades and paleogeographic history. With this caveat in mind, we discuss the results in light of recent hypotheses of South American paleogeography.

*4.2.1. Amazonian clades*

There are currently two competing hypotheses for Amazonian paleogeography that largely center on the recession of a wetland system that had existed since 23 Ma and the establishment of the Amazon drainage system as it stands today. As rivers are significant barriers for the dispersal of primates (Hershkovitz, 1977), the appearance of the present-day hydrology of the basin was influential in determining extant distributions. The first hypothesis suggests that the recession of the Pebas wetland system began as early as 10 Ma allowing for the establishment of the Amazon drainage system by 7 Ma (Hoorn et al., 1995, 2010; Figueiredo et al., 2009). The second hypothesis argues for a more recent recession of a pan-Amazonian lake so that the drainage system was more recently established, during the Pliocene (c. 2–3 Ma) (Campbell et al., 2006; Latrubesse et al., 2010; Ribas et al., 2012). Our biogeographic reconstructions support a widespread ancestor at c. 14 Ma suggesting recession may have begun even earlier than 10 Ma to allow for this pattern, or that callitrichids occupied regions on the periphery of this water body. However, major diversification does not begin until c. 5 Ma consistent with the establishment of a drainage system more recently, rather than by 7 Ma. This is further supported by the relatively widespread distributions of some extant species today (e.g., *Cebuella pygmaea* and the *nigricollis* group, which have taxa on both sides of the upper Amazon) that were likely present across regions before they were subdivided by major rivers. Crossing rivers through river bend cut-offs (and the formation of ox-bow lakes) will of course have occurred, but mainly in the white-water rivers draining the Andes (Salo et al., 1986). The rivers draining the ancient Guiana and Brazilian shields (clear water and black water rivers) are older and more stable in their courses (Sioli, 1984).

Although much of callitrichid diversification is relatively recent, the split between the pygmy marmoset and the eastern Amazonian marmosets occurs at c. 5 Ma; they are separated by the Rio Madeira, suggesting that it was in place by that time. No eastern Amazonian marmosets occur north of the lower Rio Amazonas, suggesting it too was in place before their diversification in the southeastern Amazon. The marmoset relationships are determined by the lower Rio Amazonas and the Rio Madeira, suggesting that these parts of the drainage were established first, and the major upper portions that now separate the tamarin clades (Rio Marañón, Rio Ucayali and upper Rio Amazonas, for example) may have been established later. This sequence of river establishment is suggested by patterns in other Amazonian taxa as well (Ribas et al., 2012). It would also explain the ties between the Chocó region and the Guianas that both would have been isolated from the southern Amazon by the formation of the lower Rio Amazonas. Further diversification in both marmosets and tamarins would be driven by the continued formation of smaller tributaries to these larger rivers and the progressive afforestation of the central Amazon.

Interestingly, the ranges of *Callimico goeldii* and *Cebuella pygmaea* currently overlap extensively. Izawa (1979) provided the most cogent explanation for the current range of *Callimico*: that it is a specialist on bamboo forest and dense understory, so-called "shabby" forest of terra firma, around the Amazon lake, that is now being gradually replaced by tall forest with more open understories; the patchiness of the vegetation type today determines the patchiness of the occurrence of *Callimico* in the range of *Cebuella pygmaea*. *Cebuella* by contrast is associated with river edge forest and would have moved east between the major Amazon tributaries, flowing largely west to east, along with the dispersing moustached and white-mouthed tamarins.

A hint as to the process which split the today sympatric *nigricollis* and *mystax* groups can be seen in that the first taxa to originate in the *nigricollis* group (*nigricollis* and *graellsi*) and *fuscus* are those in the northwestern Amazon, north of the Rio Amazonas-Solimões, 2.4 Ma (Rylands et al., 2011; see also Cropp et al., 1999; see Fig. 7). The saddlebacks to the south begin to appear later. The *mystax* group tamarins appear a little earlier, first with differentiation of *S. imperator* in Inambari. It is reasonable to believe that a broadly distributed tamarin was caught (like the gracile capuchins, *Cebus*; see Boubli et al., 2012) in an arc along the west bank of the Amazon lake, before 11 Ma, and gradually speciated at the two ends of the range. The subsequent afforestation of the central Amazon from 5 Ma onwards allowed them to invade—the *nigricollis* group to the south and east and the *mystax* group to north and east—between the developing tributaries of the Amazon, oriented essentially west to east. To the south of the Amazon, both were stopped by the long, south-to-north barrier of the Rio Madeira (as was the marmoset genus *Mico* entering from the south in its northwesterly trajectory, see Section 4.2.2). (We note that Weddell's saddleback tamarin, *S. weddelli weddelli*, crossed to the east of the Rio Madeira in one locality, resulting in a small enclave where it is sympatric with *Mico rondoni*; see Ferrari and Martins, 1992.) *Saguinus mystax* evidently crossed the Amazon more than 1 Ma; its isolation resulting in *S. inustus* that dispersed northwest into southern Colombia between the rios Solimões-Japurá and Negro. Before 6 Ma the ancestral large-bodied tamarin was split into the clade in the south west (Inambari) and one isolated north of the Rio Amazonas to give rise to the Guianan (*midas* and *bicolor* groups) and northern Andean forms (*oedipus* group).

*4.2.2. Atlantic Forest, Cerrado and Caatinga clades*

The paleogeographic history of eastern South America is poorly known. We thus base our discussion more directly on our biogeographic reconstruction and the current geographic configuration of the region. The lack of genetic sequences available for species of *Leontopithecus* in addition to their very recent radiation less than 1.5 Ma has made details about their history difficult to uncover. Both biogeographic reconstructions identify *Leontopithecus* as resulting from a relict population isolated in the northern Atlantic Forest. However, the reconstruction at that node invoking a population dispersing to become widespread across the north and south Atlantic Forest and subsequently divided by vicariance (DIVAj model) has the greatest support (Fig. S1). Presently the four species of lion tamarins all have disjunct and minute distributions and are threatened. It is likely that their current decline is an issue of suitable habitat rather than competitive exclusion by late arriving marmosets. Given the very recent radiation of the extant lion tamarin species, fluctuating sea-level and climate cycling during the Quaternary would have caused reductions and expansions of the available lowland habitat, possibly isolating lion tamarin populations in small patches of forest for periods that allowed for speciation (Hershkovitz, 1977; Câmara, 1991; Mundy and Kelly, 2001; Perez-Sweeney et al., 2008).

The separation of *Mico*/*Cebuella* on the one hand and *Callithrix* on the other occurred a little more than 5 Ma, with the former moving north into the Amazon; *Cebuella* being isolated to the west of the Rio Madeira and *Mico* to the east. The two biogeographic models disagree about the nature of the distributional history of Atlantic forest marmosets. While the DIVAj model favors a widespread Atlantic Forest marmoset population undergoing diversification by vicariance, the DECj model invokes a founder event in the northern Atlantic Forest by invaders from the southern Atlantic Forest. In either case, the presence of Atlantic Forest marmosets in the southern Atlantic forest is clear, as the earliest branching

lineage occurs there, *Callithrix aurita*. Their phylogenetic relationships and our biogeographic reconstruction clearly indicate a south to north trajectory through the Atlantic Forest, with their gum-feeding proclivities, obtained while occupying highly seasonal pre-montane/montane habitats in the southeast, gradually evolving to the extent that it allowed them to spread into the highly seasonal habitats of the Cerrado in Central Brazil, and even the Caatinga environments of northeast Brazil (Rylands, 1984, 1996; Rylands et al., 1996). Mycophagy has been documented in some *Callithrix aurita* and *C. flaviceps*, also believed to constitute an essential food at times of seasonal fruit shortage (Hilário, 2008). Invasion of the Cerrado and Caatinga in both reconstructions proceeds from the northern Atlantic forest. The occupation of the driest habitats in the callitrichid range are the most recent expansions, probably due to delayed adaptation to the extremes of these environments.

Note that the extreme behavioral, physiological and morphological adaptations to gum feeding must have arisen twice in the marmosets, once leading to *C. jacchus* and *C. penicillata* in the Atlantic forest group, and in the second case to *Cebuella* and the dwarf marmoset, *M. humilis*, in the Amazon, in different selective regimes (habitat in the Atlantic forest marmosets and competition with tamarins, associated with reduced body size).

## 5. Conclusion

Our study largely corroborates the topologies of recent molecular studies of callitrichid phylogenetics, with divergence times falling within 1–2 million years of previous estimates. The biogeography of the marmosets and tamarins appears to have begun with the fragmentation of a widespread ancestor across the Amazon and Atlantic forests. The establishment of the Amazon river system and the afforestation of the basin, resulting in largely centripetal dispersal with isolation between tributaries, has led to the current diversity of callitrichid species in the region. The Atlantic Forest diversity, however, seems to be the result of specialization in different ecoregions: *Leontopithecus* in the lowlands with its currently recognized species arising only in the last 1.5 my, and *Callithrix* in the highlands and seasonal forest (Atlantic Forest, Cerrado and Caatinga).

Morphology, behavior, ecology, phylogenetics and biogeography all suggest that the tamarins (genus *Saguinus*) should be divided into two genera, splitting the small-bodied and large-bodied groups. Matauschek et al. (2010) also found this early split between the small-bodied and large-bodied tamarins, estimating a divergence date of 9.2 Ma. They too suggested that this separation warrants their classification as distinct genera, following the recommendations of Goodman et al. (1998). *Leontocebus* is a genus created by Wagner (1840; p.9; described on page 248) based on *Simia leonina* Humboldt, thought to be a lion tamarin but subsequently found to be Lesson's saddleback tamarin *Saguinus fuscus* (Lesson, 1840) from the Colombian Amazon (Hershkovitz, 1957; Coimbra-Filho and Mittermeier, 1972). *Leontocebus* is, as far we have been able to discern, the earliest genus name with a type species of a member of Hershkovitz's (1977) *nigricollis* group of saddleback (*fuscicollis*) and black-mantle (*nigricollis*) tamarins, and we recommend it be used to distinguish them from the tamarins of Hershkovitz's moustached, mottle-faced (*inustus*), *midas*, *bicolor* and *oedipus* groups (Groves, 2001).

This study provides a starting point for future research that aims to understand the relationship between historical biogeography and speciation in Callitrichidae. Phylogeographic studies that include dense sampling of the various species would provide valuable information to begin discerning between scenarios such as those provided by the DIVAj and DECj models. More explicit paleogeographic hypotheses for South America would allow for more direct tests of alternative scenarios by allowing for the direct implementation of landscape changes in biogeographic analyses.

## Acknowledgments

Thanks to Graham Slater, Francesco Santini, Alexa Golden, Mericien Venzon, Ryan Ellingson and David Gold for assistance and advice with analyses. Special thanks to David K. Jacobs for assistance with the manuscript and insight for the discussion. We are grateful to Colin P. Groves for his insights as to the appropriateness of the name *Leontocebus* Wagner, 1840, for the white-mouthed tamarins of the Amazon basin. Funding for the presentation of this research at the 2012 meeting of the International Primatological Society in Cancún, Mexico was provided by the Department of Ecology and Evolutionary Biology, UCLA. Funding for the presentation at the 2013 meeting of the Society for Integrative and Comparative Biology was provided by the Charlotte Mangum Student Support Program.

Illustrations copyright 2012 Stephen D. Nash/Conservation International; used with permission.

## Appendix A. Supplementary data

Reconstruction based on the BioGeoBEARS analyses with accompanying trees with pie charts to show the support for the reconstruction at each node (a) DIVA, (b) DIVA+J, (c) DEC, (d) DEC+J, (e) BayArea, and (f) BayArea+J.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.04.031.

## References

Ackermann, R.R., Cheverud, J.M., 2002. Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation. Am. J. Phys. Anthropol. 117, 260–271.

Araripe, J., Tagliaro, C.H., Rego, P.S., Sampaio, I., Ferrari, S.F., Schneider, H., 2008. Molecular phylogenetics of large-bodied tamarins, *Saguinus* spp. (Primates, Platyrrhini). Zool. Scripta 37, 461–467.

Boubli, J.P., Rylands, A.B., Farias, I., Alfaro, M., Lynch Alfaro, J., 2012. *Cebus* phylogenetic relationships: a preliminary reassessment of the diversity of the untufted capuchin monkeys. Am. J. Primatol. 74, 381–393.

Câmara, I.de G., 1991. Plano de Ação para a Mata Atlântica. Fundação SOS Mata Atlântica, São Paulo.

Campbell Jr., K.E., Frailey, C.D., Romero-Pittman, L., 2006. The Pan-Amazonian Ucayali Peneplain, late Neogene sedimentation in Amazonia, and the birth of the modern Amazon River system. Palaeogeogr. Palaeoclimatol. Palaeoecol. 239, 166–219.

Canavez, F.C., Moreira, M.A.M., Simon, F., Parham, P., Seuánez, H.N., 1999. Phylogenetic relationships of the Callitrichinae (Platyrrhini, Primates) based on ß2-microglobulin DNA sequences. Am. J. Primatol. 48, 225–236.

Coimbra-Filho, A.F., Mittermeier, R.A., 1972. Taxonomy of the genus *Leontopithecus* Lesson 1840. In: Bridgwater, D.D. (Ed.), Saving the Lion Marmoset. Wild Animal Propagation Trust, Wheeling, WV, pp. 7–22.

Cortés-Ortiz, L., 2009. Molecular phylogenetics of the Callitrichidae with an emphasis on the marmosets and Callimico. In: Ford, S.M., Porter, L.M., Davis, L.C. (Eds.), The Smallest, Anthropoids: The Marmoset/Callimico Radiation. Springer, New York, pp. 3–24.

Cracraft, J., 1985. Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. Am. Ornithol. Union Ornithol. Monogr. 36, 49–84.

Cropp, S.J., Larson, A., Cheverud, J.M., 1999. Historical biogeography of tamarins, genus *Saguinus*: the molecular phylogenetic evidence. Am. J. Phys. Anthropol. 108, 65–89.

Ferrari, S.F., 1993a. Ecological differentiation in the Callitrichidae. In: Rylands, A.B. (Ed.), Marmosets and Tamarins: Systematics, Ecology, and Behaviour. Oxford University Press, Oxford, pp. 314–328.

Ferrari, S.F., 1993b. The adaptive radiation of Amazonian callitrichids (Primates, Platyrrhini). Evol. Biol. 7, 81–103.

Ferrari, S.F., Martins, E.S., 1992. Gummivory and gut morphology in two sympatric callitrichids (*Callithrix emiliae* and *Saguinus fuscicollis weddelli*) from western Brazilian Amazonia. Am. J. Phys. Anthropol. 88, 97–103.

Ferrari, S.F., Sena, L., Schneider, M.P.C., Silva Jr., J.S., 2010. Rondon's marmoset, *Mico rondoni* sp. n., from southwestern Brazilian Amazonia. Int. J. Primatol. 31, 693–714.

Figueiredo, J., Hoorn, C., van der Ven, P., Soares, E., 2009. Late Miocene onset of the Amazon River and the Amazon deep-sea fan: evidence from the Foz do Amazonas Basin. Geology 37, 619–622.
Garber, P.A., 1992. Vertical clinging, small body size, and the evolution of feeding adaptations in the Callitrichinae. Am. J. Phys. Anthropol. 88, 469–482.
Garber, P.A., 1993. Feeding ecology and behaviour of the genus *Saguinus*. In: Rylands, A.B. (Ed.), Marmosets and Tamarins: Systematics, Behaviour, and Ecology. Oxford University Press, Oxford, pp. 273–295.
Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G., Groves, C.P., 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. 9, 585–598.
Groves, C.P., 2001. *Primate Taxonomy*. Smithsonian Institution Press, Washington, DC.
Groves, C.P., 2005. Order Primates. In: Wilson, D.E., Reeder, D.M. (Eds.), Mammal Species of the World: A Taxonomic and Geographic Reference, third ed., vol. 1. Johns Hopkins University Press, Baltimore, MD, pp. 111–184.
Hershkovitz, P., 1957. The systematic position of the marmoset, Simia leonina Humboldt (Primates). Proc. Biol. Soc. Wash. 70, 17–20.
Hershkovitz, P., 1977. Living New World Monkeys (Platyrrhini) with an Introduction to Primates, vol. 1. The Chicago University Press, Chicago.
Heymann, E.W., Buchanan-Smith, H.M., 2000. The behavioural ecology of mixed-species troops of callitrichine primates. Biol. Rev. Cambridge Phil. Soc. 75, 169–190.
Hilário, R.R., 2008. Padrão de actividades, dieta e uso de hábitat por *Callithrix flaviceps* na Reserva Biológica Augusto Ruschi, Santa Teresa, ES, Masters Thesis, Universidade Federal de Minas Gerais.
Hoorn, C., Guerrero, J., Sarmiento, G.A., Lorente, M.A., 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. Geology 23, 237–240.
Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., SanMartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T., Antonelli, A., 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330, 927–931.
Izawa, K., 1979. Studies on the peculiar distribution pattern of *Callimico*. Kyoto Univ. Overseas Res. Rep. New World Monkeys, pp. 23–41.
Kay, R.F., Fleagle, J.G., 2010. Stem taxa, homoplasy, long lineages and the phylogenetic position of *Dolichocebus*. J. Hum. Evol. 59, 218–222.
Kinzey, W.G., 1982. Distribution of primates and forest refuges. In: Prance, G.T. (Ed.), Biological Diversification in the Tropics. Columbia University Press, New York, pp. 455–482.
Latrubesse, E.M., Cozzuol, M., Silva-Caminha, S.A.F.da, Rigsby, C.A., Absy, M.L., Jaramillo, C., 2010. The late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. Earth-Sci. Rev. 99, 99–124.
Lesson, R.-P., 1840. Species des Mammifères Bimanes et Quadrumanes; Suivi d'un Mémoire sur les Oryctéropes. J.-B. Baillière, Paris, 292 pp.
Lynch Alfaro, J.W., Boubli, J.P., Olson, L.E., Di Fiore, A., Wilson, B., Gutierrez-Espeleta, G.A., Chiou, K.L., Schulte, M., Neitzel, S., Ross, V., Schwochow, D., Nguyen, M., Farias, I., Janson, C., Alfaro, M.E., 2012. Explosive Pleistocene range expansion leads to widespread Amazonian sympatry between robust and gracile capuchin monkeys. J. Biogeog. 39, 272–288.
Martin, R.D., 1992. Goeldi and the dwarfs: the evolutionary biology of the small New World monkeys. J. Hum. Evol. 22, 367–393.
Matauschek, C., 2010. Taxonomy, Phylogeny and Distribution of Tamarins (Genus Saguinus Hoffmannsegg, 1807). Doctoral Dissertation, Georg-August Universität, Göttingen.
Matauschek, C., Heymann, E.W., Finstermeier, M., Roos, C., 2010. Complete mitochondrial genome data reveal the phylogeny of callitrichine primates and a later Miocene divergence of tamarin species groups. In: Matauschek, C. (Eds.), Taxonomy, Phylogeny and Distribution of Tamarins (Genus *Saguinus* Hoffmannsegg, 1807). Doctoral Dissertation, Georg-August Universität, Göttingen, pp. 33–54.
Matauschek, C., Roos, C., Heymann, E.W., 2011. Mitochondrial phylogeny of tamarins with taxonomic and biogeographic implications for the *S. nigricollis* species group. Am. J. Phys. Anthropol. 144, 564–574.
Matzke, N.J., 2014. Package 'BioGeoBEARS'. BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R scripts. Version 0.2.1. 360pp. <http://cran.r-project.org/web/packages/BioGeoBEARS/BioGeoBEARS.pdf>.
Mundy, N.I., Kelly, J., 2001. Phylogeny of lion tamarins based on interphotoreceptor retinol binding protein intron sequences. Am. J. Primatol. 54, 33–40.
Napier, J.R., Napier, P.H., 1967. A Handbook of Living Primates. Academic Press, London.
Opazo, J.C., Wildman, D.E., Prychitko, T., Johnson, R.M., Goodman, M., 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Mol. Phylogenet. Evol. 40, 274–280.
Pastorini, J., Forstner, M.R.J., Martin, R.D., Melnick, D.J., 1998. A reexamination of the phylogenetic position of *Callimico* (Primates) incorporating new mitochondrial DNA sequence data. J. Mol. Evol. 47 (1), 32–41.
Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., et al., 2011. A molecular phylogeny of living primates. PLoS Genet. 7 (3), e1001342.
Perez-Sweeney, B.M., Valladares-Padua, C., Martins, C.S., Morales, J.C., Melnick, D.J., 2008. Examination of the taxonomy and diversification of *Leontopithecus* using the mitochondrial control region. Int. J. Primatol. 29, 245–263.
Ribas, C.C., Aleixo, A., Nogueira, A.C.R., Miyaki, C.Y., Cracraft, J., 2012. A paleobiogeographic model for biotic diversification within Amazonia over the past three million years. Proc. Roy. Soc. B, Biol. Sci. 279, 681–689.
Röhe, F., Silva Jr, J.de S., Sampaio, R., Rylands, A.B., 2009. A new subspecies of *Saguinus fuscicollis* (Primates, Callitrichidae). Int. J. Primatol. 30, 533–551.
Rosenberger, A.L., 1981. Systematics: the higher taxa. In: Coimbra-Filho, A.F., Mittermeier, R.A. (Eds.), Ecology and Behavior of Neotropical Primates, vol. 1. Academia Brasileira de Ciências, Rio de Janeiro, pp. 9–27.
Rosenberger, A.L., 2011. Evolutionary morphology, platyrrhine evolution, and systematics. Anat. Rec. 294, 1955–1974.
Rylands, A.B., 1984. Exudate-eating and tree-gouging by marmosets (Callitrichidae, Primates). In: Chadwick, A.C., Sutton, S.L. (Eds.), Tropical Rain Forest: The Leeds Symposium. Leeds Philosophical and Literary Society, Leeds, pp. 155–168.
Rylands, A.B., 1996. Habitat and the evolution of social and reproductive behavior in Callitrichidae. Am. J. Primatol. 38, 5–18.
Rylands, A.B., Mittermeier, R.A., 2009. The diversity of the New World primates: an annotated taxonomy. In: Garber, P.A., Estrada, A., Bicca-Marques, J.C., Heymann, E.W., Strier, K.B. (Eds.), South American Primates: Comparative Perspectives in the Study of Behavior, Ecology, and Conservation. Springer, New York, pp. 23–54.
Rylands, A.B., Schneider, H., Langguth, A., Mittermeier, R.A., Groves, C.P., Rodríguez-Luna, E., 2000. An assessment of the diversity of New World primates. Neotrop. Primates 8, 61–93.
Rylands, A.B., da Fonseca, G.A.B., Leite, Y.L.R., Mittermeier, R.A., 1996. Primates of the Atlantic forest: origin, endemism, distributions and communities. In: Norconk, M.A., Rosenberger, A.L., Garber, P.A. (Eds.), Adaptive Radiations of the Neotropical Primates. Plenum Press, New York, pp. 21–51.
Rylands, A.B., Coimbra-Filho, A.F., Mittermeier, R.A., 2009. The systematics and distributions of the marmosets and *Callimico*. In: Ford, S.M., Porter, L.M., Davis, L.C. (Eds.), The Smallest Anthropoids: The Marmoset/Callimico Radiation. Springer, New York, pp. 25–61.
Rylands, A.B., Matauschek, C., Aquino, R., Encarnación, F., Heymann, E.W., de la Torre, S., Mittermeier, R.A., 2011. The range of the golden-mantle tamarin, *Saguinus tripartitus* (Milne Edwards, 1878): distributions and sympatry of four tamarins in Colombia, Ecuador, and northern Peru. Primates 52 (1), 25–39.
Salo, J., Kalliola, R., Häkkinen, I., Mäkinen, Y., Niemalä, P., Puhakka, M., Coley, P.D., 1986. River dynamics and the diversity of Amazon lowland forest. Nature 322, 254–258.
Schneider, H., Sampaio, I., 2015. The systematics and evolution of New World primates—a review. Mol. Phylogenet. Evol. 82 (PB), 348–357.
Schneider, J., Bernardi, J.A.R., da Cunha, D.B., Tagliario, C.H., Vallinoto, M., Ferrari, S.F., Sampaio, I., 2012. A molecular analysis of the evolutionary relationships in the Callitrichinae, with emphasis on the position of the dwarf marmoset. Zool. Scripta 41, 1–10.
Sioli, H., 1984. The Amazon and its main affluents: hydrography, morphology of the river courses, and river types. In: Sioli, H. (Ed.), The Amazon: Limnology and Landscape Ecology of a Mighty Tropical River and its Basin. Junk, Dordrecht, pp. 127–165.
Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L., Stadler, T., Steiner, C., Ryder, O.A., Janečka, J.E., Fisher, C.A., Murphy, W.J., 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS ONE 7, e49521.
Tagliaro, C.H., Schneider, H., Sampaio, I., Schneider, M.P.C., Vallinoto, M., Stanhope, M., 2005. Molecular phylogeny of the genus *Saguinus* (Platyrrhini, Primates) based on the ND1 mitochondrial gene and implications for conservation. Genet. Mol. Biol. 28, 46–53.
Torres, C., 1988. Resultados preliminares de reavaliação das raças do macaco-prego *Cebus apella* (Primates: Cebidae). Rev. Nordest. Biol. 6, 15–28.
Van Roosmalen, M.G.M., Van Roosmalen, T., 2003. The description of a new marmoset genus, *Callibella* (Callitrichinae, Primates), including its molecular phylogenetic status. Neotrop. Primates 11, 1–10.
Wagner, J.A., 1840 (1839). Die Säugethiere in Abbildungen nach der Natur, Supplement 1. ix (parte). Weigel, Leipzig.

## APPENDIX

The supplementary files from the publication of chapter one are included in this appendix. These include a table of information on the genetic data used in the study. It also includes tables and figures from the full set of analyses for the biogeographical reconstructions performed in BioGeoBEARS. It also includes the parameter estimations from the various biogeographic models.

**Table 1-5.** Gene Matrix and Genbank Accession Numbers

| Taxon | 16S rRNA | ABCA1 | ADORA3 | AFF2 | COII | CytB | DLoop | Von Willebrand Factor | DMRT1 | FBN1 |
|---|---|---|---|---|---|---|---|---|---|---|
| C penicillata | | HM765304.1 | HM765166.1 | HM764949.1 | AY118196.1 | AF245062.1 | U88997.1 | | HM762552.1 | HM761916.1 |
| C kuhlii | | HM765297.1 | HM765165.1 | HM764940.1 | AY118193.1 | AF245064.1 | U88992.1 | | | |
| C aurita | | | HM765162.1 | HM764917.1 | AY118188.1 | AF245049.1 | U89001.1 | | | |
| C jacchus | U39001.1 | HM765296.1 | HM765164.1 | HM764939.1 | AB572419.1 | AB572419.1 | AY196768.1 | AF092828.1 | HM762543.1 | HM761915.1 |
| C geoffroyi | | HM765290.1 | HM765163.1 | HM764933.1 | AY118192.1 | HM368005.1 | U88994.1 | AF092823.1 | HM762537.1 | HM761914.1 |
| C pygmaea | U29002.1 | HM765307.1 | HM765167.1 | HM764952.1 | AY118176.1 | AF245080.1 | U89010.1 | AF092818.1 | HM762555.1 | HM761917.1 |
| C humilis | FJ769145.1 | | | | | | FJ769148.1 | | | |
| C goeldii | U39000.1 | HM765291.1 | HM765160.1 | HM764934.1 | AY118175.1 | AF245089.1 | Jacobs-Cropp et al (1999) | AF092824.1 | HM762538.1 | HM761912.1 |
| L rosalia | U39006.1 | HM765337.1 | HM765169.1 | HM764985.1 | | AF245084.1 | | AF092831.1 | HM762590.1 | HM761919.1 |
| L chrysopygus | | | | | | | Perez-Sweeney et al (2008) | | | |
| L chrysomelas | | HM765332.1 | HM765168.1 | HM764980.1 | | L44589.1 | | | HM762585.1 | HM761918.1 |
| L caissara | | | | | | | Perez-Sweeney et al (2008) | | | |
| M argentatus | | HM765279.1 | HM765161.1 | HM764921.1 | AY118182.1 | AF245065.1 | U89005.1 | AF092816.1 | HM762525.1 | HM761913.1 |
| M emiliae | FJ769146.1 | | | | AY118178.1 | L44587.1 | | | | |
| M saterei | | | | | AY118180.1 | | FJ769152.1 | | | |
| M humeralifer | | HM765345.1 | HM765170.1 | HM764994.1 | AY118184.1 | AF245052.1 | U89008.1 | | | |
| M mauesi | FJ769147.1 | | | | AY118187.1 | AF245051.1 | U89006.1 | | | |
| S imperator | EU497287.1 | HM765394.1 | | HM765046.1 | | HM368020.1 | DQ241249.1 | | HM762651.1 | HM761946.1 |
| S labiatus | EU497289.1 | HM765395.1 | | HM765047.1 | | HM367998.1 | HM367897.1 | | HM762652.1 | HM761947.1 |
| S mystax | EU497294.1 | HM765398.1 | HM765202.1 | HM765050.1 | | HM367983.1 | DQ257674.1 | | HM762655.1 | HM761951.1 |
| S inustus | | | | | | | Jacobs-Cropp et al (1999) | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| S leucopus | EU497286.1 | | | | | Jacobs-Cropp et al (1999) | Jacobs-Cropp et al (1999) | | | |
| S oedipus | EU497296.1 | HM765399.1 | HM765203.1 | HM765051.1 | | HM36007.1 | HM367907.1 | | HM762656.1 | HM761952.1 |
| S geoffroyi | U39008.1 | HM765392.1 | HM765198.1 | HM765044.1 | | Jacobs-Cropp et al (1999) | Jacobs-Cropp et al (1999) | | HM762649.1 | HM761945.1 |
| S niger | EU497246.1 | | | | | | DQ241234.1 | | | |
| S midas | EU497273.1 | HM765396.1 | HM765200.1 | HM765048.1 | | AJ489760.1 | DQ241240.1 | | HM762653.1 | HM761950.1 |
| S martinsi | EU497276.1 | HM765402.1 | | HM765054.1 | | | DQ241246.1 | | HM762659.1 | HM761948.1 |
| S bicolor | EU497280.1 | HM765388.1 | HM765196.1 | HM765039.1 | | Jacobs-Cropp et al (1999) | DQ257686.1 | AF092834.1 | HM762644.1 | HM761944.1 |
| S f fuscus | | | | | | Jacobs-Cropp et al (1999) | Jacobs-Cropp et al (1999) | | | |
| S graellsi | | | | | | HM368035.1 | HM367933.1 | | | |
| S n nigricollis | | | | | | HM367992.1 | HM367946.1 | | | |
| S f illigeri | | | | | | HM368066.1 | HM367899.1 | | | |
| S f lagonotus | | | | | | HM368021.1 | HM367949.1 | | | |
| S tripartitus | | | | | | HM368006.1 | HM367976.1 | | | |
| S m melanoleucus | | | | | | HM368078.1 | HM367978.1 | | | |
| S f wedelli | | | | | | HM368065.1 | HM367963.1 | | | |
| S fuscicollis | | HM765391.1 | HM765197.1 | HM765043.1 | | | | AF092835.1 | HM762648.1 | |
| S f nigrifrons | | | | | | HM368008.1 | HM367908.1 | | | |
| S f fuscicollis | | | | | | HM368072.1 | HM367972.1 | | | |
| S f leucogenys | | | | | | HM368039.1 | HM367939.1 | | | |

**Table 1-6.** Parameter estimations from the analyses in BioGeoBEARS.

| Analysis | LnL | Number of Parameters | d | e | x | j |
|---|---|---|---|---|---|---|
| DEC | -86.4 | 3 | 0.306701825 | 1.000000e-15 | -0.46135383 | 0.0000000 |
| DEC+J | -77.8 | 4 | 0.055201736 | 1.000000e-15 | -0.31629427 | 0.27071479 |
| DIVA | -83.2 | 3 | 0.036156454 | 1.000000e-15 | -0.13291172 | 0.0000000 |
| DIVA+J | -77.2 | 4 | 0.036983231 | 1.000000e-15 | -0.21069037 | 0.08733097 |
| BayArea | -107.7 | 3 | 0.014129693 | 1.890184e-01 | 0.01315059 | 0.0000000 |
| BayArea+J | -85.2 | 4 | 0.004227238 | 2.183400e-03 | -0.02550378 | 0.04601126 |

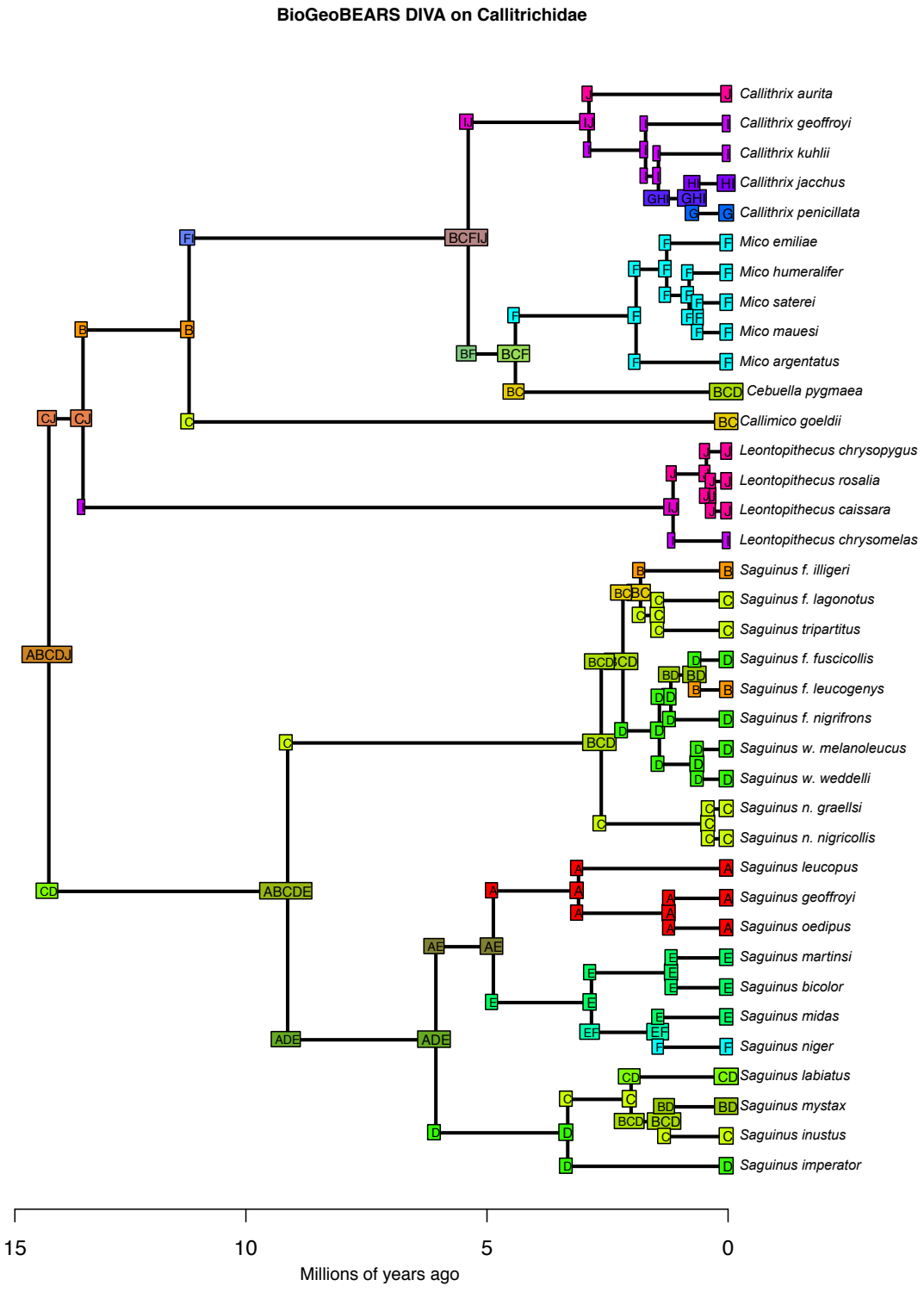


**Figure 1-8 (a).** DIVA-Like reconstruction of the biogeographic history of Callitrichidae.

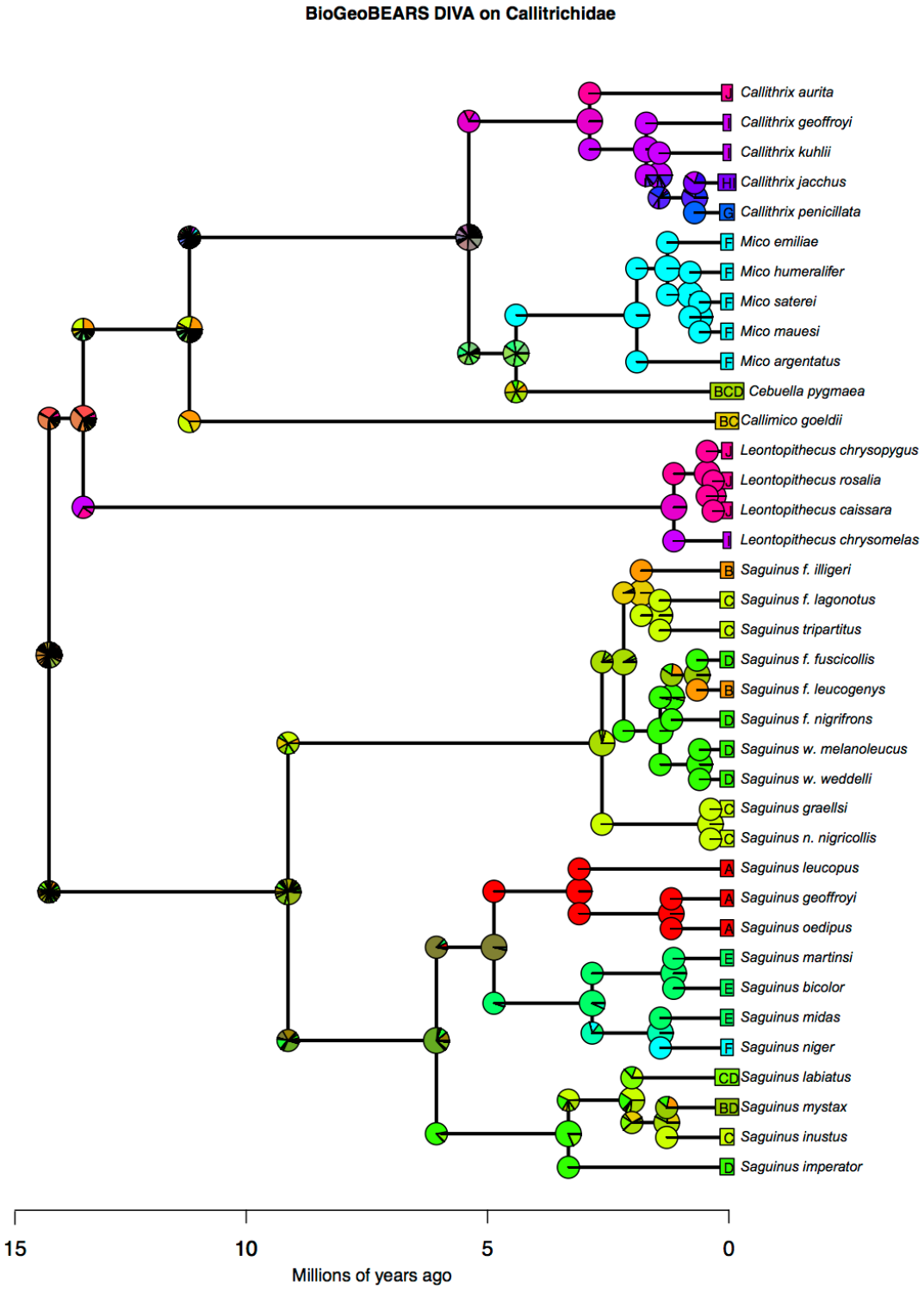


**Figure 1-8 (b).** Support for area configurations from DIVA-Like reconstruction of the biogeographic history of Callitrichidae.

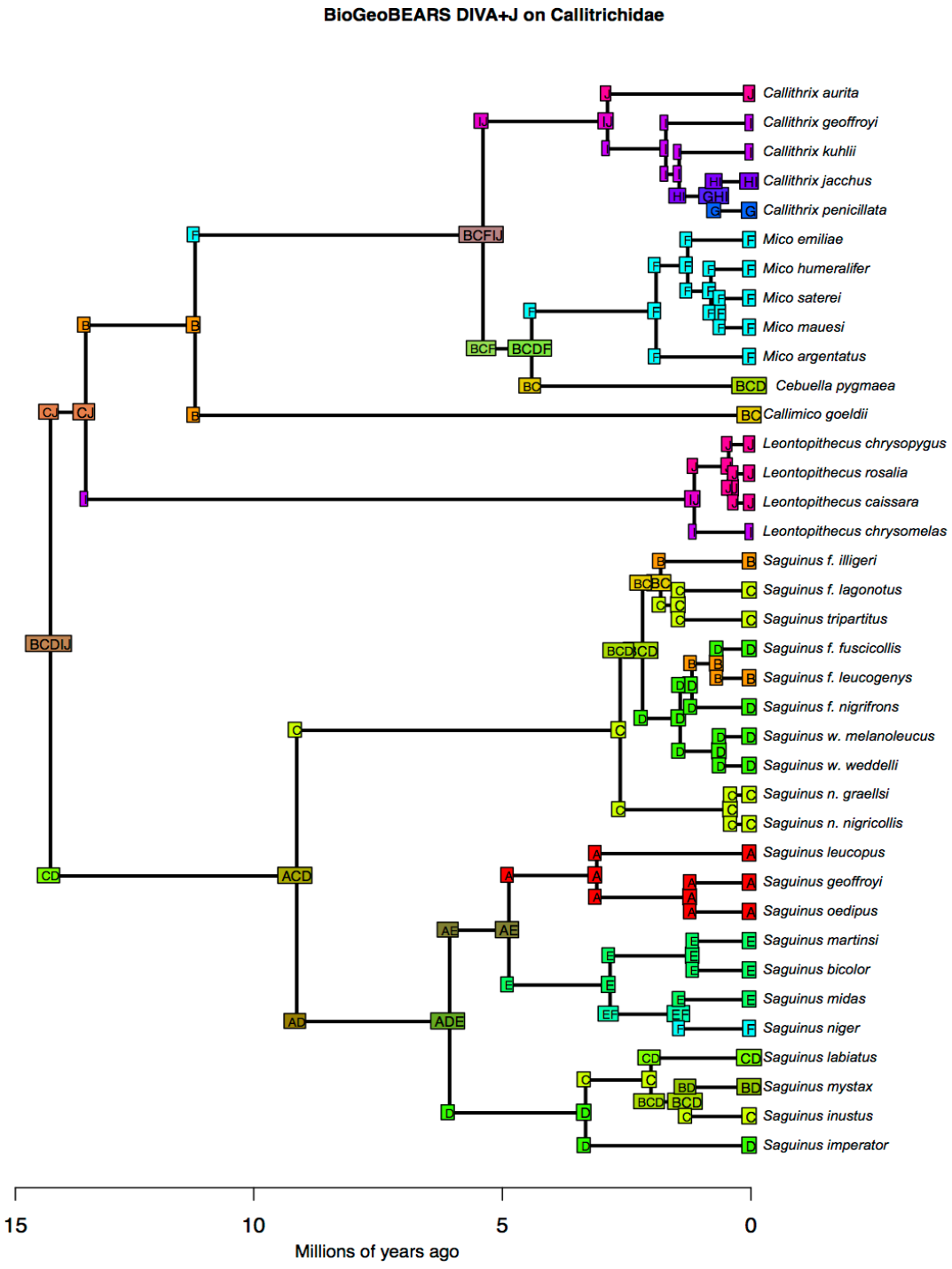


**Figure 1-8 (c).** DIVA-Like+j reconstruction of the biogeographic history of Callitrichidae.

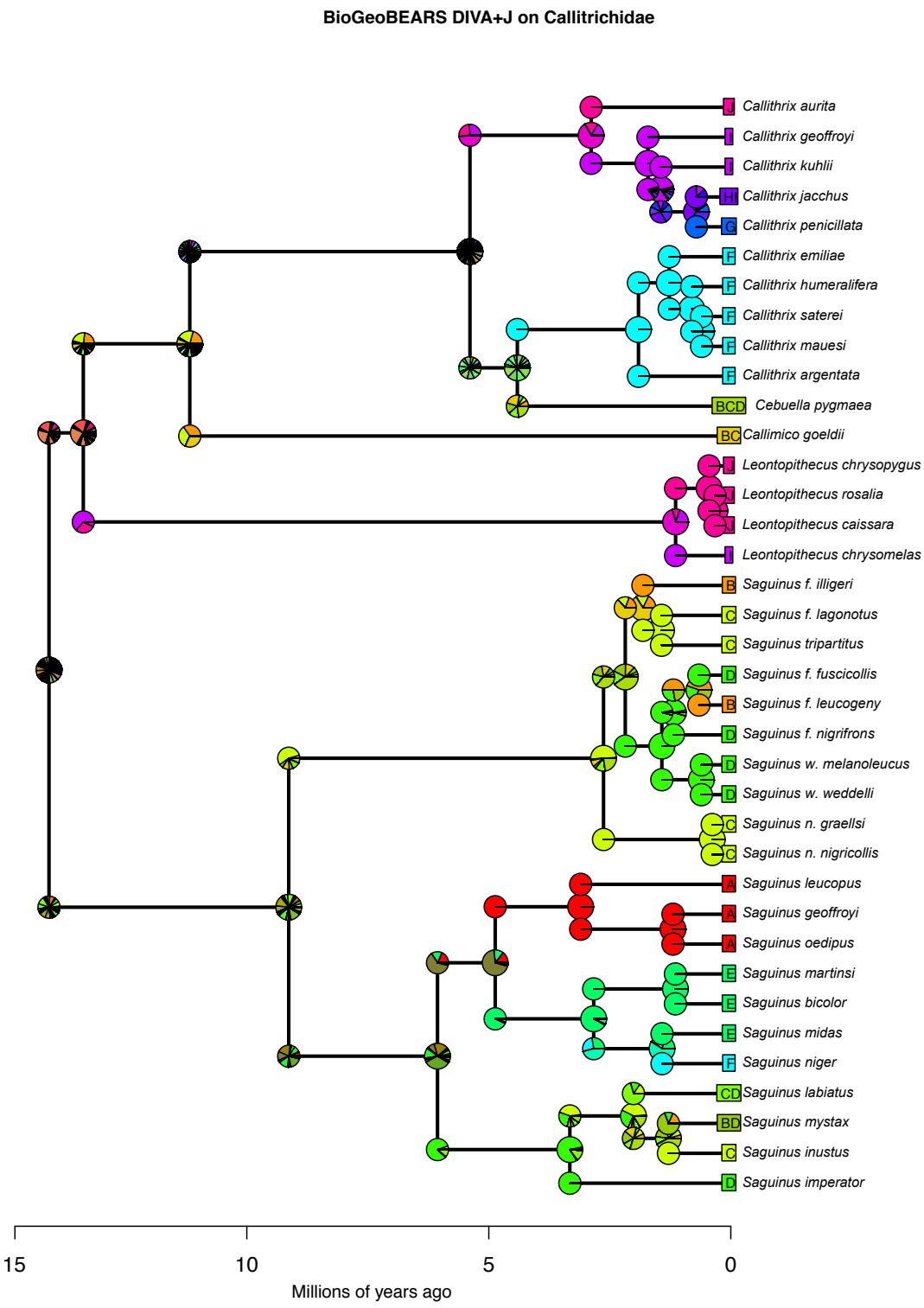


**Figure 1-8 (d).** Support for area configurations from DIVA-Like+j reconstruction of the biogeographic history of Callitrichidae.

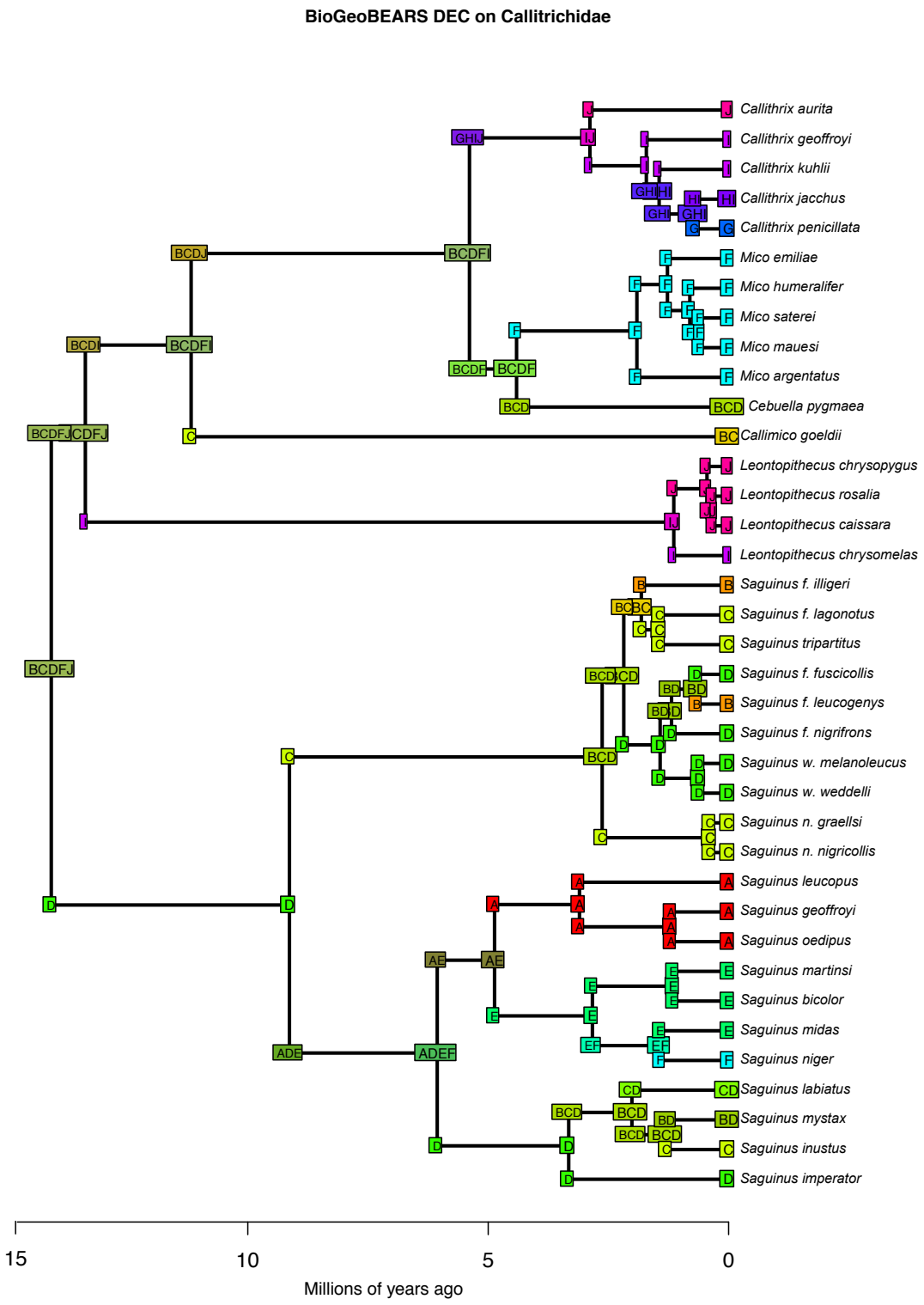


**Figure 1-9 (a).** DEC reconstruction of the biogeographic history of Callitrichidae.

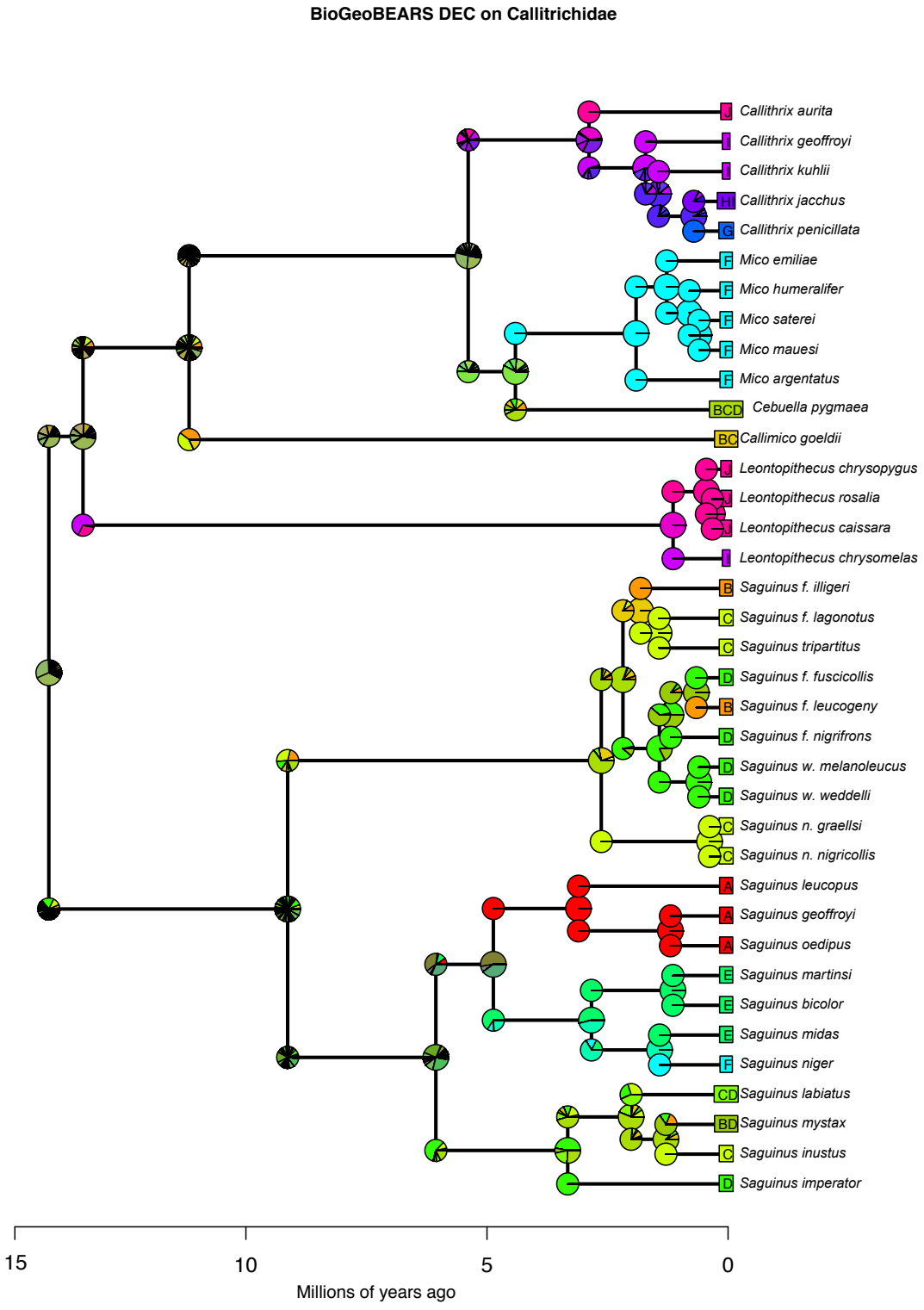


**Figure 1-9 (b).** Support for area configurations from DEC reconstruction of the biogeographic history of Callitrichidae.

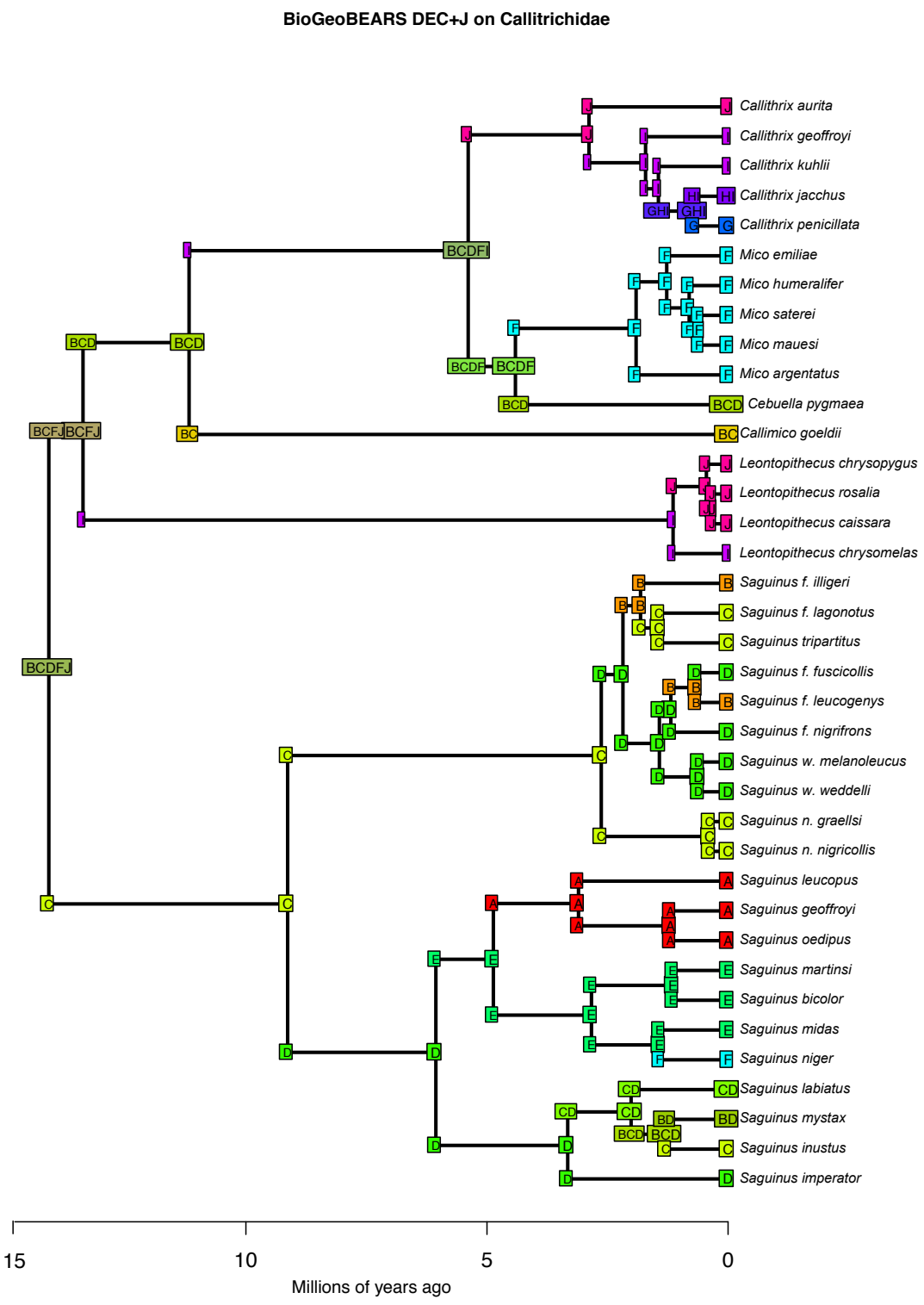


**Figure 1-9 (c).** DEC+j reconstruction of the biogeographic history of Callitrichidae.

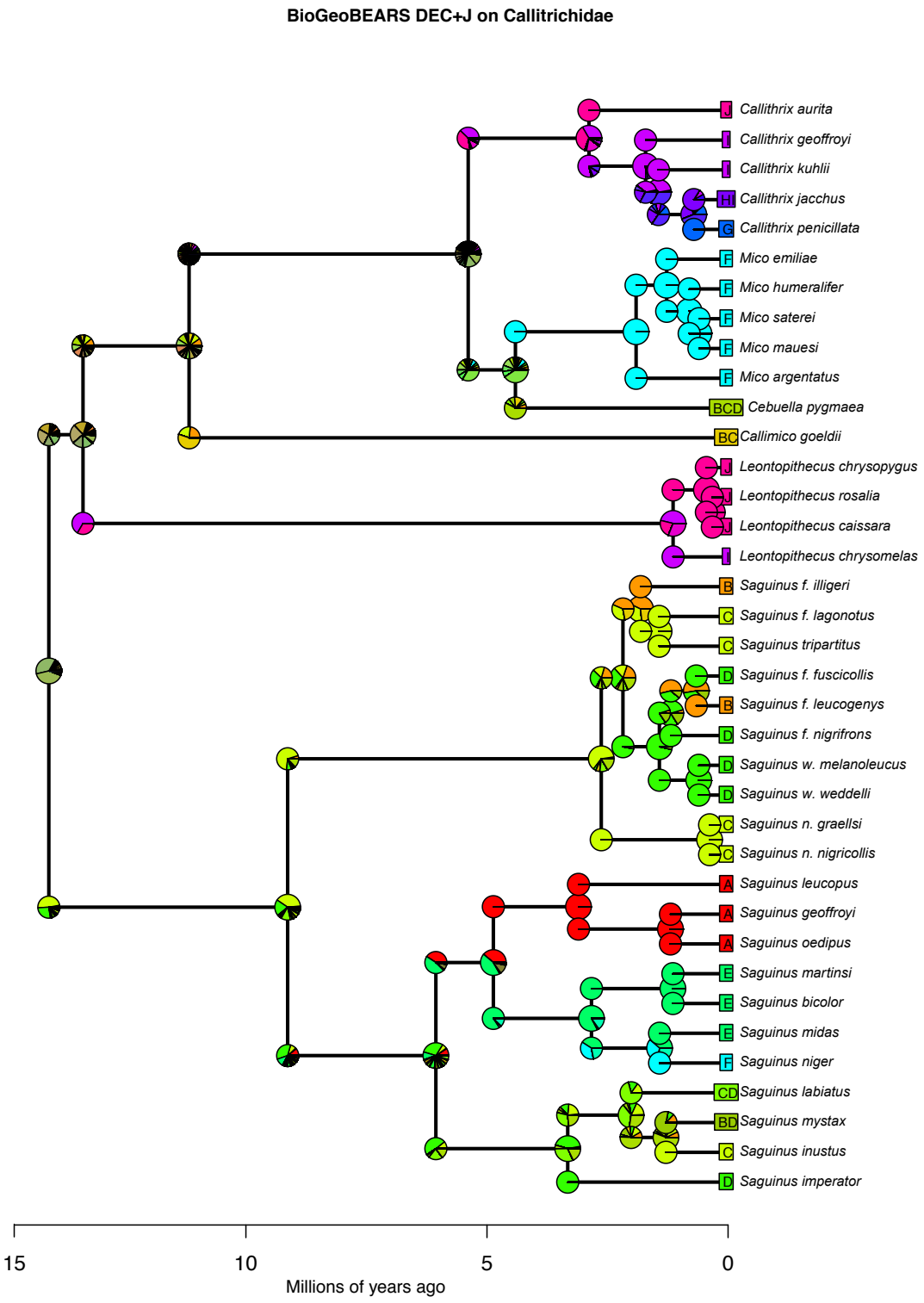


**Figure 1-9 (d).** Support for area configurations from DEC+j reconstruction of the biogeographic history of Callitrichidae.

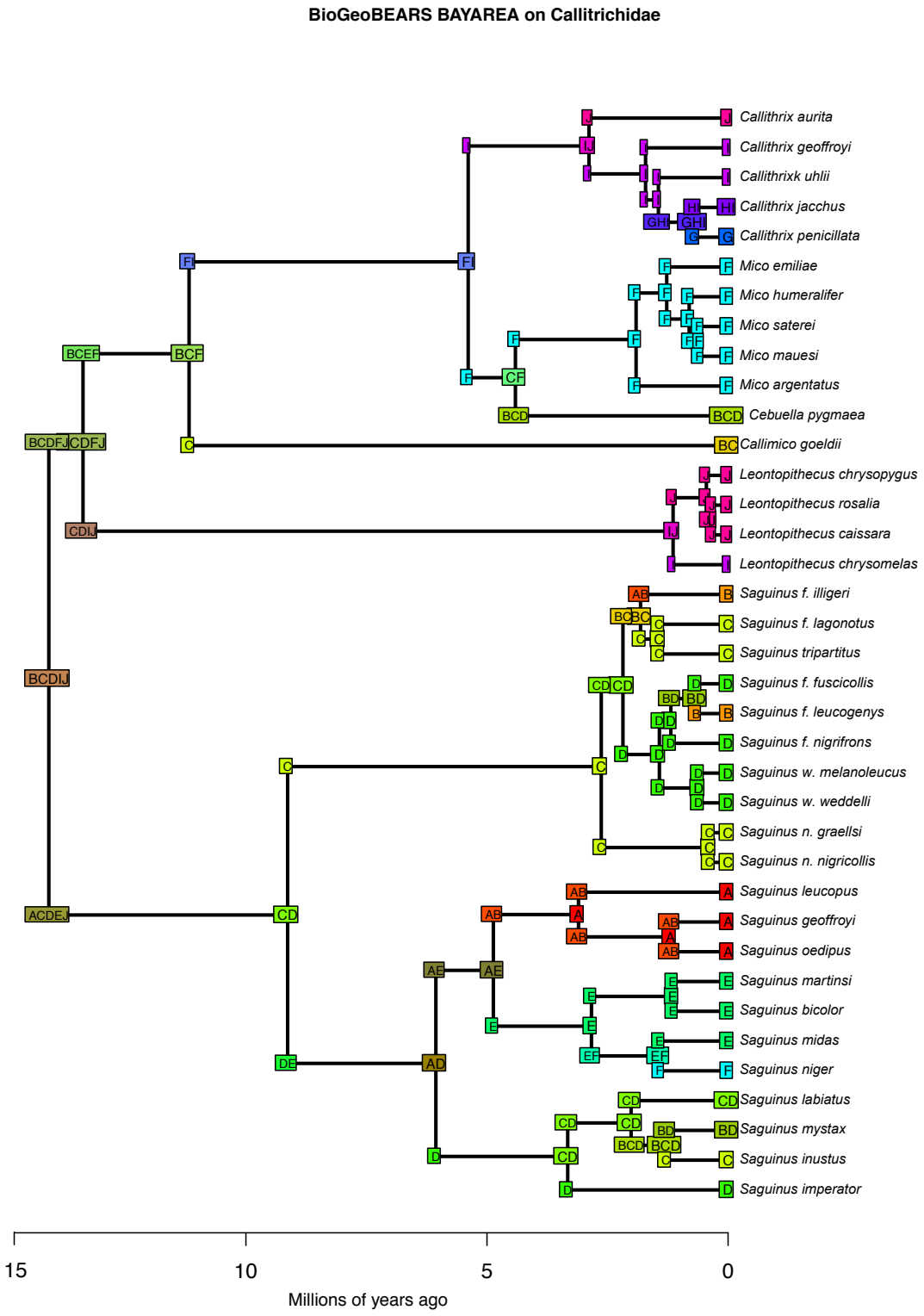


**Figure 1-10 (a).** BAYAREA-Like reconstruction of the biogeographic history of Callitrichidae.

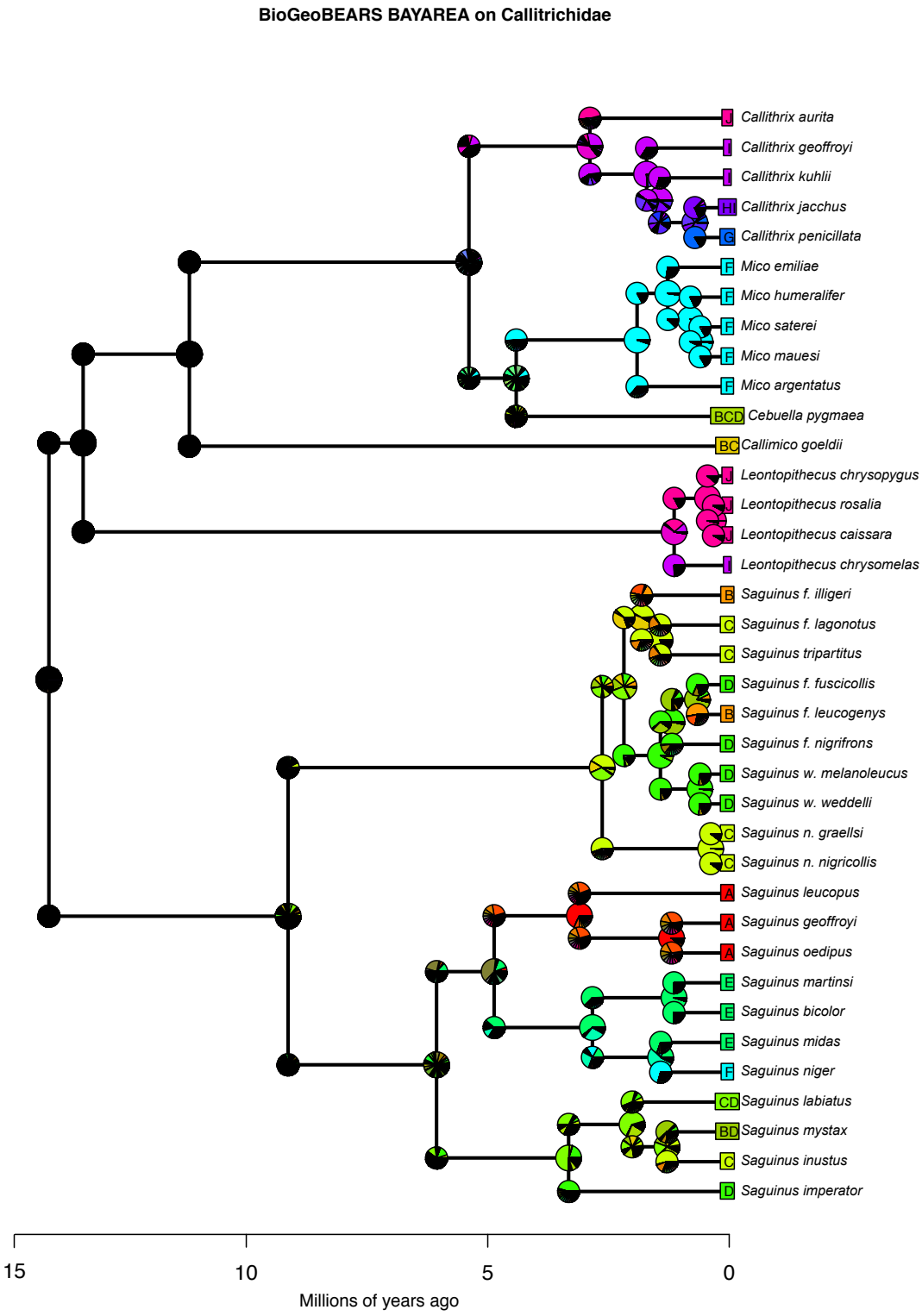


**Figure 1-10 (b).** Support for area configurations from BAYAREA-Like reconstruction of the biogeographic history of Callitrichidae.

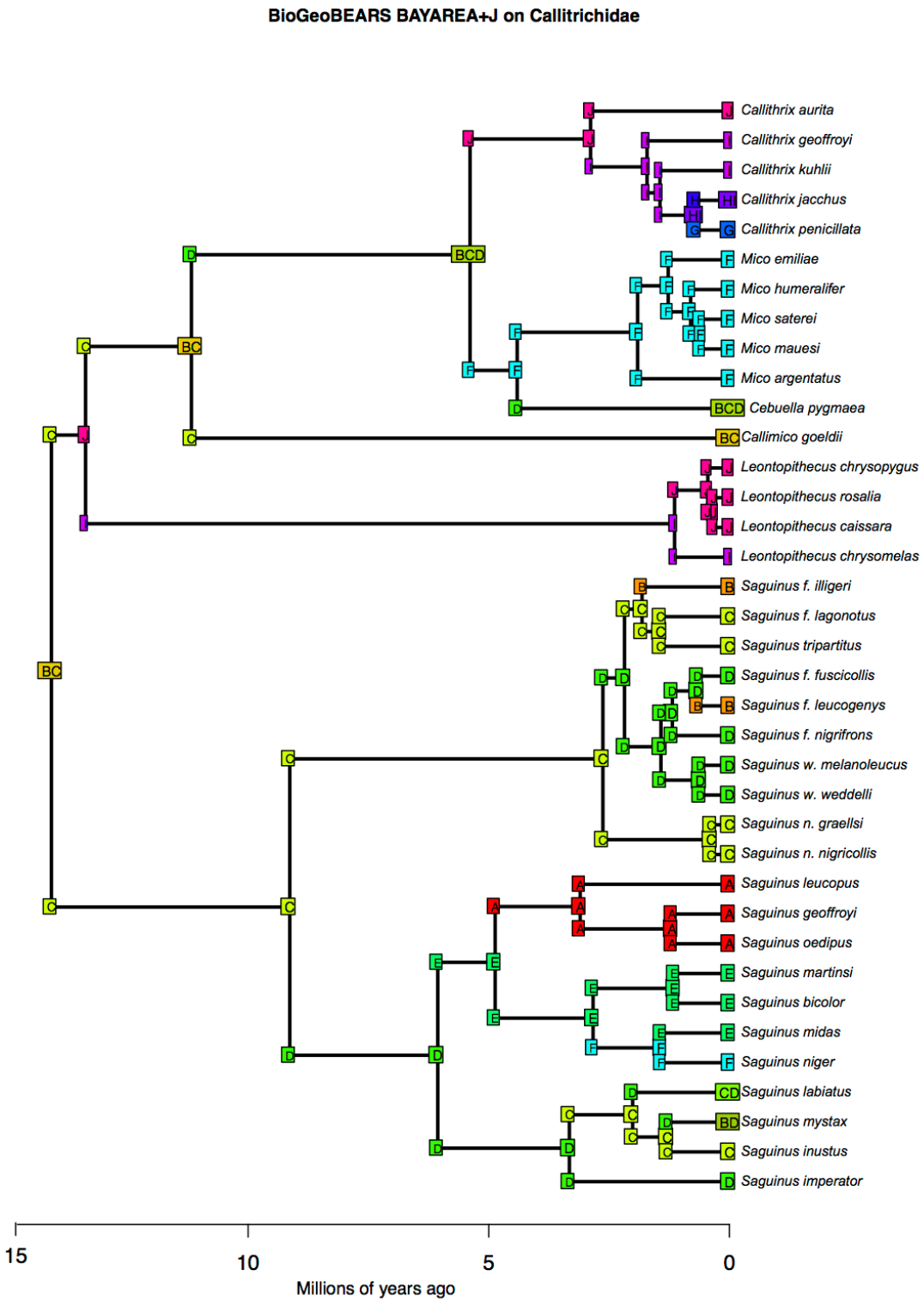


**Figure 1-10 (c).** BAYAREA-Like+j reconstruction of the biogeographic history of Callitrichidae.

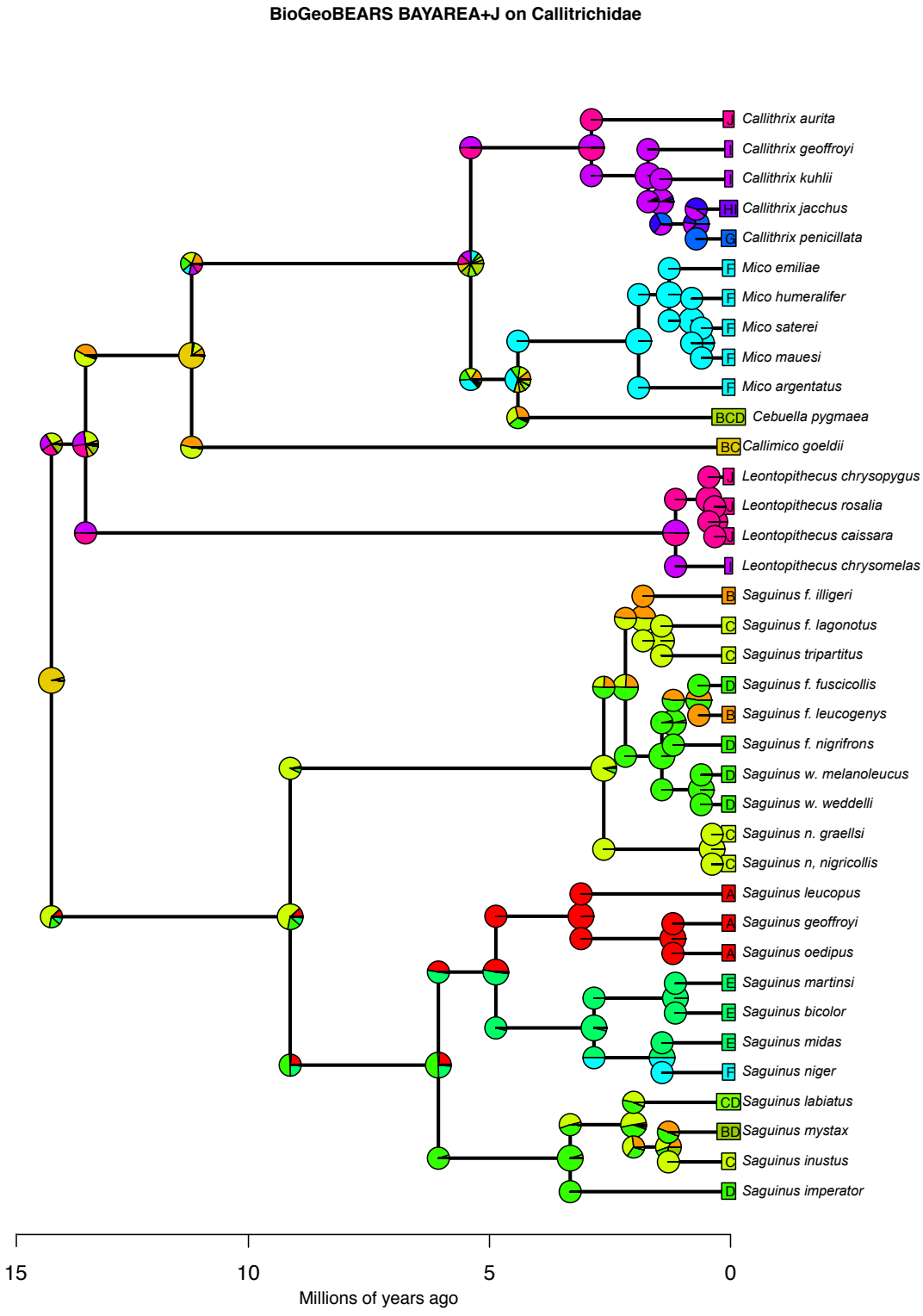


**Figure 1-10 (d).** Support for area configurations from BAYAREA-Like+j reconstruction of the biogeographic history of Callitrichidae.

# Chapter 2

## Variation and selection in MHC Class II DRB exons 2 and 3 for a wild population of white-faced capuchin monkeys (*Cebus capucinus imitator*)

**Janet C. Buckner, Amanda Melin, Katharine Jack, Gustavo Gutierrez-Espeleta, Jessica W. Lynch Alfaro**

**Abstract**

Polymorphism with the major histocompatibility complex (MHC) has been examined in several primate lineages primarily to identify appropriate model systems for humans. The high allelic diversity reported across primates in the MHC class II DRB gene suggests that the region would serve especially well to evaluate life history traits in the context of genetics and immunity. In this first of a two-part study, we use high-throughput sequencing to genotype and characterize polymorphism in two exons (2 and 3) of the DRB region in a wild population *Cebus capucinus imitator*. As in previous studies, we find evidence of multiple gene copies, extensive polymorphism and site-specific positive selection in both exons for this species. With the resulting genotypes, we can begin to assess the role of the MHC in mate-choice for a population of *Cebus capucins imitator* boasting extensive pedigree and behavioral data stemming from a 34-year long-term study.

**Introduction**

The major histocompatibility complex (MHC), or Human Leukocyte Antigen (HLA) in humans, is a vital group of genes involved in the immune response of vertebrates. There are two principal classes of MHC genes: class I molecules bind and present antigens of intracellular origin while class II molecules bind and present antigens of extracellular origin to their corresponding T cells (Janeway Jr. et. al., 2001). The MHC is particularly effective at combating pathogens because it is both highly polygenic and polymorphic - there are many different genes, each with multiple alleles, increasing the diversity of antigen-binding molecules available for recognizing potential threats to the organism (Janeway Jr. et. al., 2001). Studies of MHC genes have shown evidence of positive selection (Hughes and Nei, 1988; Yang and Swanson, 2002), further suggesting that

increased allelic diversity is beneficial to the organism, providing immunity to a wider suite of potential pathogens (Janeway Jr. et. al., 2001). Thus, the degree of allelic diversity and composition in individuals of a population may be specific to parasitic loads or other localized pathogens to which they are exposed (Garamszegi *et. al*., 2011; Borghans et. al., 2007).

Extensive polymorphism in the MHC class II DRB region is reported widely among vertebrates, particularly in comparative studies of humans and other primates (Bontrop et. al., 1999; Doxiadis et. al., 2000; Middleton et. al., 2004; Huchard *et. al.*, 2006; Knapp et. al., 2006; Suárez et al, 2006; O'Connor et. al. 2007). Although diversity in this region has been explored in several non-human primate species, studies of Neotropical primates have been limited by sample size and sequencing technology (Grahovac et. al., 1992; Trtková et. al., 2003; Antunes et. al., 1998; Middleton et. al., 2004). Also, traditionally most MHC studies of non-human primates have relied heavily on captive animals that likely experience a different set of immunological challenges from wild populations. However, recent studies have increasingly explored non-invasive sampling to genotype wild primates (Muller et. al. 2014), with some additionally using high-throughput sequencing techniques (Hans et. al., 2015; Grogan et. al., 2016).

In this study, we explore variation at exon 2 and 3 of DRB genes in 149 white-faced capuchin monkeys (*Cebus capucinus imitator*) using primarily non-invasive sampling for high-throughput MHC genotyping. White-faced capuchins belong to the Neotropical primate family Cebidae. Previous studies of MHC diversity in cebids focused on only a few individuals of *Saimiri sciureus* and *Sapajus apella* of captive origin (Trtková et al, 1993). As a result, MHC diversity across this group of primates is poorly understood. Somewhat more extensive data exists for cotton-top tamarins (*Saguinus oedipus:* Watkins et. al., 1990; Watkins et. al., 1991; Grahovac et. al., 1992; Trtková et. al., 1993; Gyllensten et. al., 1994; Cadavid et. al., 1999;

Kriener et. al., 2001; Middleton et. al., 2004) and common marmosets (*Callithrix jacchus;* Antunes et. al., 1998; Wu et. al., 2000; Shiina et. al., 2011; van der Wiel et. al., 2013; Kono et. al., 2014) of the family Callitrichidae, and night monkeys (*Aotus;* Diaz et. al., 2000; Diaz et. al., 2002; Nino-Vasquez et. al., 2000; Suarez et. al., 2003; Cardenas et. al., 2005; Baquero et. al., 2006; Suarez et. al., 2006) of the family Aotidae.

White-faced capuchins are omnivorous habitat generalists that occupy and forage in all levels of the canopy (Rylands et. al., 2013). They are also extremely social, characterized by large meta-populations with regular male dispersal across groups (Jack and Fedigan, 2004; Jack et. al., 2012). Such wide niche-breadth and high sociality results in increased parasite loads and greater opportunities for contracting pathogens (Parr et al., 2013), making the MHC repertoire vital. Additionally, white-faced capuchins have a complex hierarchical mating system that operates on at least two levels (dispersal-mediated male choice and primarily female mate choice via receptivity), with dominant males siring most of the offspring (Jack and Fedigan, 2006). The population assessed here is also the subject of a long-term study that has resulted in extensive data on relatedness, paternity, reproductive success and dispersal history for most individuals (Fedigan and Jack, 2012).

Although not always supported, some studies of primates, rodents and other vertebrates have suggested that the MHC may function in genetically based mate choice (Schwensow et. al., 2008 - fat-tailed lemur; Fickel and Weyrich, 2011 - rodents; Consuegra et. al., 2008 - Atlantic salmon; Olsson et. al., 2003 - sand lizards), with detection of genotype potentially achieved through odor (Knapp et. al., 2006). Thus, white-faced capuchins may ultimately serve as a good model for understanding how the MHC may provide insight into the susceptibility versus resistance of wild populations to disease as well as the potential role of the MHC in assortative

mating. This report is part one of a two-part study that aims to examine the possible role of the MHC diversity in mate choice among wild white-faced capuchin monkeys. The aims are to: (1) genotype DRB exon 2 and 3 for individuals from a Santa Rosa metapopulation, (2) summarize the observed polymorphism in both exons across this subspecies and (3) evaluate the evolutionary history of these exons within white-faced capuchins and in the larger context of MHC diversity across New World primates (NWP).

## Methods

**Study Site and Sampling** – We obtained fecal and tissue samples from 115 individuals from the Santa Rosa Sector of the Area de Conservacion Guanacaste, Costa Rica (SRP) dating from 2010. The full SNRP metapopulation is estimated at approximately 600 individuals in 38 groups, averaging 17 individuals per group (Fedigan and Jack, 2001). This population has been the subject of a long-term study begun in 1983. We collected blood samples from an additional 33 individuals from Costa Rican populations in San Ramón, Manuel Antonio, Damas, Liberia, San Mateo and Palo Verde. Fecal samples were preserved at 4ºC and blood samples were preserved at -80ºC until extraction.

**DNA extraction, Primer Design and Amplification** – Genomic DNA was extracted from fecal, blood and tissue samples using the QIAmp DNA Stool Kit (Qiagen, Valencia, CA) and the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), respectively, following the manufacturer's protocol with modifications from Di Fiore et al. (2015). Resulting extracts were quantified using a Qubit dsDNA BR Assay (Life Technologies, Carlsbad, CA). 5-10 uL of the resulting extractions were included in 25uL PCR reactions that included the following reagents: 12.5uL of PCR Master Mix (Promega, Madison, WI), 0.8uL of the forward primer (10uM), 0.8uL of the reverse primer (10uM), 0.8uL BSA and finally PCR-grade water to bring the final

volume to 25uL. Primers were designed from aligned sequences of published New World primate sequences (Table 1). PCR cycling conditions for all exons were as follows: (1) An initial denaturing at 94ºC for 2 minutes followed by (2) 35 cycles of: 94ºC for 30 seconds, 58ºC for 40 seconds and 72ºC for 1 minute followed by (3) a final extension at 72ºC for 7 minutes. The PCR product was visualized on a 1.5% agarose gel to confirm the presence of fragments of the expected size.

**Library preparation and sequencing** – We cleaned the resulting amplicons to remove primer dimers and fragments less than 100bp by binding DNA to 20uL of Serapure beads (Rohland and Reich, 2012) and performing ethanol washes. DNA was subsequently eluted from the beads with 50uL of 10mM Tris pH 8.5. Fragments were again visualized on a gel to ensure removal of small fragments. Following PCR cleaning, we attached unique dual index combinations to DNA from each individual using the Nextera XT indexing kit (Illumina, San Diego, CA), following the manufacturer's protocol. We then cleaned the indexed amplicons using Serapure as before. These clean indexed amplicons were then quantified using a Qubit dsDNA BR Assay (Life Technologies, Carlsbad, CA), visualized using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA) and then pooled in equimolar ratios for 2x250 sequencing on an Illumina MiSeq machine (Illumina, San Diego, CA) at the Technology Center for Genomics and Bioinformatics (University of California, Los Angeles).

**Post-Sequencing Quality Control and Genotyping** – We largely followed the quality control steps for MHC loci outlined in Grogan et al. (2016). We began by separating sequences of exons 2 and 3 based on their primer sequences. We then discarded any reads with primer sequences that differed from the original sequences by more than one base pair (bp). For the remaining reads, primer sequences were removed. We then merged the forward and reverse reads in

Geneious Pro 9.0.5 (Kearse et al., 2012) to get the full-length sequence. Only reads that were successfully paired with their partners were used in subsequent steps. Using Galaxy, we filtered out any reads shorter than 150bp or longer than 400bp and any reads where more than 5% of bp had a Phred score less than 20. The remaining reads were clustered into unique sequences using Geneious Pro 9.0.5 and singleton variants were discarded. As the number of MHC-DRB copies is unknown for white-faced capuchins, we assumed an average amplification efficiency of 0.7 and one gene copy (a maximum of 2 alleles) per individual while targeting at least three reads per allele (Galan et. al., 2010; Sommer et. al., 2013). Therefore, individuals with less than 25 reads at the end of this process were discarded as having too few reads to genotype (Sommer et. al., 2013).

We concatenated all reads, labeled with the individual's ID, into a single fasta file and aligned them using MAFFT 7.130b (Katoch, 2013). Subsequently, we separated reads by individual and corrected insertion/deletion homopolymer errors. We removed any variants that remained with incorrect reading frames, that is variants with lengths that differed from 254 (exon 2) or 217 (exon 3) by any value that was not a multiple of 3.

For each individual, we sorted variants by their relative frequency. The most frequent variant was automatically considered a putative allele. Among the remaining variants, we discarded any that did not represent more than 1% of the total proportion of reads for the individual. We then verified that the remaining sequences were not chimeras using the *uchime2_denovo* command in UCHIME. We discarded putative chimeras only if found to be a chimera in all individuals that the sequence occurred. The remaining variants were then sorted into the following two categories: (1) those with 1-2bp differences from any more frequent variant; (2) those with more than 2bp differences from any more frequent variant.

For the five individuals with replicate amplicons for both exons, if the variants were in both replicates, and were more frequent than any putative artifacts, they were classified as alleles, regardless of the number of bp differences. If they were less frequent than putative artifacts, they were marked unclassified variants. We discarded variants as artifacts if they were 1-2 bp different but were not present in both replicates. If the variant was more than 2bp different, but not present in both replicates, we discarded them as an artifact if they did not occur in any other individual.

For the remaining individuals without replicates, if any variants of that individual were not present as an allele in at least one other individual, it was discarded as an artifact. If the variant was found in another individual to be an allele, and differed by 1-2bp from more frequent variants, then it was considered an allele if it was more frequent than any artifact. If it was less frequent than any artifact, it was marked an unclassified variant. If the variant was found in another individual to be an allele or unclassified variant, and differed by more than 2bp from more frequent variants, then it was considered an allele. If it was not found in another individual as an allele or unclassified variant, then it was marked an unclassified variant. After the initial genotyping, most individuals had more than two alleles (see Results for averages) so that the threshold for minimum number of reads to genotype an individual under the aforementioned conditions increased to 120 (Sommer et. al., 2013). As a result, three individuals were dropped from the exon 2 dataset, but there was no change in the genotyping result.

At the end of sequence classification on an individual basis, variants had to be found in at least 2 individuals of *Cebus capucinus imitator* to be included in the final list of alleles, unless it was the most frequent variant in an individual. For high coverage individuals (>1000 reads), any putative allele that is present in only two individuals with absolutely no artifacts are also

classified as artifacts. If a sequence was found in only two individuals, and in only one was found to be a chimera, these were labeled unclassified variants in the chimera sample. These sequences were not included in the final list of alleles for *C. c. imitator*. Similarly, if a sequence was found in only two individuals, was an allele in one but an unclassified variant in the second, these sequences were not included in the final list of alleles for *C.c. imitator*.

**Selection and Evolution in DRB exons -** We used PAML (Yang et. al., 2007) to obtain an average dN/dS ratio (w) – the ratio of non-synonymous to synonymous nucleotide changes – for each exon. A single w shared across all sites in a protein coding gene is biologically unlikely, we therefore additionally performed likelihood ratio tests (LRT) between two sets of nested models that test for neutrality versus selection at sites while incorporating transition/transversion rates and codon usage bias.

To get a sense of the evolutionary history of these exons in the context of New World Primates (NWP), we reconstructed phylogenetic trees. We made alignments for each exon from our recovered alleles and from published DRB sequences obtained from Genbank. We used neighbor-joining in MEGA 7 with the Kimura 2-parameter model to reconstruct evolutionary relationships among sequences (Kumar et. al., 2016). When information was available to distinguish pseudogenes from published data, these were excluded from the analysis.

## Results

**Genotyping** – We amplified DRB exon 2 and 3 in 144 *C. c. imitator* individuals. We successfully genotyped exon 2 in 140 individuals, 108 of these were from SRP. We successfully genotyped exon 3 in 143 individuals, 110 of these were from SRP. We were unable to genotype individuals with less than 120 reads – 4 individuals for exon 2 and one individual for exon 3. We successfully produced replicate amplicons for both exons in five individuals.

The MiSeq produced 2,637,162 reads with DRB2 primers for *C. capucinus imitator*. All sequences retained after filtering and genotyping were 254 bp long (82 codons). Among the 1,188,308 (45.06%) reads that ultimately contributed to putative alleles, the mean number of read pairs per individual was 4,243 (min=129, max=40,239; Fig. 1a). The mean number of putative alleles per individual was 7 (min=2, max=15; Fig. 2a). For the Santa Rosa individuals, which were mostly genotyped from fecal samples, the mean number of read pairs per individual was lower at 2,704 (Fig. 1a). However, the mean number of putative alleles per individual was 7, as among all *C. c imitator* (max=14; Fig. 2a). We found that the number of alleles recovered and the number of reads obtained per individual were uncorrelated (Fig. 4a; Slope = 0.00004; $R^2$=0.00638). A linear regression analysis resulted in a p-value of 0.3480, confirming that the number of reads is not a reliable indicator of the number of alleles per individual.

The MiSeq produced 1,991,254 reads with DRB3 primers for *C. capucinus imitator*. All sequences retained after filtering were 214 or 217 bp long (71 or 72 codons). Among the 1,207,198 (60.63%) reads that ultimately contributed to recovered alleles in *Cebus capucinus imitator*, the mean number of read pairs per individual was 4,220 (min=151, max=40,269; Fig. 1b). The mean number of putative alleles per individual was 14 (min=1, max=23; Fig. 2b). Among only Santa Rosa individuals, the mean number of read pairs per individual was slightly higher 5,218 (Fig. 2b). However, the mean, minimum and maximum number of putative alleles per individual remained the same as for all *C. c. imitator*. We found that the number of alleles recovered and the number of reads obtained per individual were uncorrelated (Fig. 3b; Slope = 0.0001; $R^2$=0.01085). A linear regression analysis resulted in a p-value of 0.2148, confirming that the number of reads is not a reliable indicator of the number of alleles per individual.

**Polymorphism** - Among *Cebus capucinus imitator*, we recovered 41 unique DRB exon 2 sequences, the rarest of these occurring in two individuals (Fig. S1). All DNA sequences produced unique amino acid sequences (Fig. 4a; S2). Within the Santa Rosa population, we recovered 29 unique sequences, the rarest of these occurring in one Santa Rosa individual but found in *C. c. imitator* from another locality. Based on our sampling, we found the following 8 alleles to be unique to the Santa Rosa population: 16, 17, 18, 30, 33, 56, 63 and 78. The number of individuals with a particular exon 2 allele varied greatly across sampled *C. c. imitator* (Fig. 5a).

Among *Cebus capucinus imitator*, we recovered 69 unique DRB exon 3 sequences, the rarest of these occurring in one individual whose most frequent sequence was unique among the population (Fig. S3). These DNA sequences produced only 46 unique amino acid sequences (Fig. 4b; S4). Within the Santa Rosa population, we recovered 52 unique sequences, the rarest of these occurring in one Santa Rosa individual but found in *C. c. imitator* from another locality. Based on our sampling, we found the following 16 alleles to be unique to the Santa Rosa population: 21, 22, 25, 34, 37, 42, 43, 44, 45, 47, 52, 53, 76, 77, 84 and the one unique allele. The number of individuals with a particular exon 3 allele varied greatly across sampled *C. c. imitator* (Fig. 5b).

**Sequence Evolution** - Using the one-ratio model (M0) in CODEML, we obtained an average dN/dS (w) value of 0.73 and 0.88 for exon 2 and 3, respectively. These values suggest that the exons are under negative selection. However, likelihood ratio tests confirm that models that allow for sites under positive selection (M2a and M8) fit the data better than models that disallow positive selection for both exons. Table 2 lists putative positively selected sites (PSS) in

each exon, with Bayes Empirical Bayes posterior probability of >0.95 calculated in CODEML (Yang, 2005).

Figures 6 and 7 summarize the reconstructed evolutionary relationships for exons 2 and 3, respectively, among NWP. The topologies show seven and six groups of *Cebus capucinus imitator* sequences in DRB exon 2 and 3, respectively. These are highlighted in the phylogenetic trees. However, most groups are not supported by bootstrap values above 50.

**Discussion**

In this study, we genotype and characterize polymorphism in exon 2 and 3 of the MHC-DRB region for *Cebus capucinus imitator*, a species for which no MHC data previously existed. By using high-throughput sequencing, we successfully genotyped > 97% of our individuals with high sequencing depth. Although the sequences produced originated from DNA sequencing rather than RNA sequencing, the lack of stop codons in recovered alleles suggests functionality. BLAST searches of DRB exon 2 sequences 1, 2 and 3 produced complete matches to the published genome for the species (XM_017522743, XM_017522742, 017522746, respectively). Similar BLAST searches of DRB exon 3 produced complete matches for sequences 4 and 5 (XM_017522743 and XM_017522746, respectively).

Evidence of greater allelic diversity and substitution rates in Malagasy and New World primates relative to Old World primates was demonstrated by Garamszegi and colleagues (2011). The sampling in our study is largely restricted to a single metapopulation of *C. c. imitator*, and a more comprehensive sampling may prove that the number of alleles reported here is an underestimation of the full diversity of DRB alleles. This possibility is suggested by the increase in the number of alleles for DRB exon 2 from 29 to 41 when individuals outside SRP are included. On the other hand, we were unable to produce replicates for every amplicon

examined, so it is possible that our estimation of unique alleles for these exons is slightly inflated. However, the number of alleles (41) recovered from 140 individuals for DRB exon 2, possibly the most widely studied exon across MHC Class II loci, falls within the range of those recovered in similar studies (Table 3). The most closely related species in the Immuno Polymorphism Database for MHC in Non-human primate database (IPD-MHC NHP) is *Sapajus apella* for which there are 6 DRB alleles reported (Trtková et. al., 1993). These sequences were obtained from analysis of just two individuals. There are fewer studies specifically examining exon 3 in NWP making direct comparison more difficult: *S. oedipus* (6 indiv.), *C. jacchus* (1 indiv.), *S. sciureu*s (1 indiv.), *A. trivirgatus* (1 indiv.), *S. apella* (2 indiv.) and *C. moloch* (2 indiv.) - 8, 1, 4, 1, 1 and 4 sequences respectively. However, the amount of diversity we find for this locus exceeds that for exon 2. Therefore, more studies that include this locus, in addition to exon 2, can give us a better sense of the nature of the diversity in this gene region.

As of yet it is difficult to determine why the number of sequences sometimes differ greatly among species. This could be explained by differences in sequencing techniques or differences in life history among species. In the case of NWP, previous studies have largely used more traditional methods (cloning, etc.) and small sample sizes when characterizing MHC polymorphisms possibly explaining the typically much smaller numbers of sequences reported. However, life history of capuchin monkeys may also select for higher diversity of MHC alleles because of their greater potential exposure to a range of pathogens. A re-evaluation of traditionally examined species with more recent methods and greater sample sizes may help to determine whether such differences are due to methods or inherent within the species.

As in other studies of selection in MHC genes, we find evidence of positively selected sites in both exons (Grogan et. al., 2016). This finding corroborates previous suggestions that

species benefit from a diverse repertoire of MHC alleles which can confer immunity to a greater suite of potential pathogens. A better understanding of the protein structure resulting from the amino acid sequence is needed to understand the role of these sites and why they might be under positive selection. If future studies can establish such roles, homologous sites across species that also show such selection patterns may reveal readily evolving regions in MHC proteins that allow for improved immune system function. In addition to analyses of selection, studies toward this end will continue to benefit from understanding the evolutionary history of these genes within and across species.

Patterns of clustering in our phylogenetic trees (Figs. 6 and 7) seem to support a shared ancestor hypothesis for NWP (though direct comparison to Old World monkey sequences are needed to confirm this). Although there are large clusters of sequences that aggregate based on species, species are not all monophyletic in the tree, suggesting shared ancestry in gene lineages. It is important, however, to note that basal nodes are virtually unsupported in the tree so that the evolutionary history of this gene region remains unclear. Such trans-species polymorphism, where nearly identical sequences occur across species suggests that selection may have maintained these alleles throughout the NWP radiation (Grogan et. al., 2016). Additionally, previous studies of DRB gene phylogeny in primates also reported evidence that the genes in the Platyrrhini lineage originated from a common ancestor following the group's split from the Catarrhini (Grahovac et. al.,1992). These patterns together suggest that DRB alleles evolved after the arrival of the most recent common ancestor of NWP to the western hemisphere. Alternatively, sequence similarities could be explained rapid mutation rates or selection toward particular amino acids that confer increased defense against pathogens that threaten multiple species.

## Conclusion

In this study, we genotyped two DRB Class II exons in 144 wild white-faced capuchin monkeys. We found evidence of extensive polymorphism, particularly at exon 3. Thus, we provide more data demonstrating the diversity at exon 2 which is reportedly common across primate lineages. The extent of diversity at exon 3 is less thoroughly explored in other NWP but our results suggest that future studies would benefit from including this exon or alternatively sequencing the entire gene region when possible. As in other studies employing high-throughput sequencing technology and large sample sizes, our results suggest that such methods have greater power to increase the recovery of alleles in MHC loci over traditional methods. However, when possible all individuals should have at least two replicate sequences to avoid inflated diversity.

Although sequences cluster largely by species in the phylogenetic trees of both exons, species are not fully monophyletic, suggesting shared gene history across species or convergent evolution of sequences. We also find evidence of positively selected sites in both exons, suggesting that MHC diversity is important in white-faced capuchins for immunity, and possibly fecundity via sexual selection. We intend to now use our genotype information to test such hypotheses of MHC-mediated mating in this Santa Rosa population leveraging the pedigree and dispersal data gathered over the last 34 years. Such studies will be important in wild primates because they represent a natural state: mate choice and pathogen exposure are not limited by laboratory or captive colony conditions that potentially offer artificially increased or decreased pathogenic risks and limited copulatory opportunities. For example, mating in wild populations are impacted by migration events which regularly introduce new potential mates. Thus, the data from this study now opens the opportunity to study mate choice in a genetic context for a species with a particularly complex social structure.

**Table 2-1**. DRB primer sequences are given along with the references used to design them.

| Exon | Fwd Primer | Rev Primer | Exon Length | Reference |
|---|---|---|---|---|
| 2 | CGGATCGTTCG-TGTYCCCACAG | CTCTCCGCTGC-ACTGTGAAGCT | 248 | *Cebus capucinus imitator* genome |
| 3 | TGACTGTGTATCCT-GCMAAGACCCAG | ATTGCACTGTGR-KAGGGCTCRTCA | 217 | *Cebus capucinus imitator* genome |

**Table 2-2.** PAML CODEML Results. The critical value for the LRT comes from the $X^2$ distribution whose value for 2 degrees of freedom is 5.99 for p=0.05. $p_x$ equates to the fraction of sites evolving at the corresponding dN/dS ($w_x$). k is the transition/transversion ratio of nucleotide substitutions. The parameters p and q are the parameters of the beta distribution employed in models M7 and M8 (Yang, 2006).

| DRB exon 2 | | | | |
|---|---|---|---|---|
| Model | -LnL | LRT | Parameters | PSS |
| M1a | 1632.23 | 140.53 | K=0.75; $p_0$=0.77, $w_0$=0.03 | N/A |
| M2a | 1561.96 | | K=0.91; $p_0$=0.69, $w_0$=0.03; p1=0.24, $w_1$=1; $p_2$=0.07, $w_2$=7.32 | 4,6,52,65,81 |
| M7 | 1627.25 | 135.12 | K=0.75; p =0.013, q=0.043 | N/A |
| M8 | 1559.70 | | K=0.89; $p_0$=0.93, $p_1$=0.07, p=0.02, q=0.05; w=6.77 | 4,6,52,65,81 |

| DRB exon 3 | | | | |
|---|---|---|---|---|
| Model | -LnL | LRT | Parameters | PSS |
| M1a | 1266.80 | 123.28 | k=6.68; p0=0.83, w0=0.09 | N/A |
| M2a | 1205.15 | | K=8.29; p0=0.77, w0=0.15, p1=0.17, w1=1; p2=0.07, w2=9.02 | 1,6,57,65 |
| M7 | 1275.38 | 137.24 | K=6.83; p0=0.18, q=0.43 | N/A |
| M8 | 1206.76 | | K=8.34; p0=0.93, p1=0.07, p=0.42, q=0.86, w=9.27 | 1,6,57,65 |

PSS = Positively Selected Sites with Bayes Empirical Bayes posterior probability of >0.95.
N/A = Not Applicable

**Table 2-3.** Reported DRB exon 2 variation in NWP and *Lemur catta*.

| Species | Reference | No. Sequences Reported | No. Individuals Sampled |
|---|---|---|---|
| *Aotus nancymaae* | Suarez et al 2016 | 67 | 71 |
| | Nino-Vasquez et al 2000 | 34 | 15 |
| *A. nigriceps* | Suarez et al 2006 | 30 | 15 |
| *A. vociferans* | Suarez et al 2006 | 13 | 10 |
| *A. trivirgatus* | Trtková et al 1993 | 6 | 1 |
| *Saguinus oedipus* | Trtková et al 1993 | 16 | 6 |
| *Callithrix jacchus* | Trtková et al 1993 | 6 | 2 |
| | Antunes et al 1998 | 21 | 25 |
| *Saimiri sciureus* | Trtková et al 1993 | 11 | 3 |
| *Sapajus apella* | Trtková et al 1993 | 5 | 2 |
| *Callicebus moloch* | Trtková et al 1993 | 13 | 2 |
| *Lemur catta* | Grogan et al 2016 | 55 | 302 |
| *Cebus capucinus imitator* | This Study | 41 | 140 |

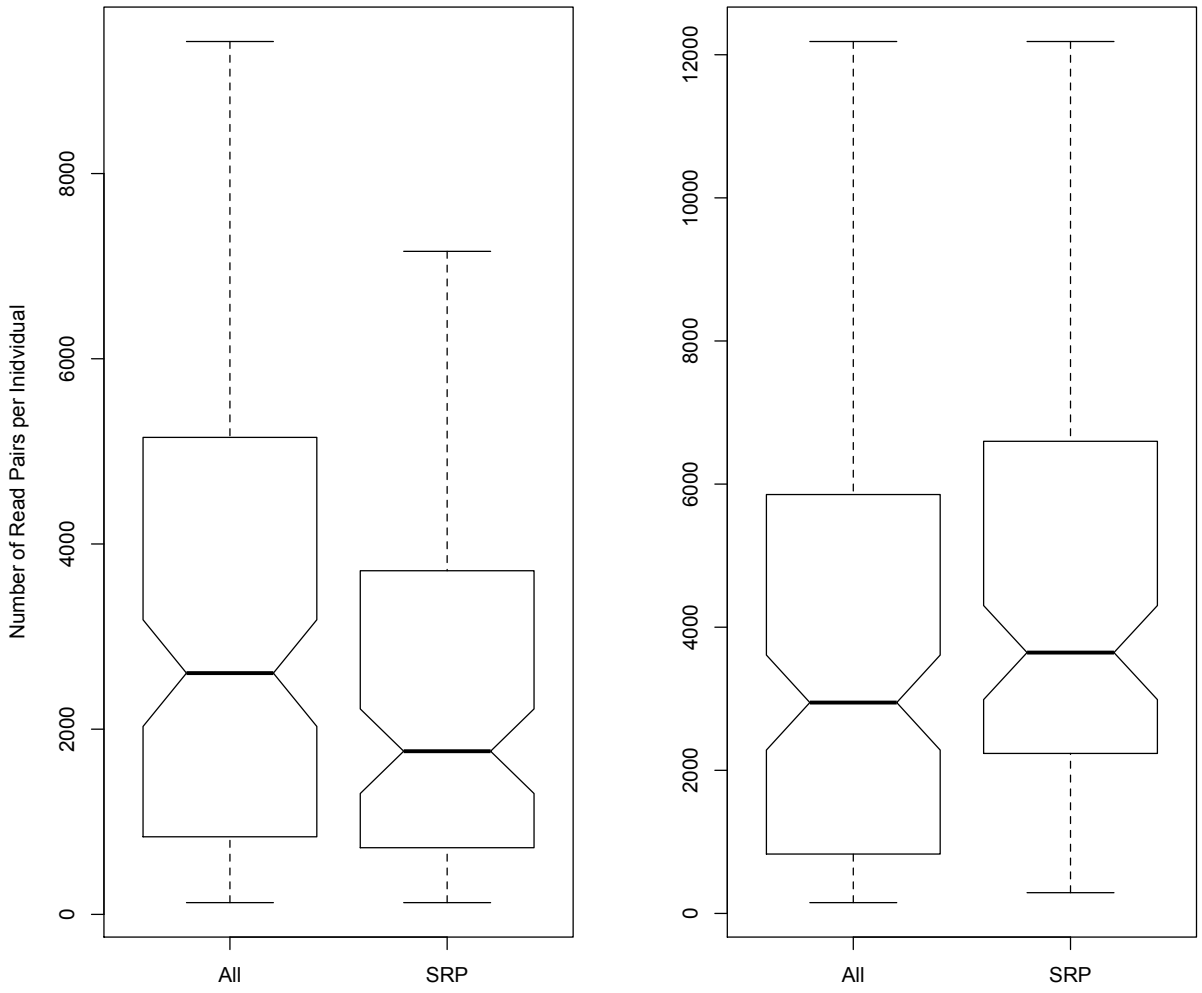


**Figure 2-1.** Box and whisker plots summarizing the distribution of read pairs per individual in all *Cebus capucinus imitator* individuals, and in Santa Rosa (SRP) individuals only, for DRB exon 2 (A) and exon 3 (B). Extreme outliers with values greater than 12,000 are not shown. Overlap across notches in plots indicates that medians are not significantly different at the 0.05 threshold (Chambers *et. al.*, 1983).

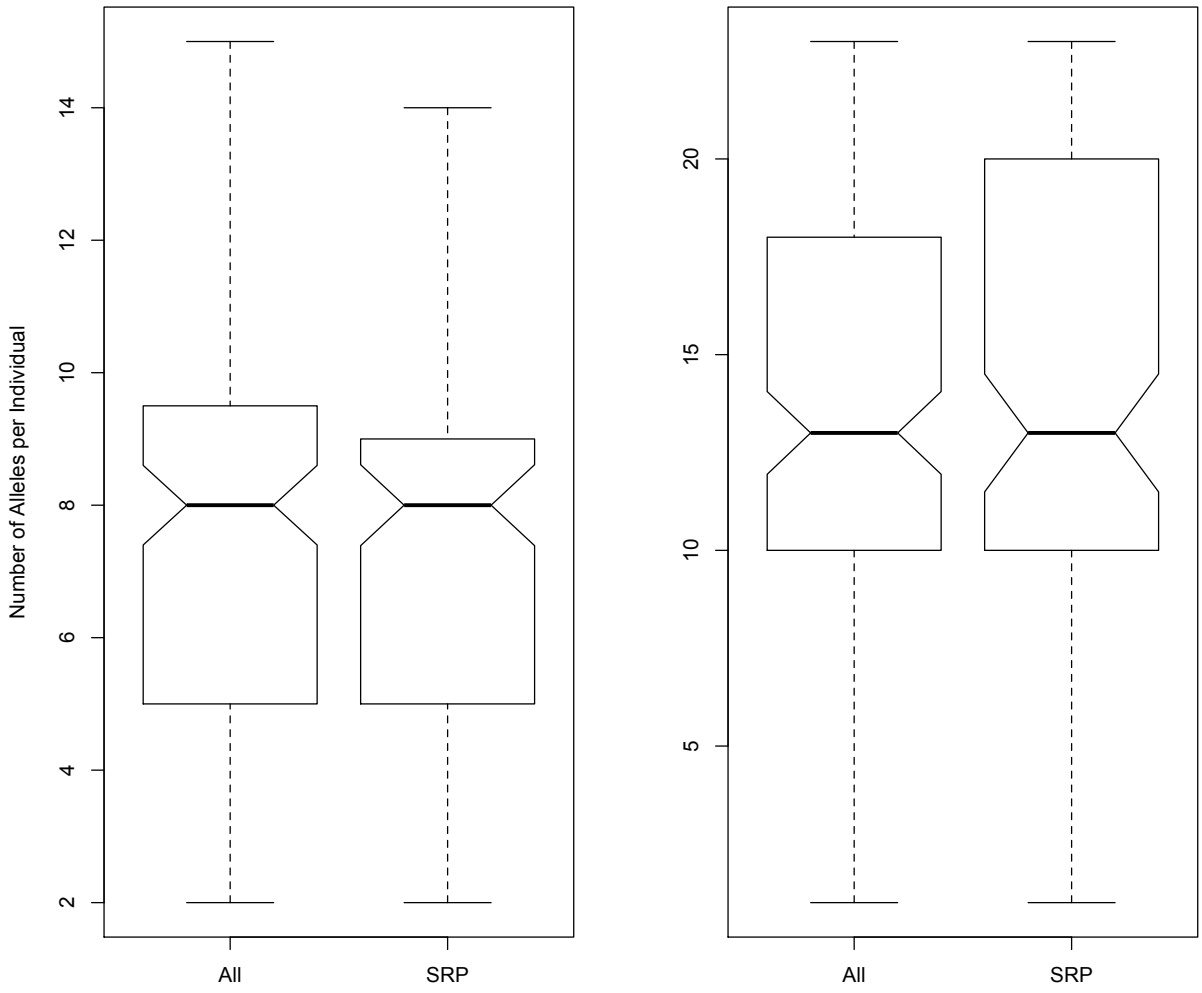


**Figure 2-2.** Box and whisker plots summarizing the distribution of putative alleles per individual in all *Cebus capucinus imitator* individuals, and in Santa Rosa (SRP) individuals only, for DRB exon 2 (A) and exon 3 (B). Overlap across notches in plots indicates that medians are not significantly different at 0.05 threshold (Chambers *et. al.*, 1983).

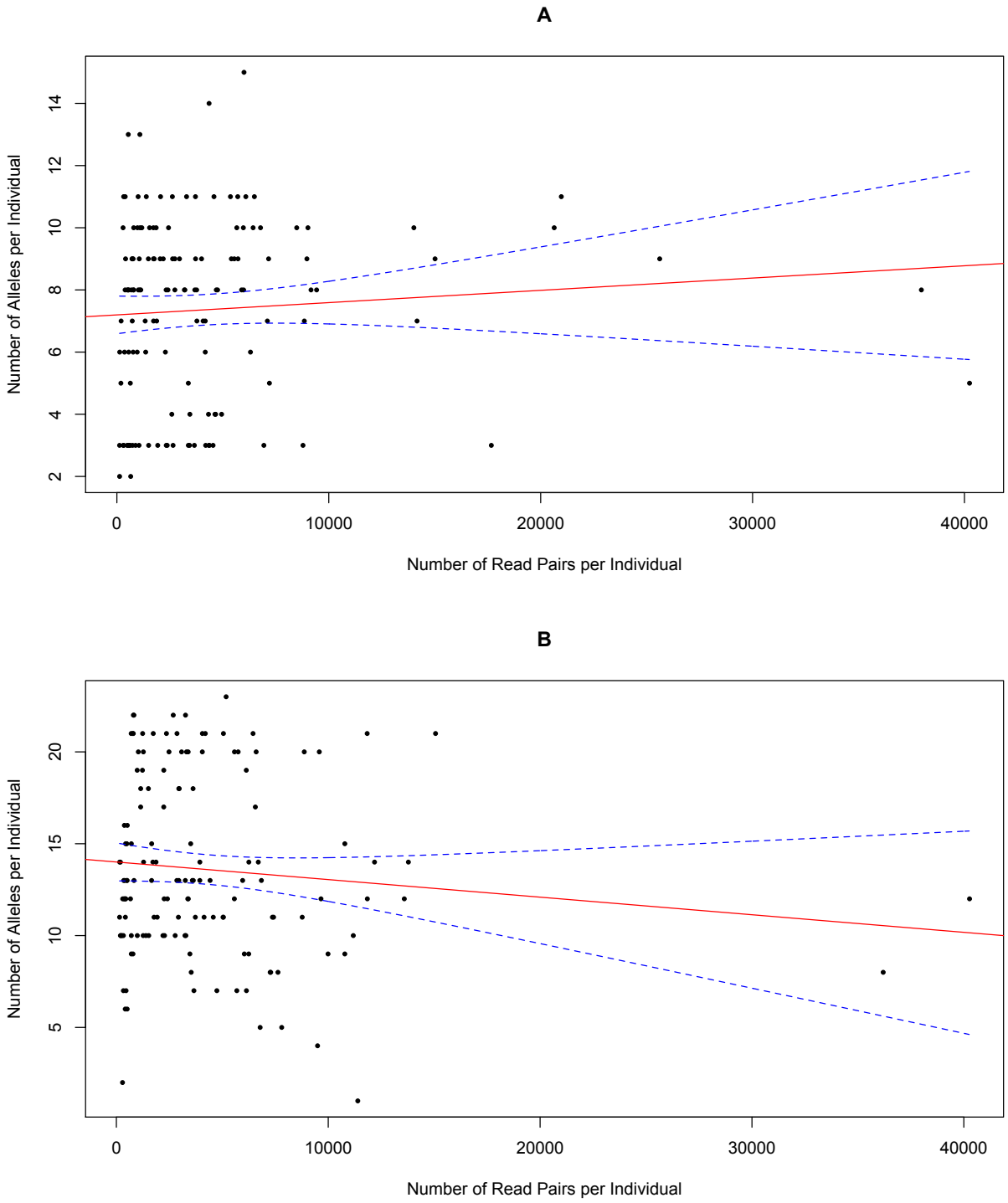


**Figure 2-3.** Scatter plots illustrating the lack of correlation between the number of reads and the number of putative alleles per individual in DRB exon 2 (A) and exon 3 (B). Red lines show the best-fit trend line and blue dotted lines indicate the 95% confidence interval.

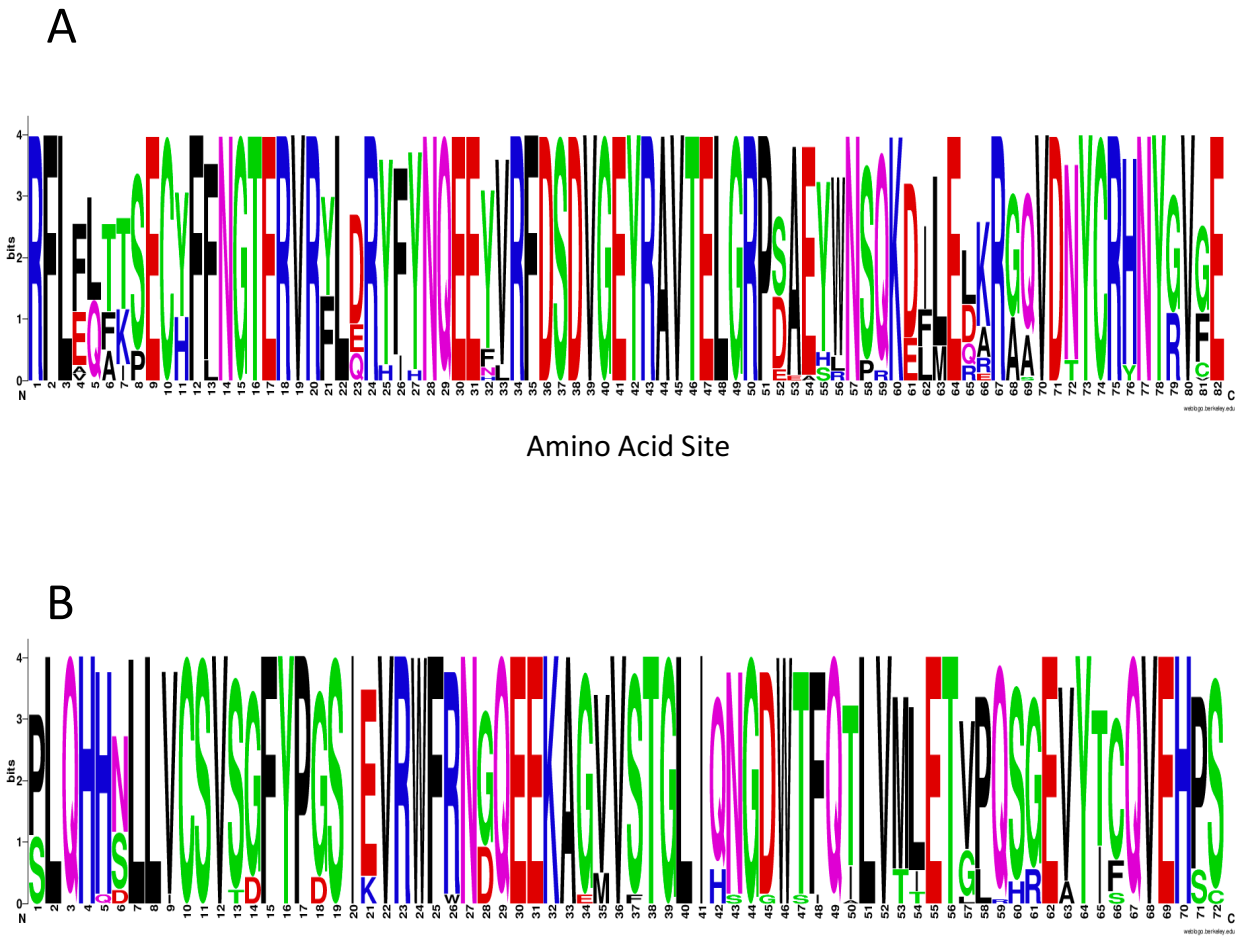


**Figure 2-4.** Sequence logos for DRB exon 2 (A) and exon 3 (B) summarizing their amino acid variation using standard single letter amino acid abbreviations. The height of each amino acid is proportional to its relative frequency across recovered sequences at the corresponding site. Sites under positive selection, as indicated in Table 2, are highlighted in gray boxes. The colors of the amino acid represent their chemical properties: Green – Polar; Blue – Basic; Red – Acidic; Black – Hydrophobic. Graphics created by WebLogo (Crooks et. al., 2004).

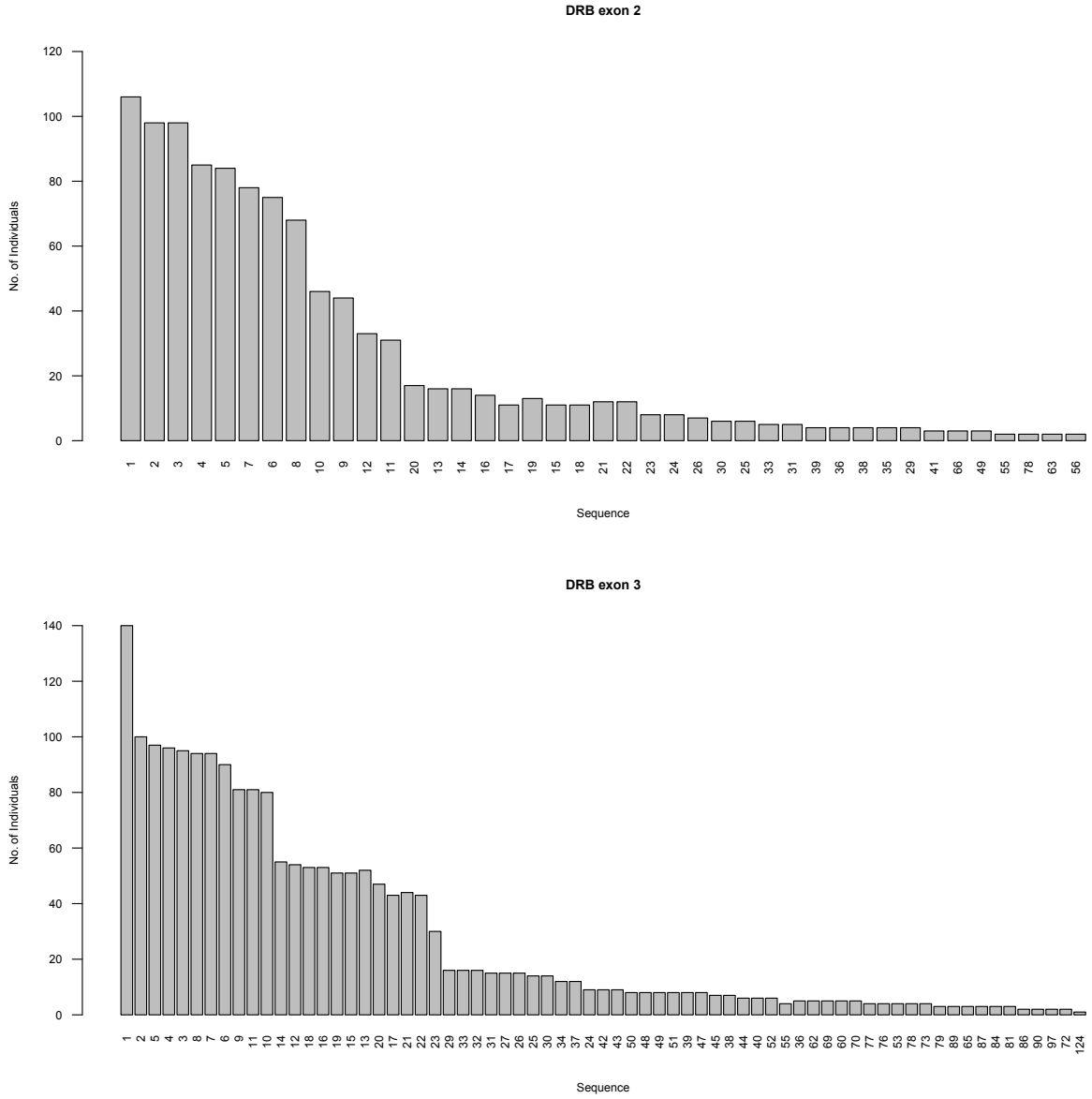


**Figure 2-5.** Bar plots showing the number of individuals that possess each of the recovered alleles for exon 2 (A) and exon 3 (B) in the study sample. Numbers on the X-axis refer to the sequence numbers of putative alleles recovered from genotyping analysis.

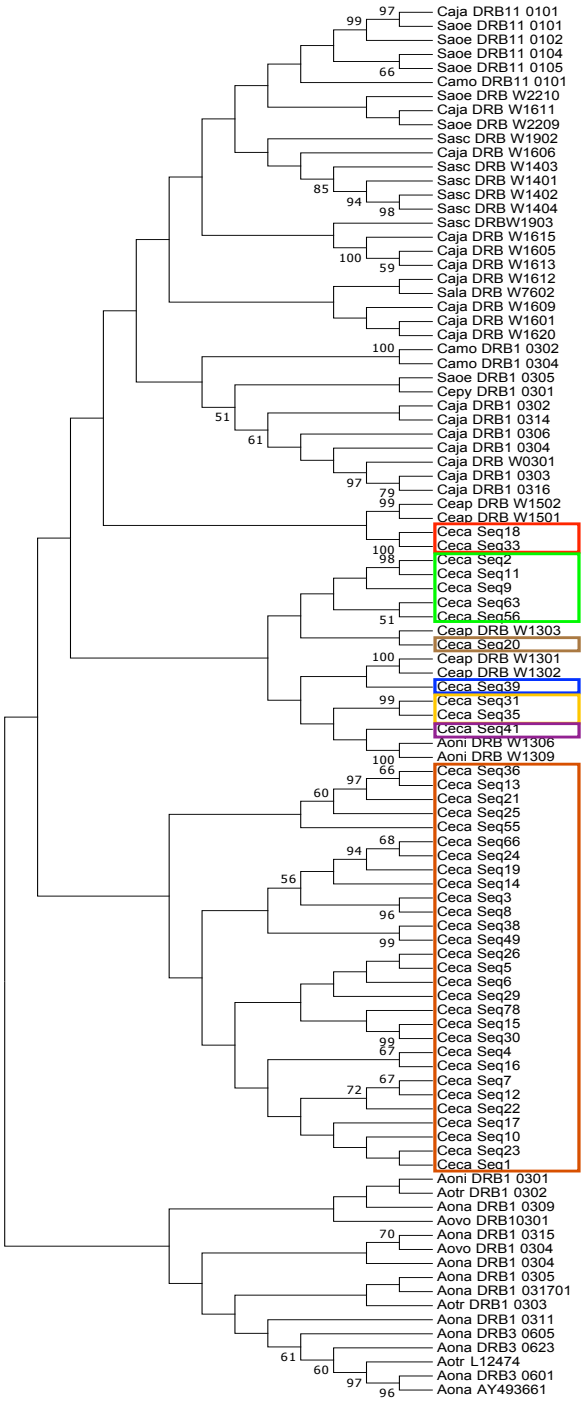


**Figure 2-6.** Phylogenetic tree of DRB exon 2 nucleotide sequences from the current and previous studies. Sequence labels follow standardized names for MHC sequences when possible, with the first four letters corresponding to the first two letters of the genus name followed by the first two letters of the species name. When a standardized name was not given, the Genbank accession number is given. Only bootstrap values ≥ 50 are shown. Colored boxes illustrate separate clusters of *Cebus capucinus imitator* sequences.

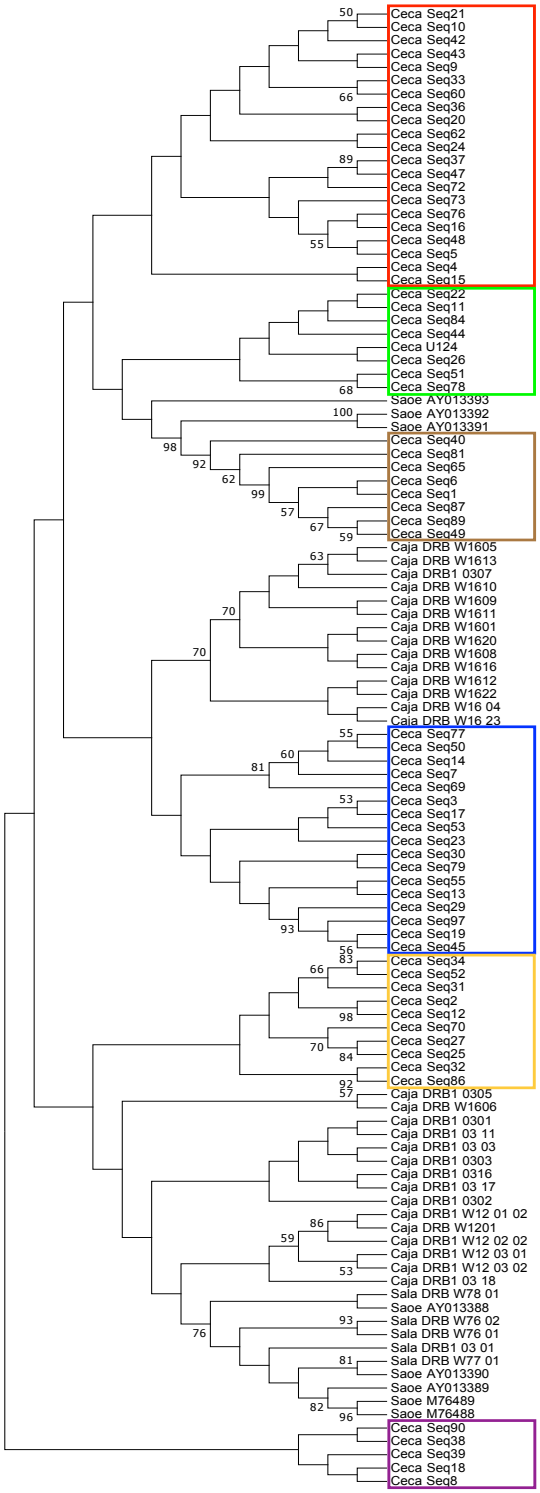


**Figure 2-7.** Phylogenetic tree of DRB exon 3 nucleotide sequences from the current and previous studies. Sequence labels follow standardized names for MHC sequences when possible, with the first four letters corresponding to the first two letters of the genus name followed by the first two letters of the species name. When a standardized name was not given, the Genbank accession number is given. Only bootstrap values ≥ 50 are shown. Colored boxes illustrate separate clusters of *Cebus capucinus imitator* sequences.

# APPENDIX

This appendix includes alignments of the nucleotide and amino acid sequences recovered from the MHC Class II DRB exons 2 and 3.

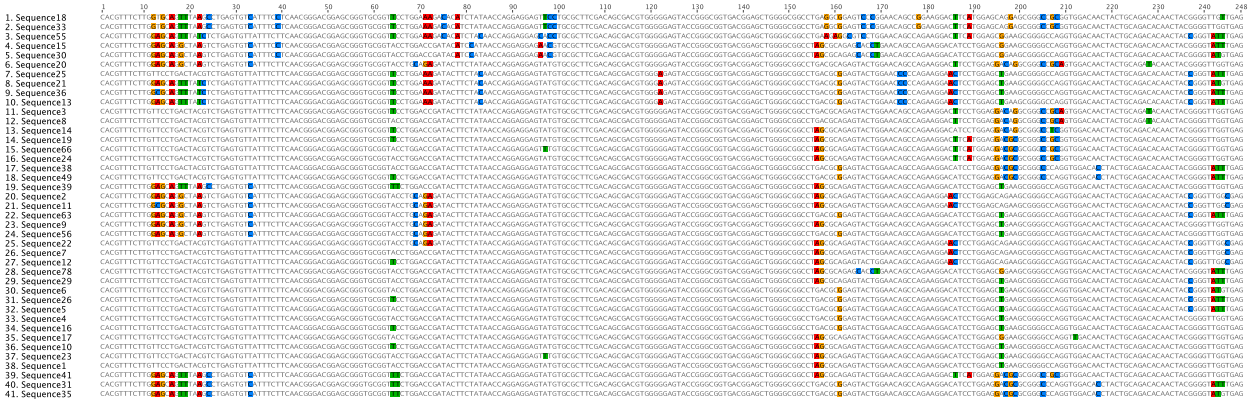

**Figure 2-8.** Nucleotide alignment of 41 putative DRB exon 2 alleles recovered in the current study. Base pair differences are highlighted with nucleotide-specific colored boxes: T - green; C - blue; A - red; and G - orange.

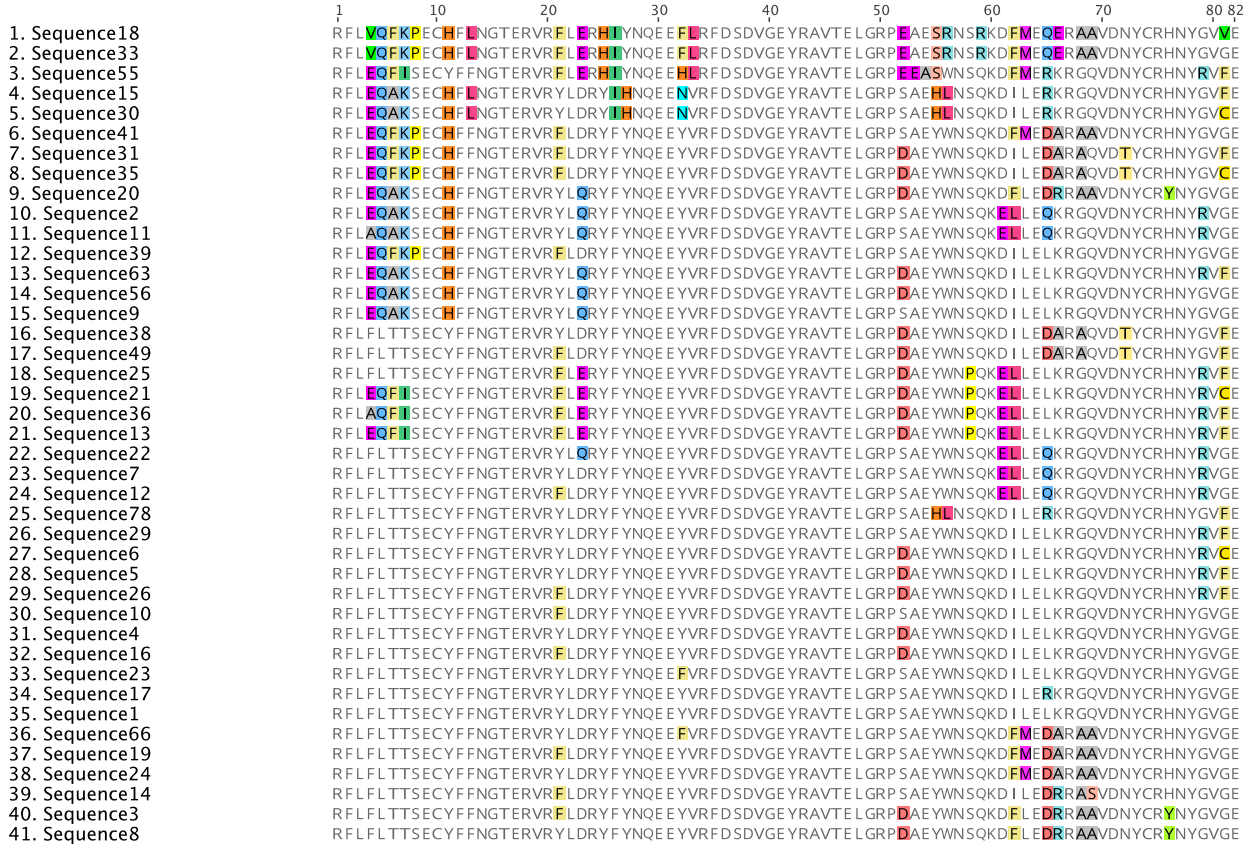

**Figure 2-9.** Alignment of 41 amino acid sequences resulting from putative DRB exon 2 alleles recovered in the current study. Amino acid changes are highlighted in residue-specific colors.

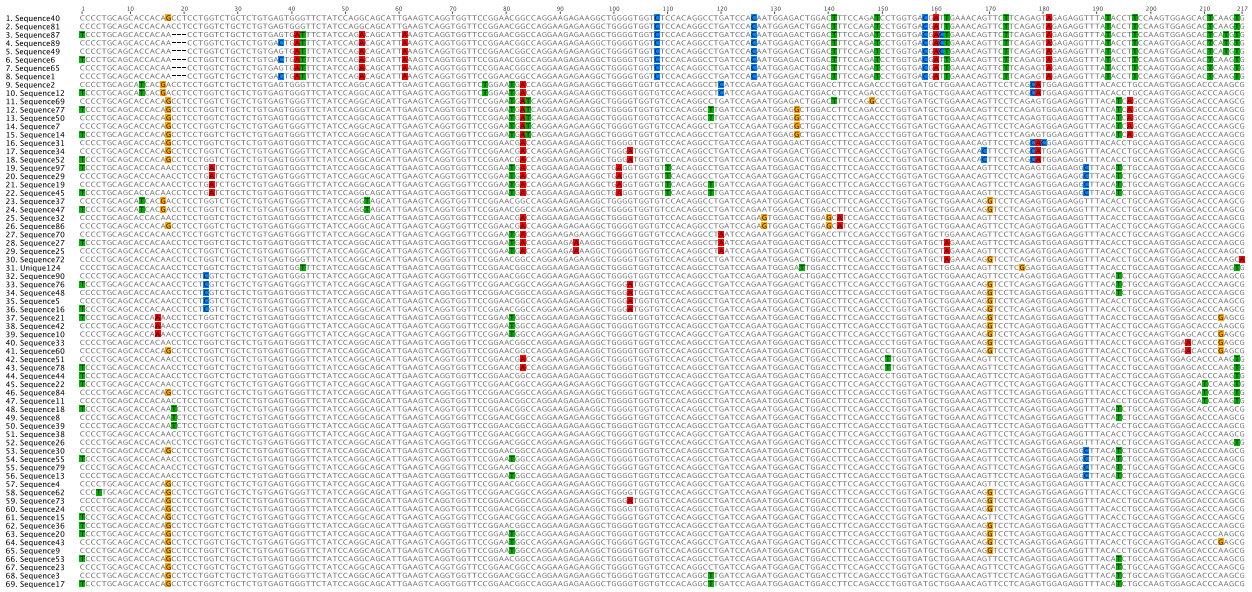

**Figure 2-10.** Nucleotide alignment of 69 putative DRB exon 3 alleles recovered in the current study. Base pair differences are highlighted with nucleotide-specific colored boxes: T - green; C - blue; A - red; and G – orange.

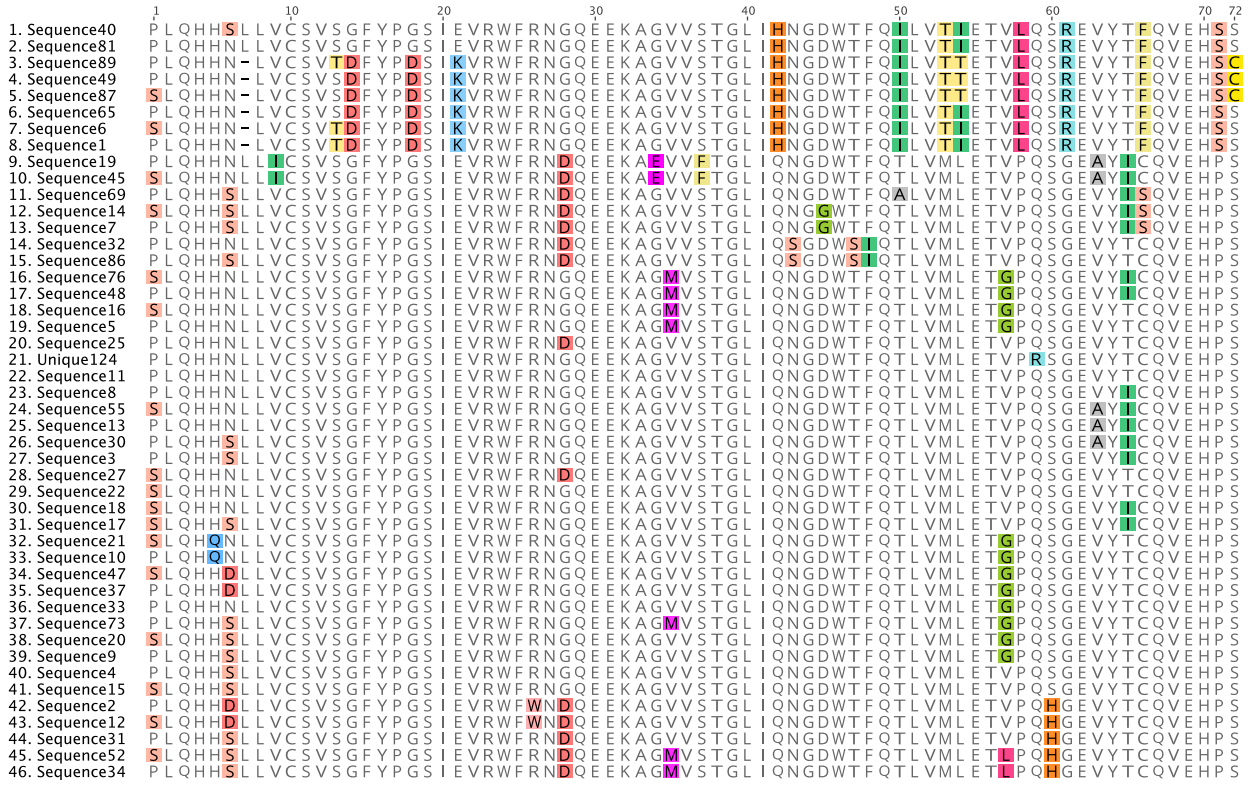

**Figure 2-11.** Alignment of 46 amino acid sequences resulting from putative DRB exon 3 alleles recovered in the current study. Amino acid changes are highlighted in residue-specific colors.

## References


Antunes, S.G., de Groot, N.G., Brok, H., Doxiadis, G., Menezes, A.A.L., Otting, N., Bontrop, R.E. (1998). The common marmoset: A new world primate species with limited Mhc class II variability. *Proc. Natl. Acad. Sci. USA*, 95, 11745-11750.

Baquero, J. E., Miranda, S., Murillo, O., Mateus, H., Trujillo, E., Suarez, C., ... & Parra-López, C. (2006). Reference strand conformational analysis (RSCA) is a valuable tool in identifying MHC-DRB sequences in three species of Aotus monkeys. *Immunogenetics*, *58*(7), 590-597.

Bontrop, R.E., Otting, N., de Groot, N.G., Doxiadis, G.G.M. (1999). Major histocompatibility complex class II polymorphisms in primates. *Immunological Reviews*, 167, 339-350.

Borghans, J.A.M., Kesmir, C., De Boer, R.J. (2007). MHC diversity in Individuals and Populations. In: Flower, D., Timmis, J. (eds) *In Silico Immunology*. Springer, Boston, MA. doi:10.1007/978-0-387-39241-7_10

Cadavid, L. F., Mejía, B. E., & Watkins, D. I. (1999). MHC class I genes in a New World primate, the cotton-top tamarin (Saguinus oedipus), have evolved by an active process of loci turnover. *Immunogenetics*, *49*(3), 196-205.

Cardenas, P. P., Suarez, C. F., Martinez, P., Patarroyo, M. E., & Patarroyo, M. A. (2005). MHC class I genes in the owl monkey: mosaic organisation, convergence and loci diversity. *Immunogenetics*, *56*(11), 818-832.

Di Fiore, A., Chaves, P. B., Cornejo, F. M., Schmitt, C. A., Shanee, S., Cortes-Ortiz, L., ... & Pacheco, V. (2015). The rise and fall of a genus: Complete mtDNA genomes shed light on the phylogenetic position of yellow-tailed woolly monkeys, *Lagothrix flavicauda*, and on the evolutionary history of the family Atelidae (Primates: Platyrrhini). *Molecular phylogenetics and evolution*, *82*, 495-510.

Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P.A. (1983). Graphical Methods for Data Analysis. Available from: https://www.researchgate.net/publication/265022942_On_the_Design_of_Self-Organized_Decision_Making_in_Robot_Swarms [accessed Dec 08 2017].

Consuegra, S., & de Leaniz, C. G. (2008). MHC-mediated mate choice increases parasite resistance in salmon. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1641), 1397-1403.

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, *14*(6), 1188-1190.

Diaz, D., Naegeli, M., Rodriguez, R., Nino-Vasquez, J. J., Moreno, A., Patarroyo, M. E., ... & Daubenberger, C. A. (2000). Sequence and diversity of MHC DQA and DQB genes of the owl monkey Aotus nancymaae. *Immunogenetics*, *51*(7), 528-537.

Diaz, D., Daubenberger, C. A., Zalac, T., Rodriguez, R., & Patarroyo, M. (2002). Sequence and expression of MHC-DPB1 molecules of the New World monkey Aotus nancymaae, a primate model for Plasmodium falciparum. *Immunogenetics*, *54*(4), 251-259.

Doxiadis, G.G.M., Otting, N., de Groot, N.G., Noort, R., Bontrop, R.E. (2000). Unprecedented polymorphism of Mhc-DRB region configurations in Rhesus Macaques. *J. Immunol*., 164(6), 3193-3199. doi:10.4049/jimmunol.164.3193

Doxiadis, G.G., van der Wiel, M.K., Brok, H.P., de Groot, N.G., Otting, N., 't Hart, B.A., van Rood, J.J., Bontrop, R.E. (2006). Reactivation by exon shuffling of a conserved HLA-DRB3-like pseudogene segment in a New World primate species. *Proc. Natl. Acad. Sci USA*, 103(15), 5864-5868. doi:10.1073/pnas.0600643103

Fedigan, L.M., Jack, K. (2001). Neotropical primates in a regenerating Costa Rican dry forest: a comparison of Howler and Capuchin population patterns. *International Journal of Primatology*, 22(5), 689-713.

Fedigan, L. M., & Jack, K. M. (2004). The demographic and reproductive context of male replacements in Cebus capucinus. *Behaviour*, *141*(6), 755-775.

Fedigan, L.M., Jack, K.M. (2012). Tracking Neotropical monkeys in Santa Rosa: lessons from a regenerating Costa Rican dry forest. In: Kappeler, P.M., Watts, D.P. (eds.) *Long-Term Field Studies of Primates.* Springer, Berlin. doi: 10.1007/978-3-642-22514-7_8

Fickel, J., & Weyrich, A. (2011). Female Mate Choice in Rodents. In *From Genes to Animal Behavior* (pp. 3-33). Springer Japan.

Galan, M., Guivier, E., Caraux, G., Charbonnel, N., Cosson, J. (2010). A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, 11, 296. doi:10.1186/1471-2164-11-296

Garamszegi, L.Z., Nunn, C.L. (2011). Parasite-mediated evolution of non-synonymous substitution rate at the functional part of the MHC in primates. *J. Evol. Biol*., 24(1): 184-95. doi: 10.1111/j.1420-9101.2010.02156.x.

Grahovac, B., Mayer, W.E., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H., Klein, J. (1992). Major-Histocompatibility-Complex DRB genes of a New-World monkey, the cotton top tamarin (*Saguinus oedipus*). *Mol. Biol. Evol.*, 9(3), 403-416.

Gyllensten, U., Bergström, T., Josefsson, A., Sundvall, M., Savage, A., Blumer, E. S., ... & Watkins, D. I. (1994). The cotton-top tamarin revisited: Mhc class I polymorphism of wild

tamarins, and polymorphism and allelic diversity of the class II DQA1, DQB1, and DRB loci. *Immunogenetics*, *40*(3), 167-176.

Hans, J.B., Haubner, A., Arandjelovic, M., Bergl, R.A., Fanfstück, T., Gray, M., Morgan, D.B., Robbins, M.M., Sanz, C., Vigilant, L. (2015). Characterization of MHC Class II B polymorphism in multiple populations of wild gorillas using non-invasive samples and next-generation sequencing. *Am. J. Primato*l., 77, 1193-1206. doi: 10.1002/ajp.22458

Huchard, E., Cowlishaw, G., Raymond, M., Weill, M., Knapp, L.A. (2006). Molecular study of Mhc-DRB in wild chacma baboons reveals high variability and evidence for trans-species inheritance. *Immunogenetics*, 58, 805-816.

Hughes, A.L., Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167-170.

Jack, K. M., & Fedigan, L. (2004). Male dispersal patterns in white-faced capuchins, Cebus capucinus: Part 2: patterns and causes of secondary dispersal. *Animal behaviour*, *67*(4), 771-782.

Jack, K. M., & Fedigan, L. M. (2006). Why be alpha male? Dominance and reproductive success in wild white-faced capuchins (Cebus capucinus). In *New perspectives in the study of Mesoamerican primates* (pp. 367-386). Springer US.

Janeway, C.A. Jr, Travers, P., Walport, M., et al. (2001) The major histocompatibility complex and its functions. In: Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27156/

Katoch, K., Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol., 30(4), 772-780. doi:10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Copper, A., Markowitz, S. Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647, 1649. doi:101093/bioinformatics/bts199

Knapp, L.A., Robson, J., Waterhouse, J.S. (2006). Olfactory signals and the MHC: a review and case study in *Lemur catta*. *Am. J. Primatol*., 68, 568-584.

Kriener, K., O'hUigin, C., Tichy, H., Klein, J. (2000). Convergent evolution of major histocompatibility complex molecules in humans and New World monkey. *Immunogenetics*, 51(3), 169-178.

Kriener, K., O'hUigin, C., Klein, J. (2001). Independent origin of functional MHC class II genes in humans and New World monkeys. *Hum. Immunol.,* 62(1), 1-14.

Kono, A., Brameier, M., Roos, C., Suzuki, S., Shigenari, A., Kametani, Y., ... & Walter, L. (2014). Genomic sequence analysis of the MHC class IG/F segment in common marmoset (Callithrix jacchus). *The Journal of Immunology*, *192*(7), 3239-3246.
Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), 1870-1874.

Middleton, S.A., Anzenberger, G., Knapp, L.A. (2004). Identification of New World monkey MHC-DRB alleles using PCR, DGGE and direct sequencing. *Immunogenetics*, 55, 785-790.

Müller, N., Ostner, J., Schalke, O., Walter, L. (2014). Towards the non-invasive assessment of MHC genotype in wild primates: analysis of wild Assamese macaque MHC-DRB from fecal samples. *Am. J. Primatol.*, 76(3), 230-238. doi:10.1002/ajp.22225

Nino-Vasquez, J.J., Vogel, D., Rodriguez, R., Moreno, A., Patarroyo, M.E., Pluschke, G., Daubenberger, C.A. (2000). Sequence and diversity of DRB genes of *Aotus nancymaae*, a primate model for human malaria parasites. *Immunogenetics*, 51, 219-230.

O'Connor, S.L., Blasky, A.J., Pendley, C.J., Becker, E.A., Wiseman, R.W., Karl, J.A., Hughes, A.L., O'Connor, D.H. (2007) Comprehensive characterization of MHC class II haplotypes in Mauritian cynomolgus macaques. *Immunogenetics*, 59(6), 449-462.

Olsson, M., Madsen, T., Nordby, J., Wapstra, E., Ujvari, B., & Wittsell, H. (2003). Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*, S254-S256.

Parr, N. A., Fedigan, L. M., & Kutz, S. J. (2013). Predictors of parasitism in wild white-faced capuchins (Cebus capucinus). *International Journal of Primatology*, *34*(6), 1137-1152.

Rohland, N., Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22(5), 939-946. doi:10.1101/gr.128124.111

Rylands, A.B., Mittermeier, R.A., Bezerra, B.M., Paim, F.P., Queiroz, H.L. (2013). Family Cebidae (Squirrel Monkeys and Capuchins). In: Mittermeier, R.A., Rylands, A.B., Wilson, D.E. (eds.) *Handbook of the Mammals of the World – Volume 3*. Lynx Edicions, Barcelona.

Schwensow, N., Fietz, J., Dausmann, K., & Sommer, S. (2008). MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology*, *22*(5), 617-636.

Shiina, T., Kono, A., Westphal, N., Suzuki, S., Hosomichi, K., Kita, Y. F., ... & Walter, L. (2011). Comparative genome analysis of the major histocompatibility complex (MHC) class IB/C segments in primates elucidated by genomic sequencing in common marmoset (Callithrix jacchus). *Immunogenetics*, *63*(8), 485-499.

Sommer, S., Courtiol, A., Mazzoni, C.J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artifacts and allelic dropout. *BMC Genomics*, 14, 542. doi:10.1186/1471-2164-14-542

Suarez, C. F., Cardenas, P. P., Llanos-Ballestas, E. J., Martinez, P., Obregon, M., Patarroyo, M. E., & Patarroyo, M. A. (2003). α1 and α2 domains of Aotus MHC Class I and Catarrhini MHC Class Ia share similar characteristics. *HLA*, *61*(5), 362-373.

Suárez, C.F., Patarroyo, M.E., Trujillo, E., Estupiñan, M., Baquero, J.E., Parra, C., Rodriguez, R. (2006). Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. *Immunogenetics*, 58, 542-558.

Trtková, K., Kupfermann, H., Grahovac, B., Mayer, W.E., O'hUigin, C., Vichy, H., Bontrop, R., Klein, J. (1993). Mhc-DRB genes of platyrrhine primates. *Immunogenetics*, 38, 210-222.

van der Wiel, M. K., Otting, N., de Groot, N. G., Doxiadis, G. G., & Bontrop, R. E. (2013). The repertoire of MHC class I genes in the common marmoset: evidence for functional plasticity. *Immunogenetics*, *65*(12), 841-849.

Watkins, D. I., Letvin, N. L., Hughes, A. L., & Tedder, T. F. (1990). Molecular cloning of cDNA that encode MHC class I molecules from a New World primate (Saguinus oedipus). Natural selection acts at positions that may affect peptide presentation to T cells. *The Journal of Immunology*, *144*(3), 1136-1143.

Watkins, D. I., Garber, T. L., Chen, Z. W., Toukatly, G., Hughes, A. L., & Letvin, N. L. (1991). Unusually limited nucleotide sequence variation of the expressed major histocompatibility complex class I genes of a New World primate species (Saguinus oedipus). *Immunogenetics*, *33*(2), 79-89.

Wu, M. S., Tani, K., Sugiyama, H., Hibino, H., Izawa, K., Tanabe, T., ... & Iseki, T. (2000). MHC (major histocompatibility complex)-DRB genes and polymorphisms in common marmoset. *Journal of molecular evolution*, *51*(3), 214-222.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.*, 24(8), 1586-1591.

Yang, Z., Swanson, W.J. (2002). Codon-substitution models to detect adaptive evolutions that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.*, 19, 49-57.

Yang, Z. (2006). *Computational molecular evolution*. Oxford University Press. London.

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular biology and evolution*, *22*(4), 1107-1118.

# Chapter 3

## Mitogenomics supports an unexpected taxonomic relationship for the extinct diving duck *Chendytes lawi* and definitively places the extinct Labrador Duck

**Janet C. Buckner, Ryan Ellingson, David A. Gold, Terry L. Jones, David K. Jacobs**

**Abstract**

*Chendytes lawi*, an extinct flightless diving anseriform from coastal California, was traditionally classified as a sea duck, tribe Mergini, based on similarities in osteological characters. We recover and analyze mitochondrial genomes of *C. lawi* and five additional Mergini species, including the extinct Labrador Duck, *Camptorhynchus labradorius*. Despite its diving morphology, *C. lawi* is reconstructed as an ancient relictual lineage basal to the dabbling ducks (tribe Anatini), revealing an additional example of convergent evolution of characters related to feeding behavior among ducks. The Labrador Duck is sister to Steller's Eider which may provide insights into the evolution and ecology of this poorly known extinct species. Our results demonstrate that inclusion of full length mitogenomes, from taxonomically distributed ancient and modern sources can improve phylogeny reconstruction of groups previously assessed with shorter single-gene mitochondrial sequences.

**Introduction**

True ducks (subfamily Anatinae) are a distinct clade of birds whose evolutionary history is valuable for understanding past and present environments. Unfortunately, the phylogenetic relationships within this group remain problematic, making it difficult to reconstruct the life history of several interesting but extinct duck species. We reconstruct the Anatinae phylogeny to systematically place the extinct diving duck *Chendytes lawi*. Miller (1925) erected the genus *Chendytes* based on Holocene fossil material from the California coast and nearby Channel Islands (Fig. 3-1). Two species are known. The goose-sized *C. lawi* has more degeneration of the wing elements than the smaller *C. milleri,* which may represent an intermediate form between a flying ancestor and the flightless *C. lawi* (Howard, 1955). Known only from the Pleistocene of San Nicolás Island, *C. milleri* is more limited in abundance and geography than *C. lawi,* which

has an extensive Holocene record extending from northern Baja to southern Oregon (Jones et. al., 2008a; Gruhn and Bryan, 2006). Carbon dating and the frequent recovery of material from middens suggest that the latter species was eventually lost to human exploitation, but unlike many other extinct Pleistocene lineages it persisted until as recently as 2,400 years ago (Jones et al., 2008a; Grayson, 2008).

*Chendytes* was traditionally classified as a sea duck, tribe Mergini. Miller (1925) allied it with the Surf scoter (*Melanitta perspicillata*), but an extended study by Livezey (1993) suggested placement in the eider genus *Somateria*. Despite uncertainty regarding the modern genus closest to *Chendytes,* previous authors consistently placed it amongst the Mergini based on osteological characters and proportions (Howard, 1947; Howard, 1955; Howard, 1964; Livezey, 1993; Miller, 1930; Miller et. al., 1961). Nevertheless, several characters used for phylogenetic placement of *Chendytes* were found to be convergent, as they also occur in other diving Anatinae clades such as *Tachyeres* (steamer ducks) and Aythyini (scaups/pochards), as well as in more basal diving anseriforms such as *Oxyura* (e.g. Ruddy duck) (Livezey, 1993; Miller, 1930).

Here, we address the systematic placement of *Chendytes lawi* using molecular data. We generated mitochondrial genome sequences for *C. lawi* and five additional sea duck species, including the extinct Labrador Duck, *Camptorhynchus labradorius*. We analyzed these in combination with other anatid mitochondrial sequences. Using maximum likelihood and Bayesian inference methods, we compare phylogenetic results from three alternative data matrices: (1) maximized taxonomic sampling with missing data (2) mitogenomes only with limited taxonomic sampling and (3) a two-gene matrix with maximized taxonomic sampling and zero missing data. Our results consistently indicate that *Chendytes lawi* is not a member of any currently recognized diving duck clade but is a stem dabbling duck with convergent osteological

adaptations for diving. *Camptorhynchus labradorius* is an eider that is sister to *Polysticta stelleri* (Steller's Eider) within the sea-duck tribe Mergini. The combination of mitogenomic and single mitochondrial gene sequences improve estimates of phylogeny within the Anatinae.

## Methods

**Sample Collection –** In this study, we follow the binomial list of Clements et. al. (2017) for species – eBird/Clements checklist of birds of the world – supplemented by the higher taxonomic classifications outlined by Cracraft (2013). Samples of *Chendytes lawi* bone fragments were provided by TLJ. The bones were recovered from archaeological site, CA-SLO-2, on the central California coast in San Luis Obispo County, midway between Avila Beach and Morro Bay (Jones et al. 2008b).[1] Two toe pad samples from separate individuals of *Camptorhynchus labradorius* (Labrador Duck) were lent by the American Museum of Natural History (AMNH). One toe pad sample from each of the extant species of Mergini (*Mergus serrator*, *Mergus merganser americanus*, *Melanitta nigra* and *Melanitta fusca deglandi*) were provided by the Donald R. Dickey Collection of Birds and Mammals at the University of California, Los Angeles (UCLA DC). All other mitochondrial sequences for anatid species were acquired from Genbank. See Table S1 for catalog numbers, accession numbers and species information.

**DNA Extraction, Library Preparation and Target Enrichment** – DNA was extracted from a left tibiotarsus shaft of *Chendytes lawi* in a dedicated ancient DNA facility at UCLA. Prior to extraction the outer surface of the bone was removed with a sterilized dremel tool and a new,

[1] The *C.lawi* fossils were excavated in 1968 (Greenwood 1972) from a 3.4m deep shell midden deposit, and were identified in 2007 by Judy Porcasi who relied on comparative materials at Los Angeles County Museum of Natural History (Paleontology Department.) and the Department of Biology at California State University, Long Beach. Thirty-three radiocarbon determinations show that the archaeological site was occupied from 10,300 to 300 cal BP. The specimen that produced the sequence was recovered from unit S4/W2, from a depth of 270-280 cm (Jones et al. 2008b). Eight radiocarbon determinations show that the levels between 280 and 200 cm date between 9000 and 5000 cal BP.

disposable rotary head to reduce exogenous contamination. The sample was subsequently ground into a coarse powder. This powder was then incubated in a solution of EDTA pH 8.0 and proteinase-K for 24 hours on a rotator followed by 3 hours of incubation at 56º C. DNA was then treated to the silica-adhesion protocol described in Rohland and Hofreiter (2007). The resulting DNA extract was quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) and then stored at –20º C prior to downstream application.

DNA extraction from all toe pads was performed in a dedicated hood for historic DNA extraction in a pre-PCR DNA free laboratory. We followed the manufacturer's protocol for the Qiagen DNEasy extraction kit (Qiagen Inc., Valencia, California, USA) with the following modifications: adding 40uL proteinase K and 20uL of 1M dithiothreitol (DTT) to the extraction buffer and incubating the sample at 50ºC until completely digested, about 48 hours (Fulton et. al., 2012a). Qubit (ThermoFisher) fluorometric quantification of double stranded DNA followed extraction and all subsequent steps of the protocol.

After extraction, we prepared libraries from all DNA extracts with end-repair and dual-indexed adapter ligation using the Kapa Biosystems LTP Library Preparation Kit and custom indexes from BadDNA UGA oligos, respectively, following the manufacturers' protocols. Resulting DNA was size selected with Seramag Speedbeads to exclude fragments less than 150 base pairs (bp) (Rohland and Reich, 2012). Note: Though libraries were prepared in the same way for all samples, the extraction and library preparation (indexing) of the *Chendytes lawi* specimen was completed in the dedicated ancient DNA facility before any handling or processing of the historic toe pad samples.

We synthesized an 80mer bait set with 8x tiling at MYcroarray based on eight published anatid mitogenomes (*Dendrocygna javanica, Cygnus atratus, Mergus squamatus, Cairina*

*moschata, Aythya americana, Anas formosa, Anas crecca, Anas platyrhynchos*) to target the full mitochondrial genome. Following library preparation and quantification, we performed target capture of mitogenomes following the manufacturer's protocol except that hybridization and subsequent washes were carried out at a temperature of 55ºC. Dual-end sequencing (2 × 300 bp) was performed on the pooled, enriched libraries on a MiSeq instrument (Illumina Inc., San Diego, California, USA).

**Read Processing** – Reads were de-multiplexed and sequence quality was evaluated using FastQC (Andrews, 2010). As the insert lengths from our ancient and historic DNA were on average several base pairs shorter than our read length, we created a first set of "QC reads" by cutting the sequences down to the first 70 bp. We also used the Trimmomatic pipeline enabled on the online platform Galaxy (version 0.32.3) on the raw reads to remove adapter contamination and sequencing artifacts to create a second set of QC reads (Bolger et. al. 2014). Poor quality leading and trailing ends were removed and sequences were trimmed based on a sliding window of 5 base pairs where windows with an average quality less than 30 were removed. Reads from Trimmomatic of fewer than 30 bp were excluded. We mapped these two sets of processed reads for each of our museum samples from extant species as well as the Labrador Duck to the single published mitogenome for a Mergini species, *Mergus squamatus,* using Geneious Pro 9.0.5 (Kearse et al., 2012). Reads from *Chendytes lawi* were verified by eye and suspected contaminants were identified using NCBI's nucleotide BLAST and removed. Once all samples were processed, contigs were assembled to produce two mitogenome sequences for each species based on the 70mer and Trimmomatic processed reads. We compared the mappings for the two datasets for each species visually by aligning the assembled contigs and looking for discrepancies. When discrepancies

were found, the highest base call quality was used to decide between nucleotides. When there was no difference in base call quality, the Trimmomatic mappings were chosen over the 70mer mappings.

**Validation of Historic DNA –** We used mapDamage 2.0 (Jónsson et. al., 2013) to identify nucleotide mis-incorporations in our ancient and historic DNA samples. Such damage patterns are typical and expected from these historical sources of genetic data and thus also serve as a method of confirming that ancient DNA (aDNA) sequences are from the target specimen and not from modern contaminants. We used Bowtie (Langmead et al. 2009) to map the MiSeq reads back to the assembled mitogenomes of *C. lawi* and *M. deglandi*. The resulting alignment (BAM) files were used as input files for mapDamage.

**Phylogenetic Reconstruction –** We aligned mitogenome data to previously published Anatid mitogenomes and mitochondrial gene sequences available from GenBank using the MUSCLE algorithm in Geneious Pro 9.0.5. Each taxon had between one and four gene fragments, or a complete mitogenome, represented in the alignment. In all, 32 taxa had complete mitogenomes, 14 had four gene fragments, 17 had three fragments, 30 had two fragments and 11 taxa had one fragment and were present in the combined matrix (matrix A) of 104 total taxa. Two additional partial matrices of 32 taxa with complete genomes (matrix B), and 72 taxa with both Cyt b and COI fragments without missing data (matrix C) were generated. In matrix C, one taxon with missing data was included, *Polysticta stelleri*, in order to confirm relationships to extinct species recovered from analyses of matrix A. Note that the control region was not included in any of the three matrices to mitigate issues related to ambiguous alignment.

Using k-means clustering in R (Ellingson et al., 2014), each of the sequence alignments were binned into five rate classes according to per-site relative rates as estimated in HyPhy

(Kosakovsky Pond et al., 2005). Sites that evolve quickly—particularly at deeper nodes of a phylogeny—potentially lead to long-branch attraction if substitution models cannot accurately account for homoplasy. It is not uncommon to omit such rapidly evolving sites from phylogenetic analyses (Pisani, 2004), a procedure analogous to (but more precise than) choosing loci based on the optimal rate of molecular evolution for a given phylogeny. In matrix A sites of the two highest rate classes ($n$ = 4 bp total) were removed and in matrix B sites of the single highest rate class (n=5 bp total) were removed as they were most likely to exhibit homoplasy. In matrix C, no sites were evolving especially rapidly to warrant removal from the alignment.

We reconstructed the phylogenetic relationships of Anatidae with all three matrices using MrBayes 3.2.6 (Huelsenbeck and Ronquist, 2001) and performed two runs of the analysis for 50 million generations, with four chains (3 heated, one cold) and Beagle options enabled on CIPRES (Miller et. al., 2010). The data in matrix A were partitioned into three rate categories, in matrix B four categories and in matrix C five categories. We applied a mixed nucleotide substitution model with gamma rate heterogeneity. We also reconstructed the phylogenies under maximum likelihood using RAxML 8.2.10 (Stamatakis, 2014) under the default settings enabled on CIPRES.

**Results**

We recovered complete mitochondrial genomes for the four extant species and two individual Labrador Duck specimens. Excluding the control region, only *Mergus serrator* was missing a total of two bases in the alignment among historic samples. We recovered a near complete mitogenome (45 missing bases, not including missing bases in the control region) for *Chendytes lawi* with 12X average coverage. The mitogenomic sequences from the two Labrador Duck toe pads were identical, we therefore include only the highest quality sequence (determined by read

depth) in the phylogeny. Further details of the quality of assembly are included in Table 1. Data matrix A included a total of 15,513 sites, 5,778 of which were variable (37.2%). Data matrix B included a total of 15, 512 sites, 5,662 of which were variable (36.5%). Data matrix C included a total 1,093 sites, 464 of which were variable (42.5%).

The topologies from our maximum likelihood (ML) and Bayesian trees based on matrix A were largely congruent. With geese and swans constrained as the outgroup, the Tadornini branch first, followed by the Mergini, then the Aythyini and finally the Anatini (Fig. 3-2). Within the tribe Anatini, the topologies differ in that there is a polytomy at the base of this clade in the Bayesian tree while in the ML tree *Spatula/Sibirionetta* branch first, leaving the South American dabblers sister to *Mareca/Anas*. Consequently, placement of the South American clade consisting of *Amazonetta*, *Speculanas* and *Tachyeres* is uncertain (bootstrap support (bs) = 39) and differs between the two methods (Fig. 3-2)

The extinct Labrador Duck, *Camptorhynchus labradorius*, is strongly supported— bs = 100; posterior probability (pp) = 1.0—as sister to Steller's Eider, *Polysticta stelleri*, consistent with previous placement of the Labrador Duck within the Mergini based on morphology (Livezey, 1995). This *Polysticta*/*Camptorhynchus* grouping falls within a monophyletic eider group, and all modern sea ducks sampled form a well-supported (bs = 66; pp = 1.0) monophyletic tribe Mergini (Fig. 3-2). Sea duck genera are monophyletic except for *Mergus,* which includes *Lophodytes* but with low support in the maximum likelihood tree (bs = 42) and forming a polytomy in our Bayesian tree (Fig. 3-2).

The topologies resulting from matrix B (mitogenomes only) support the relationships among tribes obtained from analyses of matrix A and the placement of *Chendytes lawi* (Appendix Fig. 3-5). The Labrador Duck remains closely related to the Mergini, but more specific comment on

this relationship is not possible with this reduced taxonomic sampling. In the analyses of matrix C (Cyt b and COI, no missing data), taxonomic placement of both extinct species is largely maintained (Appendix Fig. 3-6). The Labrador Duck is again confirmed as sister to Stellar's eider (*P. stelleri*) with high support (bs=96; pp=0.99). *Chendytes lawi* is maintained as basal to the Anatini (bs=71; pp=0.93). However, while matrix C contained almost no missing data, there is substantial degradation of resolution in terms of the relationships among the duck tribes and significant decreases in statistical support generally across nodes relative to analyses of the combined matrix A.

## Discussion

**Anatinae phylogeny –** Our approach to the overall topology of the Anatinae phylogeny involved (1) combining a well-supported backbone of complete mitogenomes from a limited number of taxa across true ducks with (2) shorter single-gene sequences representing more thorough taxonomic sampling (Fig. 3-2). Analysis of this whole matrix compared with analyses of each of the reduced matrices shows the importance of the mitogenomic backbone in generating strong support for the basal arrangement of tribes as well as many internal nodes of the tree. Further evidence of the utility of analyzing a patchwork of gene fragments with the aid of a strong mitogenomic backbone is shown by the inability to place the South American duck clade (*Amazonetta, Tachyeres, Speculanus*). There are no mitogenomes available for any member of this group, or an apparent close relative. Beyond resolution of the base of the Anatinae phylogeny and placement of the extinct taxa, our effort produced a comprehensive and well-supported hypothesis that bears some similarity to previous studies.

Our work is largely in agreement with Mitchell et. al. (2014) although they did not include representatives of Mergini. Our basal topology contrasts with Fulton et. al. (2012), where

the first split among the subfamily Anatinae is between Anatini and the remaining tribes. Within the latter, Mergini branches first leaving Aythini sister to Tadornini. Our results resemble those of Gonzalez et. al. (2009) with the exception that in their phylogeny the branching order of Tadornini and Mergini are reversed: Mergini branches first, followed by Tadornini, then Aythyini and finally Anatini. This latter relationship is also recovered in the two-gene phylogeny presented in Sun et. al. (2017), however in the same publication the authors recover a branching order equivalent to ours in their mitogenome only tree, supporting our contention that inclusion of these complete mitogenomes provides important resolution in the base of the tree. Our own two-gene phylogenies, where the data are most like previous mitochondrial (mtDNA) gene-based studies, did not support any hypothesis of relationships among major tribes (Appendix Fig. 3-6).

The implication then is that the mitogenomes we included in our combined matrix (A) provide a strongly supported backbone. This topology is also recovered from the mitogenome data matrix with virtually no missing data and a much larger number of variable sites than would be available from any two-gene alignment. Inclusion of two genes has been the dominant pattern among published anatid phylogenies. Inclusion of additional fragmented mtDNA sequence data with this strong backbone then allows a better supported phylogeny with more species aiding in the placement of our extinct ducks. Wiens (2006) demonstrated that inclusion of more taxa often benefits phylogenetic results even at the cost of including large amounts of missing data, consistent with our results. Thus, sequencing of more anatid mitogenomes may improve our understanding of their evolutionary history with minimal cost, while allowing the effective inclusion of samples from a variety of source material. This approach does have limits where whole genomes are poorly distributed or where gene fragments are too small to allow effective inclusion of taxa in the analyses. Thus, researchers must always take care to evaluate the effect

of varying configurations of data on analyses and results. Additionally, mitochondria provide an important but limited basis for phylogenetic analysis, which should ultimately be compared to other data sources. However, such alternative data may be less easily obtained from degraded material such as fossil bone or museum specimens.

**Phylogenetics and inferred ecology of *C. lawi* and *C. labradorious* –** In previous phylogenetic studies of Anseriformes, non-monophyletic assemblages share morphological adaptations for diving, dabbling, or grazing, indicating convergence of these feeding modalities (Faith, 1989; Gonzalez et. al., 2009; McCracken et. al., 1999; Olsen, 2015). Molecular phylogenetics strongly supports placement of *Chendytes lawi* in a novel position as sister to the Anatini, not with Mergini or any other diving clade within Anatidae (Fig. 3-2). Thus, this lineage likely represents a fourth independent evolution of diving within true ducks and *C. lawi* appears hyper-adapted to this feeding modality.

Head and beak morphology suggest *C. lawi* was an invertivore (Miller *et. al.*, 1961). Based on the large, robust morphology of the cervical vertebrae, skull, and bill, Miller further argued that *C. lawi* specialized on sessile invertebrates commenting that the species likely possessed "a remarkable ability to wrench off invertebrate animals attached to hard substrate" (Miller *et. al*., 1961). The lack of obvious adaptations for piscivory and the presence of a robust supraorbital process, which likely protects the eyes and salt glands during benthic foraging, further support such an invertebrate diet (Livezey, 1993; Raikow, 1970). Mussels and urchins are often abundant on the California coast and are amongst the diverse invertebrate prey of other large diving ducks (Lovvorn *et al.* 2003; Savard and Petersen 2015). The largest modern sea ducks regularly dive to 60-m depths (Žydelis & Richman, 2015) and with its larger body size, *C. lawi* presumably could have taken advantage of prey in deeper habitats (Miller *et. al.*, 1961).

Larger body size and foot-propelled diving were achieved at the expense of wing degeneration and likely resulted in high wing loadings in *C. lawi* (Howard, 1947; Livezey 1993). The close relationship between *C. lawi* and the Pleistocene *C. milleri* with its intermediate reduction in flight (Howard 1955) appears comparable to the differential loss of flight in lineages of South American Steamer ducks *(Tachyeres* - Fulton et. al., 2012b). This may suggest a dynamic or iterative late-Pleistocene evolution of a flightless form in the *Chendytes* lineage facilitated by their presence on islands. Flightlessness has strong implications for reproductive biology in birds, rendering them obligate residents and typically ground-nesting species. In *Chendytes*, extensive recovery of fossil birds and eggshells from the Channel Islands confirms ground nesting and further suggests that they mainly bred on offshore islands (Guthrie, 1992; Miller et. al., 1961). *Tachyeres* species display a similar pattern where they are distributed in southern South America and offshore on the Falkland Islands, displaying the inferred reproductive behavior evident in the fossil record for *Chendytes*.

Steamer ducks (*Tachyeres*) are the only extant flightless diving anseriforms, and *Chendytes lawi* and *Shiriyanetta hasegawai* are the only flightless divers reported from the fossil record. The latter is described as a large sea duck of the tribe Mergini but given the emphasis of its morphological similarity to *Chendytes* (Watanabe and Matsuoka, 2015), it could represent a third example of a derived Anatinae lineage. Molecular systematics are required to support or refute this hypothesis. However, if *S. hasegawai* is indeed more closely related to *Chendytes* than to Mergini, a repeated pattern could emerge of resident, large, flightless Anatinae on mid-latitude insular coasts. Competition in species-rich habitats might have driven repeated adaptation of diving specialists capable of exploiting deep-water resources inaccessible to their primarily dabbling sister species. Thus, our result that *Chendytes* spp. are not sea ducks (Mergini) but

dabbling ducks (Anatini) with derived anatomical features specialized for diving could be the beginning of broader evolutionary interpretation if more data can be generated.

Although *Chendytes lawi* is known exclusively from fossil material and has been extinct for millennia, aspects of its biology and ecology, (e.g., breeding range) are more easily established than those of the much more recently extinct (c. 1875) Labrador Duck. Relatively specific locality data are only established for a small number of the roughly forty known museum specimens and these appear to pertain exclusively to wintering birds (see Fig. 3-1). Purported eggs of the species were later shown through genetic analyses to be those of other ducks (Chilton and Sorenson, 2007). Given the lack of verified information on the Labrador Duck beyond the winter range, an improved understanding of Mergini relationships may help interpret its biology.

In contrast with our result for *C. lawi*, the Labrador Duck (*Camptorhynchus labradorius*) does fall within the Mergini as previously argued. However, it is sister to Steller's Eider, not a scoter of the genus *Melanitta* as suggested by Livezey (1995). Similarities of the bill and plumage reflect this sister taxon relationship. Steller's Eider, Long-tailed Duck (*Clangula hymelis*) and Harlequin Duck (*Histrionicus histrionicus*), which are basal to the remaining eiders, have more sensitive bills to feed on aquatic insects in the breeding season and feed more extensively on softer bodied arthropod (amphipod) prey in the winter than *Somateria* or scoters. Steller's Eider also appears to have a lower dependence on bivalves and a greater dependence on gastropods than these larger invertivore Mergini (Cottam, 1939; Bustnes & Systad, 2001; Lovvorn et al. 2003; Ouellet, et al. 2013). Studies further suggest that Steller's Eider feeds extensively in kelp and eelgrass beds (Bustnes et al. 2000; Metzner 1993).

Steller's Eider use of distinct breeding, molting and wintering habitat may allow us to generate more informed hypotheses regarding the Labrador Duck range and biology. Steller's Eider breeding occurs around freshwater ponds in the arctic coastal plains of the Alaskan North Slope and in Siberia; they then migrate to shallow protected embayments to molt in the autumn (Petersen et al 2006; Martin et al. 2015), and ice free shallow water habitats in winter. A major Steller's Eider molting ground on Novoya Zemla (Petersen et. al., 2006), has extensive shallow glacial formed embayments very similar to those of the Labrador coast potentially suggesting that Labrador was a significant molting area. Coastal plain breeding may then have occurred further north in coastal plain ponded habitats in the Fox Basin or Eastern Ungava as suggested by some earlier authors (Audubon 1843; Cottam 1939). Labrador Ducks recovered from the sea bird markets suggest the south shore of Long Island as the most extensive source of known specimens (Dutcher, 1891; 1894). This coast has large bay/lagoon systems behind barrier islands, such as Great South Bay that once contained vast eel grass beds (Cottam & Addy 1947; Jones & Schubel 1978). This and similar habitats in other shallow vegetated marine bays along the coast were likely winter habitats of the Labrador duck given the Steller's eider analogue.

**Conclusion**

Our phylogenetic analyses placed *C. lawi* as sister to the dabbling group Anatini, contrary to its traditional morphology-based placement within the Mergini (sea ducks). Its large body size, exaggerated hind limbs, robust head and neck, high degree of wing degeneration and inability to fly point to an emphasis on foot-propelled diving to procure invertebrate prey. Its taxonomic placement and adaptations to dive-foraging suggest an evolutionary pattern more similar to steamer ducks (*Tachyeres*) than any other living anatid group. The sister taxon

relationship between the extinct Labrador Duck and Steller's eider may ultimately help clarify ecological evolution of eiders while permitting the reconstruction of aspects of Labrador Duck biology. Our study emphasizes the need, where possible, to use molecular phylogenetics to corroborate morphological systematic studies, particularly in groups prone to convergent evolution such as the Anseriformes. It also strongly suggests that inclusion of taxa with incomplete sequence representation is beneficial if there is a sufficient backbone of shared data across the taxa included.

**Table 3-1.** Summary of sequencing coverage for anatid mitogenomes

| Species | **Sequence length** | **Total Reads in Contig** | **Total Bases** | **Avg. bp coverage** | **Min. bp coverage** | **Max bp coverage** |
|---|---|---|---|---|---|---|
| *C. lawi* | 16,647 | 2,867 | 217,221 | 12.1 | 0 | 40 |
| *C. labradorius (45802)* | 16,624 | 1,074,706 | 75,245,945 | 4,631.4 | 26 | 11988 |
| *C. Labradorius (45803)* | 16,611 | 654,259 | 45,814,655 | 2,819.8 | 12 | 8225 |
| *M. f. deglandi* | 16,608 | 841,326 | 58,909,345 | 3,605.7 | 15 | 8832 |
| *M. nigra* | 16,625 | 219,134 | 15,355,905 | 942.6 | 6 | 3711 |
| *M. m. americana* | 16,613 | 439,024 | 30,748,205 | 1,874.6 | 35 | 5080 |
| *M. serrator* | 16,632 | 575,166 | 40,278,145 | 2,470.9 | 40 | 8055 |

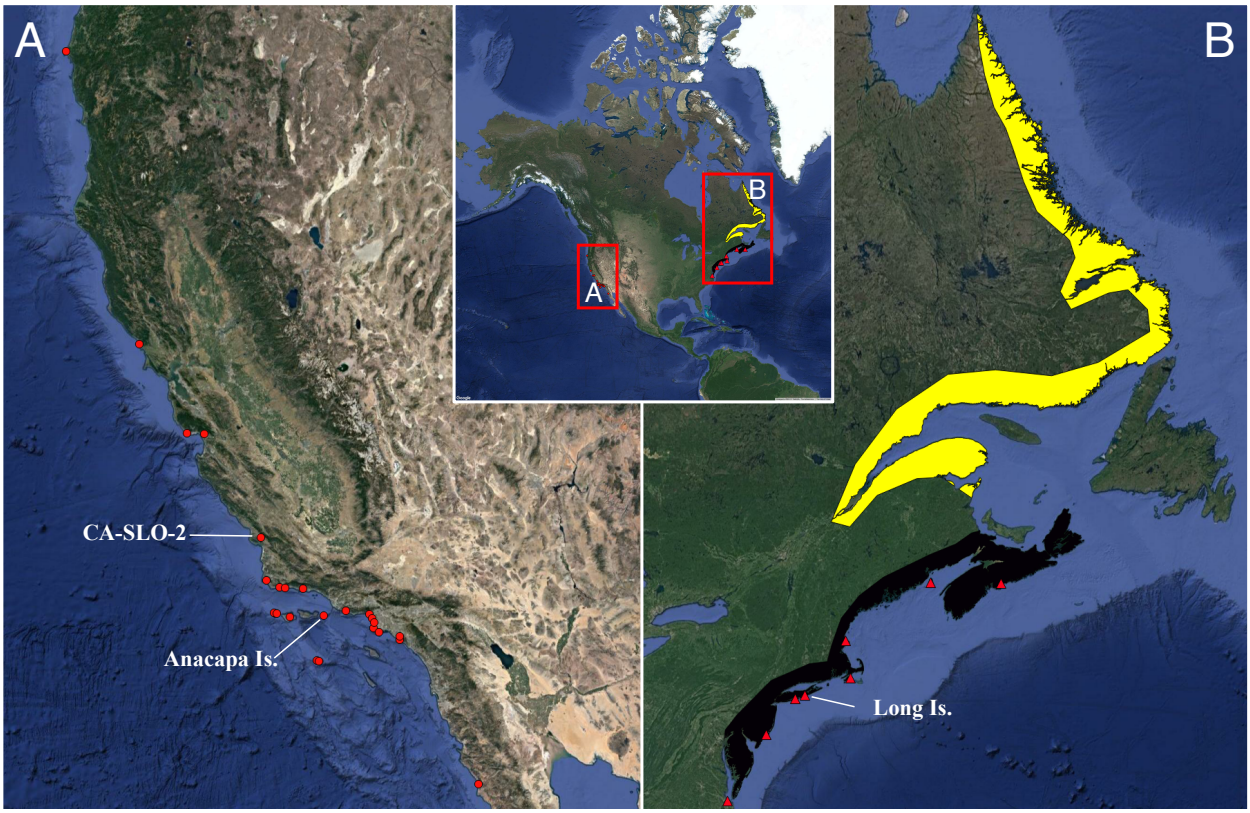


**Figure 3-1.** Distribution maps for *Chendytes sp.* (A) and *Camptorhynchus labradorius* - Labrador Duck (B). Red circles demarcate localities of fossil recovery for *Chendytes lawi* and *Chendytes milleri.* The locality (CA-SLO-2) for the sample used in this study is labelled. The only known locality for *C. milleri* (Anacapa Island) is also labelled. Red triangles demarcate verified locality records of the Labrador Duck. However, these are all winter records. The reconstructed breeding range (yellow) and wintering range (black) according to birdlife.org is shown. Note that the breeding and molting ranges of this taxon are necessarily speculative as indicated by our question mark (see Discussion). The locality (Long Island, NY) for the two sampled Labrador ducks used in this study is labelled.

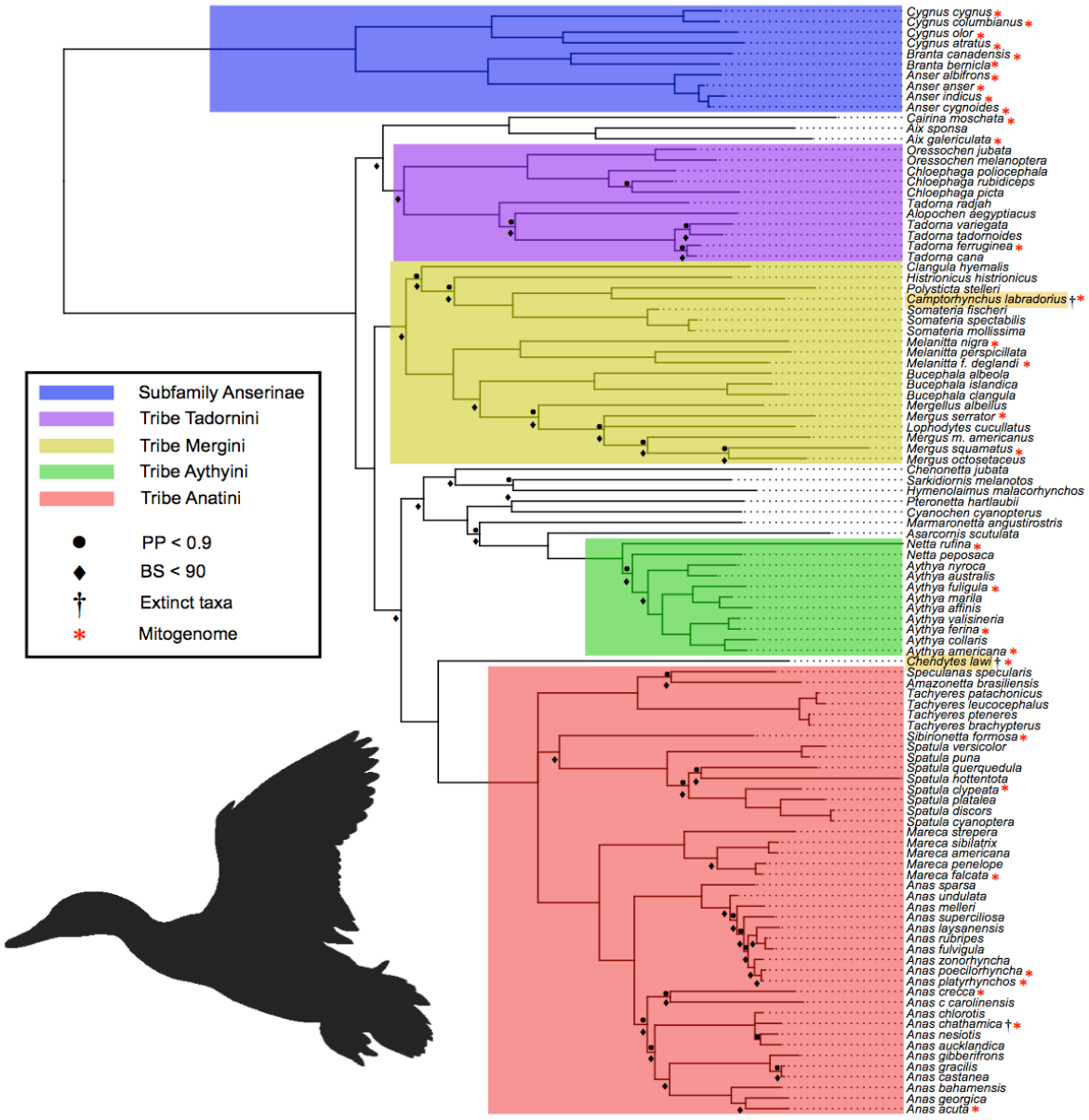


**Figure 3-2.** Bayesian inference tree based on the combined data matrix A. Circles above nodes represent low posterior probabilities from the MrBayes analysis (<0.9). Diamonds below nodes represent low bootstrap values from the RAxML analysis (<90). *Chendytes lawi* and *Camptorhynchus labradorius* (Labrador Duck) are marked with an asterisk (*). A dagger (†) indicates an extinct species. Major subfamilies and tribes are color coded (see legend).

# APPENDIX

The supplementary files from the publication in press for chapter three are included in this appendix. These include two tables of information on the genetic data used in the study. It also includes graphics that demonstrate nucleotide mis-incorporations in historic DNA samples to verify their source. Scatter plots showing per-site relative rates for all matrices used to reconstruct phylogenetic trees. Tree reconstructions from alternative DNA matrices B and C are also included. Finally, a graphic illustrating the distribution of missing data across matrix A.

**Table 3-2.** Genbank accession numbers for genes used in the phylogeny reconstruction.

| Species | mtGenome | CytB | COI | 12S | ATP8/ATP6 |
|---|---|---|---|---|---|
| Aix galericulata | NC023969 | | | | |
| Aix sponsa | | AF059053 | AY666569 | HM063543 | |
| Alopochen aegyptiaca | | HM063576 | | HM063553 | |
| Amazonetta brasiliensis | | HM063571 | JN801484 | HM063533 | AF173731 |
| Anas acuta | NC024631 | | | | |
| Anas aucklandica | | AF059056 | | AF173482 | AF173490 |
| Anas bahamensis | | EU914147 | FJ027081 | | |
| Anas castanea | | AF059065 | | AF173481 | AF173494 |
| Anas chathamica | KF562761 | | | | |
| Anas chlorotis | | AF059061 | | AF173484 | AF173491 |
| Anas crecca | NC022452 | | | | |
| Anas c. carolinensis | | AF059063 | DQ434280 | HM063539 | AF173720 |
| Anas fulvigula | | AF059074 | DQ432723 | | |
| Anas georgica | | AF059075 | FJ027092 | | |
| Anas gibberifrons | | | JQ174015 | AF173687 | AF173721 |
| Anas gracilis | | AF059076 | | AF173480 | AF173493 |
| Anas laysanensis | | AF059078 | JF498830 | | |
| Anas melleri | | AF059080 | | | |
| Anas nesiotis | | AF059057 | | AF173483 | AF173489 |
| Anas platyrhynchos | NC009684 | | | | |
| Anas poecilorhyncha | NC022418 | | | | |
| Anas rubripes | | AF059088 | AY666211 | | |
| Anas sparsa | | AF059091 | | | |
| Anas superciliosa | | AF059092 | JN801396 | AF173486 | AF173488 |
| Anas undulata | | AF059093 | | | |
| Anas zonorhyncha | | AF059095 | | | |
| Anser albifrons | NC004539 | | | | |
| Anser anser | NC011196 | | | | |
| Anser cygnoides | NC023832 | | | | |
| Anser indicus | NC025654 | | | | |
| Asarcornis scutulata | | AF059099 | | HM063548 | |
| Aythya affinis | | EU585621 | DQ434306 | | |
| Aythya americana | NC000877 | | | | |
| Aythya australis | | EU585622 | | | |
| Aythya collaris | | | DQ434322 | | |

| | | | | | |
|---|---|---|---|---|---|
| Aythya ferina | NC024602 | | | | |
| Aythya fuligula | NC024595 | | | | |
| Aythya marila | | EU585625 | DQ434331 | | |
| Aythya nyroca | | EU585626 | GQ481388 | | |
| Aythya valisineria | | | DQ434336 | | |
| Branta bernicla | NC027066 | | | | |
| Branta canadensis | NC007011 | | | | |
| Bucephala albeola | | EU585633 | DQ432777 | | |
| Bucephala clangula | | AF515261 | GU571288 | AF173712 | AF173749 |
| Bucephala islandica | | EU585635 | DQ433386 | | |
| Cairina moschata | NC010965 | | | | |
| Chenonetta jubata | | AF059100 | JN801436 | HM063545 | AF173737 |
| Chloephaga picta | | AF515262 | FJ027352 | | |
| Chloephaga poliocephala | | EU585637 | FJ027354 | | |
| Chloephaga rubidiceps | | | JN801562 | | |
| Clangula hyemalis | | EU585638 | GU571339 | | |
| Cyanochen cyanoptera | | AF059101 | | HM063550 | AF173740 |
| Cygnus atratus | NC012843 | | | | |
| Cygnus columbianus | NC017604 | | | | |
| Cygnus cygnus | NC027095 | | | | |
| Cygnus olor | NC027096 | | | | |
| Histrionicus histrionicus | | AF173766 | DQ433685 | AF173713 | AF173750 |
| Hymenolaimus malacorhynchos | | | | HM063552 | |
| Lophodytes cucullatus | | EU585650 | DQ434635 | | |
| Mareca americana | | AF059103 | DQ432718 | HM063537 | AF173722 |
| Mareca falcata | NC023352 | | | | |
| Mareca penelope | | AF059107 | GU571239 | AY164518 | |
| Mareca sibilatrix | | AF059108 | FJ027103 | | |
| Mareca strepera | | AF059109 | JN703210 | AF173689 | AF173723 |
| Marmaronetta angustirostris | | AF059104 | KP252212 | HM063551 | AF173736 |
| Melanitta perspicillata | | EU585652 | DQ434654 | | |
| Mergellus albellus | | EU585653 | GU571480 | AY164516 | |
| Mergus octosetaceus | | KM347976 | KM896444 | | |
| Mergus squamatus | NC016723 | | | | |
| Netta peposaca | | EU585656 | JN801862 | | |
| Netta rufina | NC024922 | | | | |
| Oressochen jubata | | HM063577 | KF446137 | HM063554 | AF173746 |

| | | | | | |
|---|---|---|---|---|---|
| Oressochen melanopterus | | AF173763 | | HM063555 | AF173747 |
| Polysticta stelleri | | AY351682 | GU571592 | | |
| Pteronetta hartlaubii | | AF059110 | | HM063549 | AF173739 |
| Sarkidiornis melanotos | | AF059111 | FJ028237 | HM063541 | AF173738 |
| Sibirionetta formosa | NC015482 | | | | |
| Somateria fischeri | | AY244331 | DQ434101 | U83738 | |
| Somateria mollissima | | AF515264 | AY666247 | | |
| Somateria spectabilis | | EU585662 | DQ434104 | | |
| Spatula clypeata | NC028346 | | | | |
| Spatula cyanoptera | | AF059067 | FJ027082 | AF173690 | AF173724 |
| Spatula discors | | AF059068 | AY666325 | | |
| Spatula hottenta | | AF059077 | | AF173692 | AF173726 |
| Spatula platalea | | AF059084 | FJ027098 | | |
| Spatula puna | | AF059085 | FJ027101 | | |
| Spatula querquedula | | AF059086 | GQ481326 | AF173691 | AF173725 |
| Spatula versicolor | | AF059094 | FJ027114 | | |
| Speculanas specularis | | HM063572 | | HM063534 | AF173728 |
| Tachyeres brachypterus | | HM063574 | | HM063529 | |
| Tachyeres leucocephalus | | HM063569 | | HM063530 | |
| Tachyeres patachonicus | | | JN802005 | HM063531 | |
| Tachyeres pteneres | | AF059112 | JN802006 | HM063532 | AF173730 |
| Tadorna cana | | EU585663 | | | |
| Tadorna ferruginea | NC024640 | | | | |
| Tadorna radjah | | EU585665 | | AF173708 | AF173745 |
| Tadorna tadornoides | | EU585666 | | | |
| Tadorna variegata | | AF173760 | | AF173706 | AF173743 |

**Table 3-3.** Museum records and locality information for samples sequenced in these studies.

| Species | Locality Information | Catalog Information |
|---|---|---|
| Camptorhynchus labradorius | Long Island, NY, USA | AMNH 45802 |
| Camptorhynchus labradorius | Long Island, NY, USA | AMNH 45803 |
| Chendytes lawi | San Luis Obispo County, CA, USA | SLO-2-1-S4/W2 270-280 |
| Melanitta fusca deglandi | Kings Park, Long Island, NY, USA | UCLA DC13,921 |
| Melanitta nigra | Washington State, USA | UCLA DC22,024 |
| Mergus merganser americanus | Kiona, Benton County, WA, USA | UCLA DC21412 |
| Mergus serrator | Tacoma, Pierce County, WA, USA | UCLA DC21410 |

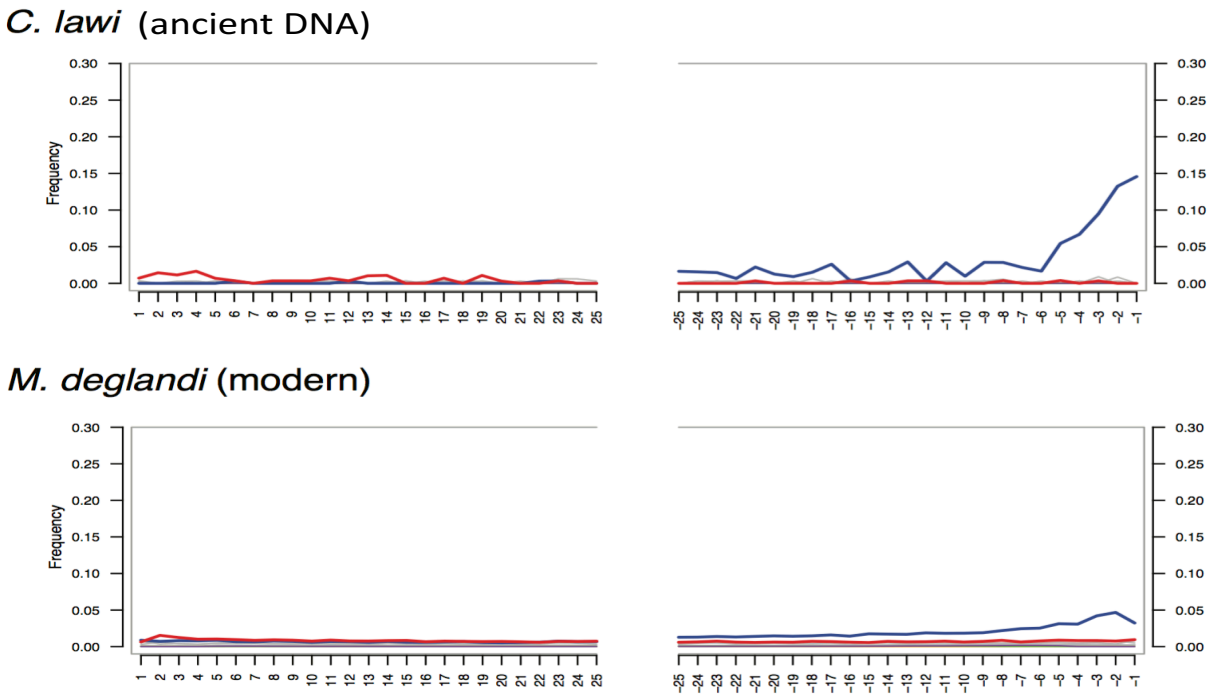


**Figure 3-3.** Nucleotide mis-incorporation plots produced using MapDamage (v2.0). Each row of plots demonstrates position-specific substitutions from the 5' (left) and the 3' (right) ends of DNA strands. Blue lines represent G to A substitutions, red lines represent C to T substitutions. These results suggest that the DNA extracted from *C. lawi* has a greater rate of 3' G to A mis-incorporation than *M. deglandi*. This supports the ancient origin and authenticity of *C. lawi* DNA.

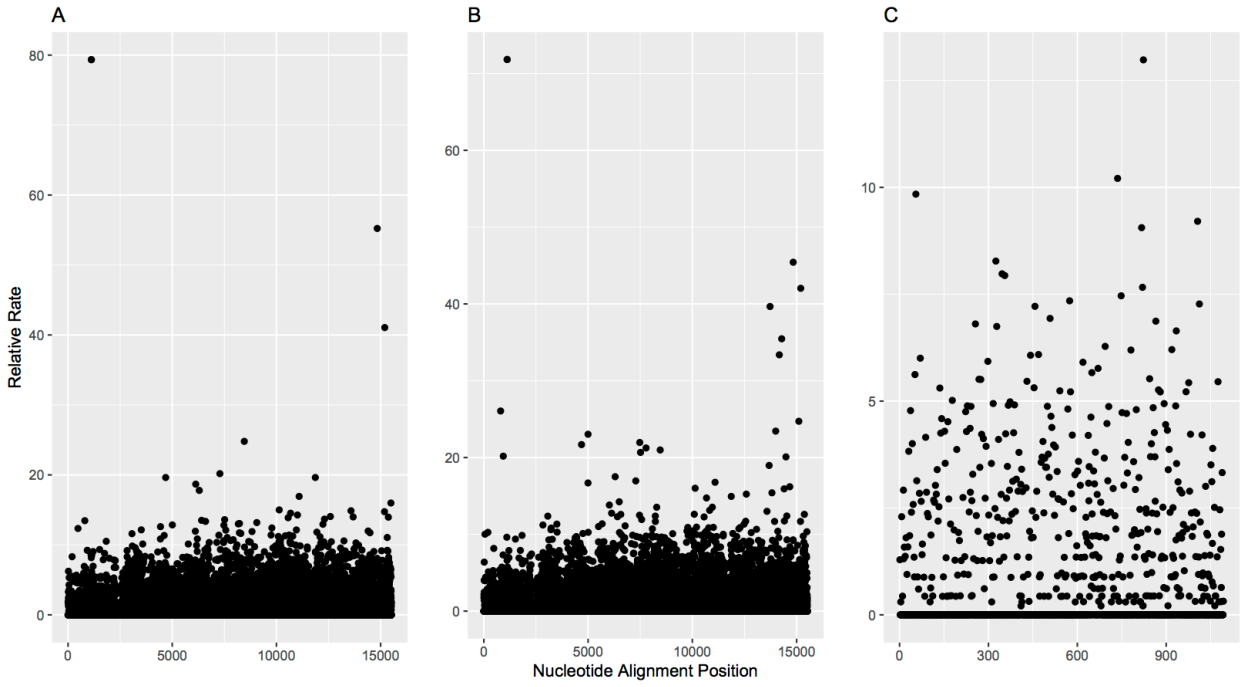


**Figure 3-4.** Scatter plots showing per-site relative rates (as estimated by HyPhy) for each nucleotide position in the DNA sequence alignments. A – Combined mitogenome and mitochondrial gene sequences; B – Mitogenomes only; C – COI and Cytochrome B only. The four sites in matrix A and five sites in matrix B with the highest rates were removed from further analyses due to their high probability of exhibiting homoplasy. The dot representing the highest rate in A is two adjacent sites (1118 and 1119), but appear as a single dot due to their proximity.

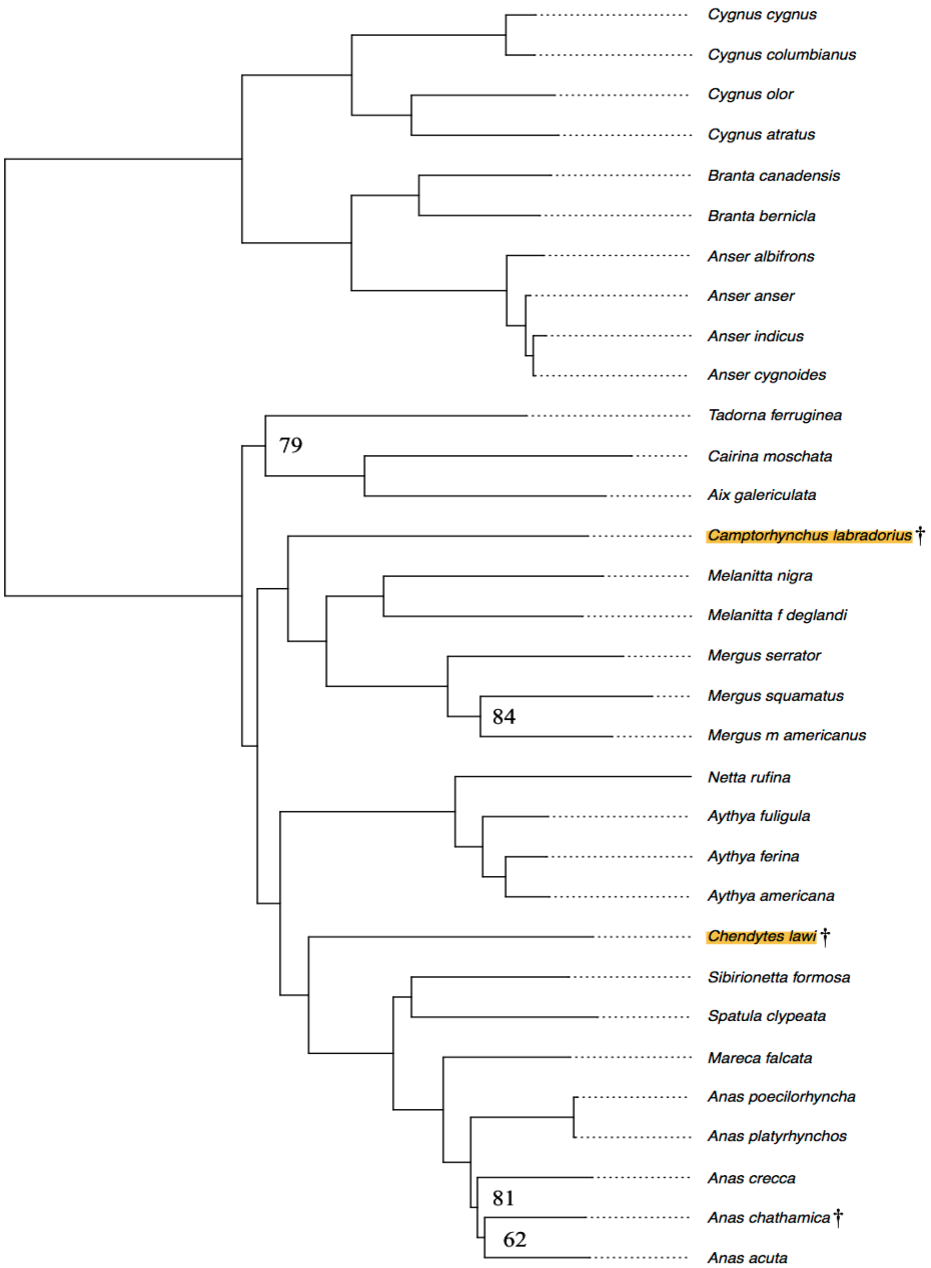


**Figure 3- 5.** Summarized Bayesian inference and maximum likelihood result based on mitogenome only DNA matrix. Bootstrap values below 90 are reported at the corresponding nodes. All nodes across the tree are supported by posterior probabilities of 0.99 or higher. The extinct taxa *Chendytes lawi* and *Camptorhynchus labradorius* are highlighted on the tree. A dagger (†) indicates and extinct species.

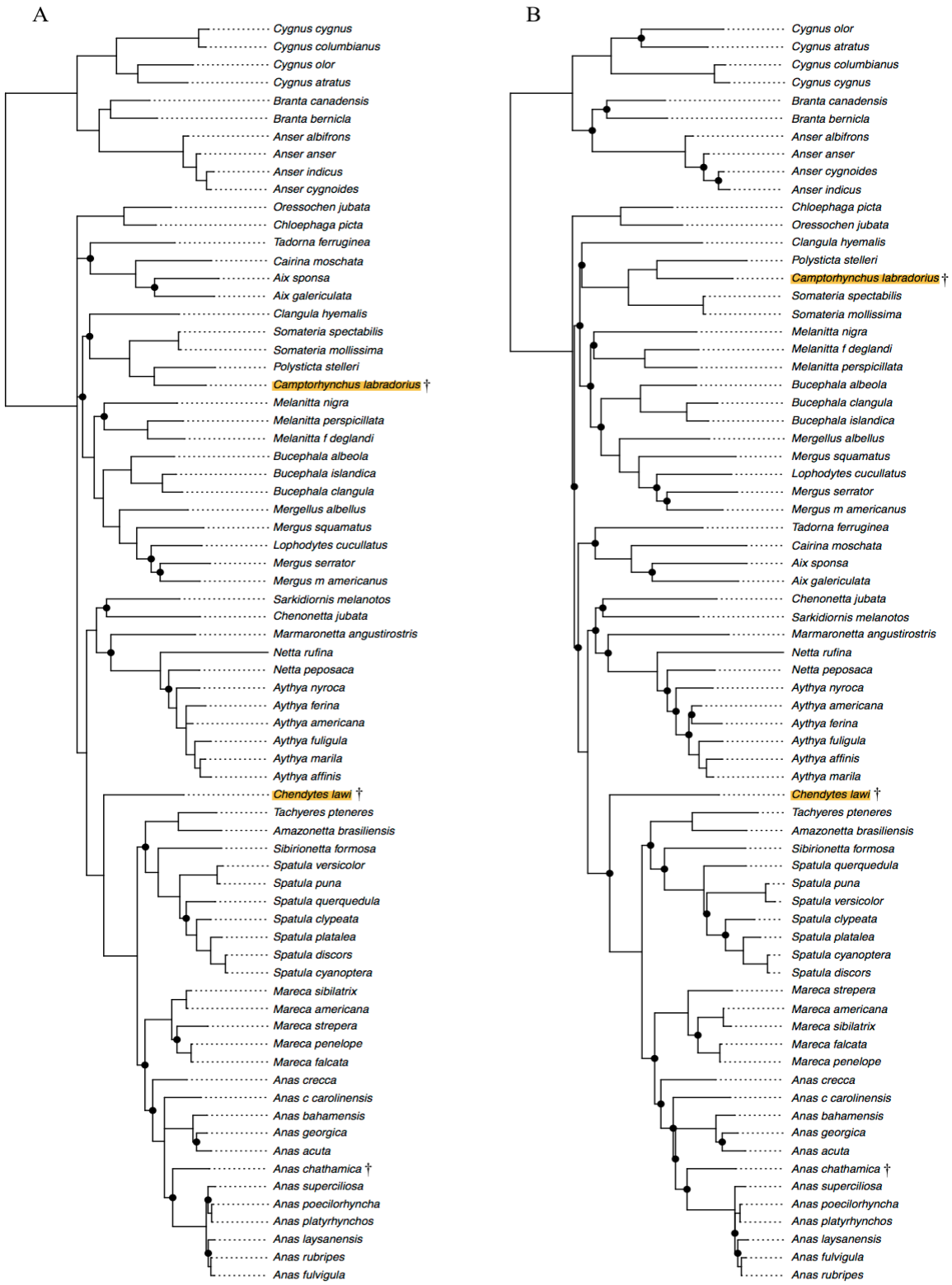


**Figure 3-6.** Bayesian inference (A) and maximum likelihood (B) trees based on reduced DNA matrix with COI and Cytochrome b sequences with no missing data. The extinct taxa Chendytes lawi and Camptorhyncus labradorius are highlighted on the tree. A dagger (†) indicates an extinct species.

**Figure 3-7.** Graphic showing the distribution of missing data across matrix A for subfamily Anatinae (the ten outgroup samples of the family Anserinae are not shown as these are all represented by complete mitogenomes). Horizontal gray bars represent the nucleotide bases represented for each sample. Horizontal black bars represent ambiguous bases and missing data for each sample.

**References**

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc\

Audubon, J.J. (1843). *The birds of America: from drawings made in the United States and their territories Vol 6.* Audubon, New York, NY & Philadlphia, PA. doi:10.5962/bhl.title.61411

Bolger, A.M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114 – 2120. doi:10.1093/bioinformatics/btu170

Bulgarella, M., Sorenson, M.D., Peters, J.L., Wilson, R.E., McCracken, K.G. (2010). Phylogenetic relationships of *Amazonetta*, *Speculanas*, *Lophonetta* and *Tachyeres*: four morphologically divergent duck genera endemic to South America. *Journal of Avian Biology, 41*(2), 186-199. doi:10.1111/j.1600-048X.2009.04819.x

Bustnes, J. O., Systad, G. H. (2001). Comparative feeding ecology of Steller's eider and long-tailed ducks in winter. *Waterbirds, 24*(3), 407-412. doi:10.2307/1522072

Bustnes, J. O., Asheim, M., Bjørn, T. H., Gabrielsen, H., Systad, G. H. (2000). The diet of Steller's Eiders wintering in Varangerfjord, northern Norway. *The Wilson Bulletin*, *112*(1), 8-13. doi:10.1676/0043-5643(2000)112[0008:TDOSSE]2.0;2

Chaves, B.R., Chaves, A.V., Nascimento, A.C., Chevitarese, J., Vaconcelos, M.F., Santos, F.R.(2015). Barcoding Neotropical birds: assessing the impact of nonmonophyly in a highly diverse group. *Molecular Ecology Resources, 15*(4), 921-931. doi: 10.1111/1755-0998.12344

Chilton, G., Sorenson, M.D. (2007). Genetic identification of eggs purportedly from the extinct Labrador Duck (*Camptorhynchus labradorius*). *The Auk 124*(3), 962-968. doi:10.1642/0004-8038(2007)124[962:GIOEPF]2.0.CO;2

Clements, J.F., Schulenberg, T.S., Iliff, M.J., Roberson, D., Fredericks, T.A., Sullivan, B.L., Wood, C.L. (2017) The eBird/Clements checklist of birds of the world: v2016. Retrieved from http://www.birds.cornell.edu/clementschecklist/download/

Cottam, C. (1939). Food habits of North Amercian ducks. *United States Department of Agriculture Technical Bulletin, 643*, 1-129.

Cottam, C., Addy, C. E. (1947). Present eelgrass condition and problems on the Atlantic coast of North America. In *Proc. 12th North Amer. Wildlife Conference.* San Antoniao, Texas (pp. 387-398).

Cracraft, J. (2013). Avian higher-level relationships and classification: nonpasseriforms. In E.C. Dickinson, J.V Remsen (Eds.), *The Howard and Moore Complete Checklist of the Birds of the World, 4th edn, Vol. 1. Non-passerines* (pp. 21-47) Eastbourne, UK: Aves Press Limited.

Ding, L., Zhang, C., Chang, Q., Yan, L., Zhang, B. (2016). The complete mitochondrial genome of Aythya fuligula (Anatidae: Aythya). *Mitochondrial DNA, 27*(2), 1003-1004. doi:10.3109/19401736.2014.926525

Donne-Gousse, C., Laudet, V., Hanni, C. (2002). A molecular phylogeny of Anseriformes based on mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution, 23*(3), 339-356. doi:10.1016/S1055-7903(02)00019-2

Dutcher, W. (1891). The Labrador Duck: A Revised List of the Extant Specimens in North America, with Some Historical Notes. *The Auk*, *8*(2), 201-216. doi:10.2307/4068077

Dutcher, W. (1894). The Labrador Duck: Another Specimen, with Additional Data Respecting Extant Specimens. *The Auk*, *11*(1), 4-12. doi:10.2307/4067622

Ellingson, R.A., Swift, C.C., Findley, L.T., Jacobs, D.K. (2014). Convergent evolution of ecomorphological adaptations in geographically isolated Bay gobies (Teleostei: Gobionellidae) of the temperate North Pacific. *Molecular Phylogenetics and Evolution, 70,* 464-477. doi:10.1016/j.ympev.2013.10.009

Faith, D.P. (1989). Homoplasy as pattern: multivariate analysis of morphological convergence in Anseriformes. *Cladistics, 5*(3), 235-258. doi:10.1111/j.1096-0031.1989.tb00488.x

Fulton, T.L., Wagner, S.M., Shapiro, B. (2012a). Case Study: Recovery of Ancient Nuclear DNA from Toe Pads of the Extinct Passenger Pigeon. In: Shapiro B., Hofreiter M. (Eds.), *Ancient DNA. Methods in Molecular Biology (Methods and Protocols), vol 840* (pp 29-35.). New York, NY: Humana Press.

Fulton, T.L., Letts, B., Shapiro, B. (2012b). Multiple losses of flight and recent speciation in steamer ducks. *Proceedings of the Royal Society B, 279*, 2339-2346. doi: 10.1098/rspb.2011.2599

Gonzalez, J., Düttman, H., Wink, M. (2009). Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *Journal of Zoology, 279*(3), 310-318. doi:10.1111/j.1469-7998.2009.00622.x

Grayson, D.K. (2008). Holocene underkill. *Proceedings National Academy of Sciences, 105*(11), 4077-4078. doi:10.1073/pnas.0801272105

Gruhn, R., Bryan, A. (2006). An Interim Report on Two Rockshelter Sites with Early Holocene Occupation in the Northern Baja California Peninsula. *Pacific Coast Archaeological Society Quarterly 42*(2), 1-16.

Guthrie, D.A. (1992). A late Pleistocene avifauna from San Miguel Island, California. *Papers on Avian Paleontology. Sci Ser, 36*, 319-327.

Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M. (2004). Identification of birds through DNA Barcodes. *PLoS Biol. 2*(10), E312. doi: 10.1271/journal.pbio.0020312.
Howard, H. (1947). Wing elements assigned to *Chendytes*. *The Condor, 49*(2), 76-77.

Howard, H. (1955). New records and a new species of *Chendytes*, an extinct genus of diving geese. *The Condor, 57*(3), 135-143. doi:10.2307/1364861

Howard, H. (1964) Futher discoveries concerning the flightless "diving geese" of the genus *Chendytes*. *The Condor, 66*(5), 372 – 376. doi:10.2307/1365429

Huelsenbeck, J.P., Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics, 17*(8), 754-755.

Jiang, F., Miao, Y., Liang, W., Ye, H., Liu, H., Liu, B. (2010). The complete mitochondrial genomes of the whistling duck (*Dendrocygna javanica*) and black swan (*Cygnus atratus*): dating evolutionary divergence in Galloanserae. *Molecular Biology Reports 37*(6), 3001-3015. doi:10.1007/s11033-009-9868-9

Johnson, K.P., Sorenson, M.D. (1998). Comparing molecular evolution in two mitochondrial protein coding genes (cytochrome b and ND2) in the dabbling ducks (Tribe: Anatini). *Molecular Phylogenetics and Evolution 10*(1), 82-94.

Jones, C.R., Schubel, J.R. (1978). Distribution of surficial sediments and eelgrass in New York's south shore bays: an assessment from the literature. *Marine Science Resources Center Spec. Rep., No. 13*, Stony Brook, New York.
Jones, T.L., Porcasi, J.F., Erlandson, J.M., Dallas, Jr. H., Wake, T.A., Schwaderer, R. (2008a). The protracted Holocene extinction of California's flightless sea duck (*Chendytes lawi*) and its implications for the Pleistocene overkill hypothesis. *Proceedings National Academy of Sciences, 105*(11), 4105-4108. doi:10.1073/pnas.0711140105

Jones, T.L., Porcasi, J.F., Gaeta, J., Codding, B.F. (2008b). Diablo Canyon Fauna: A course-grained record of trans-Holocene foraging from the central California mainland coast. *American Antiquity 73*(2), 289-316.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics, 29*(13), 1682-1684. doi:10.1093/bioinformatics/btt193

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Kennedy, M., Spencer, H.G. (2000). Phylogeny, Biogeography, and Taxonomy of Australasian Teals. *The Auk 117*(1), 154-163.

Kerr, K.C., Stoeckle, M.Y., Dove, C.J., Wigt, L.A., Francis, C.M., Hebert, P.D. (2007). Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes 7*(4), 535-543. doi:10.1111/j.1471-8286.2007.01670.x

Kerr, K.C., Lijtmaer, D.A., Barreira, A.S., Hebert, P.D., Tubaro, P.L. (2009). Probing evolutionary patterns in neotropical birds through DNA barcodes. *PLoS ONE 4*(2), E4379. doi: 10.1371/journal.pone.0004379. doi:10.1371/journal.pone.0004379

Kosakovsky Pond, S.L., Frost, S.D.W., Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics, 21*(5), 676-679. doi:10.1093/bioinformatics/bti079

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Lee, J.H., Ryu, S.H., Kang, S.G., Hwang, U.W. (2012). Complete mitochondrial genome of the Bewick's swan *Cygnus columbianus bewickii. Mitochondrial DNA 23*(2), 129-130. doi:10.3109/19401736.2011.653808

Liu, G., Zhou, L.Z, Gu, C.M. (2012). Complete sequence and gene organization of the mitochondrial genome of scaly-sided merganser (*Mergus squamatus*) and phylogeny of some Anatidae species. *Molecular Biology Reports 39*(3), 2139-2145. doi:10.1007/s11033-011-0961-5

Liu, G. Zhou, L., Li, B. Zhang, L. (2014). The complete mitochondrial genome of *Aix galericulata* and *Tadorna ferruginea*: bearings on their phylogenetic position in the Anseriformes. *PLoS ONE 9*(22), E109701. doi: 10.1371/journal.pone.0109701.

Livezey, B.C. (1989). Phylogenetic relationships and incipient flightlessness of the extinct Auckland Islands merganser. *Wilson Bulletin, 101*(3), 410-435.

Livezy, B.C. (1993). Morphology of flightlessness in *Chendytes*, fossil seaducks (Anatidae: Mergini) of coastal California. *Journal of Vertebrate Paleontology, 13*(2), 185-199.

Livezey, B.C. (1995). Phylogeney and evolutionary ecology of modern seaducks (Anatidae: Mergini). *The Condor, 97*(1), 233-255.

Lovvorn, J.R., Richman, S.E., Grebmeier, J.M., Cooper, L.W. (2003). Diet and body condition of spectacled eiders wintering in pack ice of the Bering Sea. *Polar Biology 26*(4), 259-267.

Martin, P.D., Douglas, D.C., Obritschkewitsch, T., Torrence, S. (2015). Distribution and movements of Alaska-breeding Steller's Eiders in the nonbreeding period. *The Condor*, 117(3), 341-353.

McCracken, K.G., Harshman, J., McClellan, D.A., Afton, A.D. (1999). Data set incongruence and correlated character evolution: an example of functional convergence in the hind-limbs of stifftail diving ducks. *Systematic Biology, 48*(4), 683-714.

Meng, S., Wu, Z., Pan, T., Yan, Ls, Bei, Y., Li, G., Zhang, B. (2016). Mitochondrial genome of the *Aix galericula* (Anatidae: Aix). *Mitochondrial DNA 27*(1), 318-319. doi:10.3109/19401736.2014.892096

Metzner, K.A. (1993). Ecological strategies of wintering Steller's Eider on Izembek Lagoon and Cold Bay, Alaska. Master's thesis, Univ of Missouri, Columbia.

Miller, L. (1925) *Chendytes*, a diving goose from the California Pleistocene. *The Condor, 27*, 145-147.

Miller, L. (1930). Further bird remains from the upper San Pedro Pleistocene. *The Condor, 32*(2), 116-118. doi:10.2307/1363530

Miller, L., Mitchell, E.D., Lipps, J.H. (1961). New light on the flightless goose *Chendytes lawi*. *Los Angeles County Museum Contribution in Science,* 43, 3-10.

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 Nov. 2010, New Orleans, LA pp. 1 - 8.

Mindell, D.P., Sorenson, M.D., Huddleston, C.J., Miranda, H.C. Jr., Knight, A., Sawchuk, S.J., Yuri, T. (1997). Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In D.F. Mindell (Ed.), *Avian Molecular Evolution and Systematics* (pp. 211-245). Cambridge, MA: Academic Press.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T. (1998). Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Systematic Biology, 48*(1), 138-152.

Mitchell, K.J., Wood, J.R., Scofield, R.P., Llamas, B., Cooper, A. (2014). Ancient mitochondrial genome reveals unsuspected taxonomic affinity of the extinct Chatham duck (*Pachyanas chathamica*) and resolves divergence times for New Zealand and sub-Antarctic brown teals. *Molecular Phylogenetics and Evolution 70,* 420-428. doi:10.1016/j.ympev.2013.08.017

Olsen, A.M. (2015). Exceptional avian herbivores: multiple transitions toward herbivory in the bird order Anseriformes and its correlation with body mass. *Ecology and Evolution, 5*(21), 5016-5032. doi:10.1002/ece3.1787

Ouellet, J. F., Vanpé, C., & Guillemette, M. (2013). The body size-dependent diet composition of North American sea ducks in winter. *PloS one*, *8*(6), e65667. doi:10.1371/journal.pone.0065667.

Pan, T., Ren, L., Wang, H., Chen, J., zhang, B. (2014). Mitochondrial genome of the *Anas falcata* (Anatidae: Anas). *Mitochondrial DNA 25*(2), 111-112. doi:10.3109/19401736.2013.786709.

Park, C.E., Park, G.S., Kwak, Y., Hong, S.J., Rahim Khan, A., Kwon Jung, B., Park, Y.J., Kim, J.., Cheon Park, H., Shin, J.H. (2016). Complete mitochondrial genome of *Cygnus cygnus* (Aves, Anseriformes, Anatidae). *Mitochondrial DNA, 27*(4), 2907-2908. doi:10.10.3109/19401736.2015.1060433

Pearce, J.M., Talbot, S.L., Petersen, M.R., Rearick, J.R. (2005). Limited genetic differentiation among breeding, molting, and wintering groups of the threatened Stellar's eider: the role of historic and contemporary factors. *Conservation Genetics, 6*(5), 743-757. doi:10.1007/s10592-005-9034-4

Petersen, M.R., Bustnes, J.O., Systad, G.H. (2006). Breeding and moulting locations and migration patterns of the Atlantic population of Steller's eiders *Polysticta stelleri* as determined from satellite telemetry. *Journal of Avian biology*, *37*(1), 58-68. doi:10.1111/j.2005.0908-8857.03472.x

Pisani, D. (2004). Identifying and removing fast-evolving sites using compatibility analysis: an example from the Arthropoda. *Systematic Biology, 53*(6), 978-989. doi:10.1080/1063515049088877

Raikow, R.J. (1970). The function and evolution of the supraorbital process in ducks. *The Auk, 87*(3), 568-572. doi:10.2307/4083799

Rohland, N., Hofreiter, M. (2007). Comparison and optimization of ancient DNA extraction. *BioTechniques, 42*(3), 343-352.

Rohland, N., Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research, 22*(5), 939-946. doi:10.1101/gr.128124.111

Ryu, S.H., Hwang, U.W. (2011). Complete mitochondrial genome of the Baikal teal *Anas formosa* (Aves, Anseriformes, Anatidae). *Mitochondrial DNA, 22*(4), 74-76. doi:10.3109/19401736.2011.624600

Savard, J.P.L., Petersen, M.R. (2015). Remigial molt of sea ducks. In J.L Savard, D.V. Derksen, D. Esler, J.M. Eadie (Eds.) *Ecology and Conservation of North American Sea Ducks. Studies in Avian Biology, vol. 46* (pp. 305-336). Boca Raton, FL: CRC Press.

Schindel, D.E., Stöckle, M.Y., Milensky, C., Trizna, M., Schmidt, B., Gebhard, C., Graves, G. (2011). Project description: DNA barcodes of bird species in the national museum of natural history, Smithsonian institution, USA. *Zookeys, 152*, 87-92. doi:10.3897/zookeys.152.2473

Slack, K.E., Janke, A., Penny, D., Arnason, U. (2003). Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene* 302(1-2), 43-52.

Sorenson, M.D., Cooper, A., Paxinos, E.E., Quinn, T.W., James, H.F., Olson, S.L., Fleischer, R.C. (1999). Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. *Proc. R. Soc. Lond., B, Biol. Sci. 266*(1434), 2187-2193. doi:10.1098/rspb.1999.0907

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Sun, Z., Wang, B., Sun, X., Yan, L., Pan, T., Zhang, B. (2016). Phylogenetic studies of *Anas clypeata* (Anatidae: Anas) based on complete mitochondrial DNA sequences. *Mitochondrial DNA A DNA Map Seq Anal. 27*(6), 4320-4321. doi:10.3109/19401736.2015.1089482

Sun, Z., Pan, T., Hu, C., Sun, L., Ding, H., Wang, H., Zhang, C., Jin, H., Chang, Q., Kan, X., Zhang, B. (2017). Rapid and recent diversification patterns in Anseriformes birds: Inferred from molecular phylogeny and diversification analyses. *PloS ONE 12*(9), e0184529. doi: 10.1371/journal.pone.0184529

Tavares, E.S., Gonçalves, P., Miyaki, C.Y., Baker, A.J. (2011). DNA barcode detects high genetic structure within neotropical bird species. *PLoS One 6*(12), e28543. doi: 10.1371/journal.pone.0028543.

Tu, J., Si, F., Wu, Q., Cong, B., Xing, X., Yang, F.H. (2014). The complete mitochondrial genome of the Muscovy duck, *Cairina moschata* (Anseriformes, Anatidae, Cairina). *Mitochondrial DNA 25*(2), 102-103. doi:10.3109/19401736.2013.784756

Watanabe, J., Matsuoka, H. (2015). Flightless diving ducks (Aves, Anatidae) from the Pleistocene of Shiriya, northeast Japan. *Journal of Vertebrate Paleontology 35*(6), e994745. doi:10.1080/02724634.2014.994745.

Weins, J.J. (2006). Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics, 39*(1), 34-42. doi:10.1016/j.jbi.2005.04.001

Zhou, W., Zhang, C., Pan, T, Yan, L., Hu, C., Sue, C., Can, Q., Zhang, B. (2015). The complete mitochondrial genome of *Anas poecilorhyncha* (Anatidae: Anas). *Mitochondrial DNA 26*(2), 265-266. doi:10.3109/19401736.2013.823191

Zhou, W., Zhang, C., Chang, Q., Yan, L., Pan, T., Zhang, B. (2016). The complete mitochondrial genome of *Aythya ferina* (Anatidae: Aythya). *Mitochondrial DNA A DNA Mapp Seq Anal., 27*(2), 968-969. doi:10.3109/19401736.2014.926509

Žydelis, R., Richman, S.E. (2015). Foraging behavior, ecology, and energetics of sea ducks. In J.L Savard, D.V. Derksen, D. Esler, J.M. Eadie (Eds.) *Ecology and Conservation of North American Sea Ducks. Studies in Avian Biology, vol. 46* (pp. 243-268). Boca Raton, FL: CRC Press.