



UNIVERSIDADE DE BRASÍLIA  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
DEPARTAMENTO DE GENÉTICA E  
MORFOLOGIA



PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

**Filogeografia e estrutura genética do  
Arapaçu-do-Cerrado (*Lepidocolaptes  
angustirostris*): influência dos biomas e do  
clima na diversificação da avifauna nas  
áreas abertas da América do Sul.**

**Amanda Vaz Rocha**  
**Orientador: Dr. Renato Caparroz**  
**Coorientador: Dr. Gustavo S. Cabanne**

**Brasília, Junho de 2018**

**AMANDA VAZ ROCHA**

**Filogeografia e estrutura genética do Arapaçu-  
do-Cerrado (*Lepidocolaptes angustirostris*):  
influência dos biomas e do clima na  
diversificação da avifauna nas áreas abertas da  
América do Sul.**

Tese apresentada ao Programa de Pós-Graduação em Biologia Animal da Universidade de Brasília como requisito para a obtenção do Título de Doutora em Biologia Animal.

**Orientador: Dr. Renato Caparroz**

**Brasília, Junho de 2018**

*“Porque a vida só se dá pra quem se deu, pra quem amou, pra quem chorou,  
pra quem sofreu.”*

**Vinícius de Moraes**

## AGRADECIMENTOS

- A Universidade de Brasília e ao Programa de Pós-Graduação em Biologia Animal. Todos os professores do programa assim como todos aqueles que estiveram envolvidos para que eu tivesse a oportunidade de obter esse conhecimento.
- A CAPES pela minha bolsa nos dois primeiros anos deste doutorado e pela minha bolsa de Doutorado Sanduíche; ao CNPq pela minha bolsa nos anos finais e pelo financiamento deste projeto; e a FAPDF pelo auxílio em visitas técnicas que enriqueceram meu crescimento acadêmico. Ao MINCYT da Argentina pelo apoio financeiro na obtenção dos dados genéticos deste trabalho e pelo financiamento da minha visita técnica ao Museo Argentino de Ciencias Naturales.
- Ao Laboratório de Genética e Biodiversidade da Universidade de Brasília, pela infraestrutura e por possibilitar todas as condições necessárias para o desenvolvimento nesta tese. Agradecer a todos os integrantes deste Laboratório em todos esses anos de doutorado, especialmente aos professores que fazem desse laboratório uma realidade Prof. Renato Caparroz, Prof. Fernando R. Pacheco e Prof. Lilian Giugliano além de todos os seus técnicos e alunos.
- Ao Grupo de Pesquisa da qual fiz parte por tantos anos, o grupo do Professor Renato Caparroz, agradeço a todos aqueles que fizeram parte das discussões e trabalhos de campo que sem dúvida foram fundamentais para este trabalho.
- A Banca examinadora desta tese Dr. Alexandre Aleixo, Dra. Lilian Giugliano, Dra. Silviene F. de Oliveira e Dra. Rosana Tidon por aceitarem prontamente o convite para participar desta avaliação e contribuir para o enriquecimento deste trabalho.
- Aos meus pais e irmã, que me apoiaram emocional e financeiramente em todos os anos da minha vida acadêmica por muitas vezes abrindo mão de muitas coisas em prol das minhas decisões para enriquecer este processo de aprendizagem. Meu muito obrigada sem vocês nada disso seria possível.
- Ao meu orientador de iniciação científica, de mestrado e doutorado que me formou, me orientou e mais do que tudo me deu caminhos e ferramentas para que toda minha formação acadêmica pudesse acontecer. Professor Doutor Orientador Chefe e Parainfo Renato Caparroz sem você certamente não conseguiria ter cumprido toda essa jornada, minha admiração por você é eterna e não existem palavras suficientes para te agradecer por tudo que você representa na minha vida.

- Ao Dr. Gustavo S. Cabanne meu coorientador, que por muitas vezes nos últimos anos fez papel de orientador, um exemplo de cientista e retidão, e um mais que isso, um exemplo pra mim de excelência que eu busco alcançar enquanto pesquisadora.
- Ao Dr. Brian Tilston Smith por me aceitar prontamente em seu laboratório no American Museum of Natural History, por todas as discussões, por todos os ensinamentos técnicos e científicos ao longo dos 10 meses de convivência e por acreditar e investir na Ideia deste trabalho, meu muito obrigada por toda hospitalidade em NYC.
- A todo o departamento de Ornitologia do American Museum of Natural History, pela hospitalidade, pelos inúmeros conhecimentos científicos e técnicos adquiridos neste departamento, em especial ao Dr. Joel Cracraft, Gabi, Bill, Jonas, Jessica, Glenn, Kaiya, Jon, Peter, Paul Sweet, Lucas, Lais, Luke e Tom.
- Ao Dr. Marcelo Gehara por toda atenção e contribuição no desenvolvimento das análises de cenários, as discussões, resoluções de problemas e a forma cordial e prestativa com que fui tratada. Muito obrigada Marcelo.
- Aos coautores dos trabalhos que serão fruto desta tese, Dr. Luis Fábio Silveira, Dr. Pablo Tubaro, Dr. Gustavo Cabanne, Dr. Renato Caparroz, Dra. Cassia Alves Lima, Dr. Alan Lemmon, Dra. Emilly M. Lemmon, Dr. Alexandre Aleixo e Dr. Brian T. Smith por toda contribuição na realização deste trabalho.
- A todas as pessoas que contribuíram com as amostras deste trabalho, curadores do Museu Argentino de Ciencias Naturales, do Museu Paraense Emílio Goeldi, do Museu de Zoologia da Universidade de São Paulo, do Burke Museum, todas as pessoas que contribuíram nas expedições de campo realizadas pelo Laboratório de Genética e Biodiversidade da Universidade de Brasília, aos responsáveis pelos Parques Nacionais em que coletamos, os responsáveis pelas fazendas em que coletamos, enfim a todos aqueles que contribuíram para a obtenção das amostras deste trabalho.
- Ao Dr. Alexandre Aleixo, por toda cordialidade a cada favor solicitado nesta tese que foram muitos, ao acesso indiscriminado à coleção do Museu Paraense Emílio Goeldi, o financiamento do trabalho de campo, as avaliações dos meus trabalhos durante este doutorado, as discussões e conselhos científicos e todo o apoio a mim dado, a minha eterna admiração a você, gostaria que soubesse que toda a sua atenção foi fundamental para o desenvolvimento desta tese.

- A Dra. Lilian Giugliano que me acompanha desde o mestrado, muito obrigada por toda ajuda e por todas as discussões científicas que tivemos, muito obrigada por prontamente aceitar contribuir com este trabalho, tenho uma enorme admiração por você.
- Por fim e não menos importante aos amigos que fiz, a Cássia Lima-Rezende por me ajudar em todas as etapas deste trabalho, por contribuir com o campo, com as análises, com as discussões e com o andamento deste trabalho desde o começo. Ao Bruno Almeida por toda ajuda com o campo, e com as análises biogeográficas deste trabalho. A Lucas Moreira por ter sido peça fundamental em possibilitar meu doutorado sanduiche e toda ajuda com a parte de bioinformática deste trabalho; A Mateus Ferreira pela essencial introdução à parte bioinformática desta tese e todas discussões acerca dos resultados. A Lukas Musher por todas as discussões e por me ensinar os fundamentos das análises no R que serão eternamente úteis, além de todas as ajudas com problemas de bioinformática; A Laís Coelho por me ensinar toda a parte de análise de cenários e toda ajuda com a parte de bioinformática deste trabalho. Além de toda parte científica gostaria de agradecer a vocês por fazer deste doutorado uma experiência em que levarei amigos pra minha vida. Meu eterno muito obrigada!

# SUMÁRIO

## **CAPÍTULO I - INTRODUÇÃO GERAL**

Hipóteses biogeográficas para a América do Sul	9
Soerguimento dos Andes	9
Rios como barreiras	10
Refúgios as flutuações climáticas do Pleistoceno	11
Hipótese do Arco Pleistocênico	12
Modelos de evolução da diagonal de vegetações abertas da América do Sul	12
Filogeografia	13
Sequenciamento de Nova Geração (NGS)	14
Uso dos modelos de distribuição de espécies associados à filogeografia	16
Arapaçu-do-cerrado: a espécie modelo	17
REFERÊNCIAS	22

## **CAPÍTULO II - FILOGEOGRAFIA DO ARAPAÇU-DO-CERRADO (*LEPIDOCOLAPTES ANGUSTIROSTRIS*): INSIGHTS SOBRE A DIVERSIFICAÇÃO DA AVIFAUNA NEOTROPICAL ASSOCIADA ÀS FORMAÇÕES DE VEGETAÇÕES ABERTAS**

RESUMO	30
INTRODUÇÃO	31
MATERIAL E MÉTODOS	35
RESULTADOS	51
DISCUSSÃO	64
CONCLUSÕES	67
REFERENCIAS	68

## **CAPÍTULO III - ANÁLISE GENÔMICA DA ESTRUTURA POPULACIONAL DE *LEPIDOCOLAPTES ANGUSTIROSTRIS*: A INFLUÊNCIA DAS FLUTUAÇÕES CLIMÁTICAS E DOS BIOMAS SOB A AVIFAUNA DAS ÁREAS ABERTAS DA AMÉRICA DO SUL.**

RESUMO	75
INTRODUÇÃO	76
MATERIAL E MÉTODOS	78
RESULTADOS	87
DISCUSSÃO	98
CONCLUSÕES	102
REFERÊNCIAS	103

## **CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS**

<b>PRODUÇÃO RELACIONADA:</b> Mitogenomes of two neotropical bird species and the multiple independent origin of mitochondrial gene orders in Passeriformes	109
--	-----

## CAPÍTULO I

### INTRODUÇÃO GERAL

#### *Hipóteses biogeográficas para a América do Sul*

A América do Sul abriga a maior biodiversidade do planeta (Turchetto-Zolet et al., 2013), sendo sua origem e diversificação possivelmente influenciadas pela associação de efeitos promovidos por diversos eventos geológicos e ecológicos ao longo de milhões de anos. Diversas hipóteses biogeográficas vêm sendo propostas com o objetivo de explicar os processos envolvidos na diversificação fauna da América do Sul. As hipóteses sobre os eventos geológicos consideram os eventos vicariantes como os principais agentes responsáveis pela diversificação, dentre os quais podemos destacar o soerguimento dos Andes (Chapman, 1917), os refúgios pleistocênicos (Haffer, 1969; Vanzolini & Williams, 1970), e os rios como barreiras (Wallace, 1854; Sick, 1967). No entanto, tais hipóteses tem sido mais estudadas em áreas florestais, como as Florestas Amazônica e Atlântica (Turchetto-Zolet et al., 2013). As hipóteses associadas às formações de áreas abertas (Caatinga, Cerrado e Chaco) estão fortemente relacionadas às flutuações climáticas (Turchetto-Zolet et al., 2013). Como exemplo desta influência, podemos citar a hipótese do Arco Pleistocênico, proposto na diversificação das Florestas secas decíduas e semidecíduas (Prado & Gibbs, 1993; Toby Pennington et al., 2000) e a formação dos corredores florestais nas áreas abertas que possam ter conectado áreas florestais via áreas abertas (Batalha-Filho et al., 2013; Ledo & Colli, 2017).

#### *Soerguimento dos Andes*

O soerguimento dos Andes pode ter contribuído na diversificação de aves neotropicais promovendo o isolamento geográfico entre as populações remanescentes em lados opostos da cordilheira emergente, criando um mosaico de vales e picos propícios para a ocorrência de especiação e alterando a hidrologia e o clima de todo o continente sul-americano (Chapman, 1917). Diversos estudos vêm evocando esta hipótese biogeográfica para explicar os padrões de diversidade encontrados. Sedano e Burns (2010), com base na filogenia molecular dos traupídeos, sugerem que a o tempo de diversificação destas espécies é coincidente com o soerguimento principal da parte norte dos Andes, evidenciando a importância deste evento na diversificação desta família de Aves. O soerguimento da parte central dos Andes também pode ter contribuído com a diversificação em Psittacidae, separando espécies de terras altas (Hapalopsittaca) e baixas dos Andes (Pyrilia, Quintero et al., 2012). E ainda, os Andes



formaram efetiva barreira aos ventos do Pacífico, promovendo mudança de habitats e provavelmente exercendo uma forte influência na especiação da região oriental da América do Sul (Riccardi & Rolleri, 1980). Em um trabalho com beija-flores, foram encontradas evidências da relação entre o soerguimento dos Andes e a diferenciação no gênero *Adelomyia* (Chaves et al., 2011). Além disso, o soerguimento dos Andes parece ter contribuído com a separação do complexo *Xiphorhynchus guttatus/sussurans* (Rocha et al., 2015).

### *Rios como barreiras*

Quanto à hipótese dos rios como barreiras é postulado que os rios sejam eventos vicariantes que possibilitaram processos de diversificação, Esta hipótese foi primeiramente proposta por Wallace (1854) quando propôs a existência de quatro distritos, estes abrigam espécies endêmicas de macacos, sendo as fronteiras desses distritos definidas pelos rios Amazonas, Madeira e Negro. Haffer (1997), em uma revisão sobre os modelos de especiação, aborda os efeitos dos rios Amazônicos na diversificação de aves. Trabalhos na Amazônia, avaliando o gênero *Psophia* (Ribas et al., 2011) e um complexo do gênero *Xiphorhynchus* (Aleixo, 2004) evidenciaram a forte influência da rede fluvial amazônica na diversificação destes gêneros de Aves. Esta influência dos rios pode ser observada também na Mata Atlântica (Cabanne et al., 2007, 2008). Além disso, este mesmo estudo sugere a possível influência do rio Doce na separação em *Xiphorhynchus fuscus* com base nos sinais de expansão das populações, e propõem também uma possível influência do rio São Francisco na separação de populações desta espécie. Esta influência do Rio São Francisco já é mais bem elucidada para répteis (Werneck et al., 2015) e mamíferos (Nascimento et al., 2013). Para um complexo de espécies de lagartos endêmicos da caatinga (*Tropidurus semitaeniatus*) foi identificado que o rio São Francisco atuou na diversificação do grupo, sendo que tanto o paleo curso, quanto o curso atual, participaram deste processo (Werneck et al., 2015). Já para um grupo de roedores (*Thrichomys*), o rio São Francisco parece ter contribuído com a diversificação deste grupo, porém as mudanças climáticas parecem ser o fator mais importante nesta diversificação, já que controlaram o aumento e a diminuição da vazão do rio ao longo do tempo. Em situações de alta vazão houve o impedimento do fluxo gênico, enquanto em condições de menor vazão, a formação de bancos de areia permitiram que indivíduos atravessassem o rio possibilitando o contato entre as duas margens, o que determinou o padrão de diversificação encontrado para esses pequenos mamíferos (Nascimento et al., 2013). Contudo, foi observado uma baixa estrutura populacional no gênero *Poliioptila* entre um lado e outro do atual curso do rio São Francisco (Moura et al., 2018).

### *Refúgios e as flutuações climáticas do Pleistoceno*

A hipótese dos refúgios pleistocênicos foi elaborada para explicar o elevado número de espécies endêmicas de aves (Haffer, 1969) e lagartos (Vanzolini & Williams, 1970) em determinadas regiões amazônicas. Esta hipótese propõe que mudanças climáticas e vegetacionais durante os períodos Terciário e Quaternário produziram ciclos de retração dos ambientes florestais durante os períodos frios e secos, seguido de subsequente expansão destes ambientes durante os períodos quentes e úmidos. Isto sugere que durante os períodos frios ocorreu a ruptura dos ambientes florestais gerando os “refúgios”. Estes estariam inseridos em uma matriz de vegetação aberta, isolando populações de espécies restritas a florestas, o que pode ter induzido processos de cladogênese (Haffer, 1969; Vanzolini & Williams, 1970). Além disso, há evidências de que a parte sudeste e oeste do Cerrado atuaram como refúgios importantes para espécies de plantas de savanas (Vitorino et al., 2018).

As flutuações climáticas do Pleistoceno parecem ser responsáveis pela variação do nicho de um gênero de Beija-flores (*Adelomyia*), o que promoveu em alguns clados eventos de *bottleneck* seguidos de expansões recentes, exemplificando a atuação das mudanças climáticas em aves (Chaves et al., 2011). Em um estudo com duas espécies irmãs da família Thamnophilidae: *Myrmeciza squamosa* e *Myrmeciza loucata* foi observado que o isolamento devido aos refúgios climáticos para a Mata Atlântica parece ter sido responsável tanto pela divergência basal, classificada como divergência genética profunda associada a baixo ou nenhum fluxo gênico, quanto pelos padrões de evolução intraespecíficos (Carnaval et al., 2009). E ainda, este trabalho corrobora a idéia de que as flutuações climáticas parecem ter sido importantes mesmo antes do Pleistoceno (Amaral et al., 2013).

A influência dessas flutuações também é evocada na diversificação de organismos das áreas abertas da América do Sul, como exemplo, as flutuações climáticas contribuíram para a conexão da distribuição de algumas espécies amplamente distribuídas da Caatinga, Cerrado e Chaco (Muller, 1973). As flutuações climáticas podem ter sido importantes ao favorecer o aumento da aridez e sazonalidade na Caatinga propiciando a co-expansão de répteis e anfíbios com necessidades fisiológicas e ecológicas heterogêneas (Gehara et al., 2017). O processo de expansão demográfica foi observado também em grupo de plantas no Cerrado como *Handroanthus ochraceus* e *Tabebuia aurea*, onde em períodos quentes foi observada a expansão populacional, já em períodos glaciais percebe-se a retração destes grupos (Collevatti et al., 2015; Vitorino et al., 2018).

Outra evidência do efeito das mudanças climáticas foi favorecer o intercâmbio da avifauna florestal via áreas abertas (Batalha-Filho et al., 2013). Tais corredores vêm sendo confirmados

por análises de modelagem de nicho, onde são validadas rotas de conexão pelo Cerrado em períodos glaciais e interglaciais e uma rota de conexão via Caatinga em períodos glaciais (Ledo & Colli, 2017). Além disso, a filogeografia de uma espécie de ave (*Arremon flavirostris*) confirma a existência de uma rota de conexão florestal via Cerrado durante o Último Máximo Glacial (Trujillo-Arias et al., 2017), e a similaridade da herpetofauna da Amazônia e Mata Atlântica são indícios dessa conexão entre os biomas (Prates et al., 2016).

#### *Hipótese do Arco Pleistocênico*

A hipótese do Arco Pleistocênico postula que a atual distribuição disjunta dos núcleos das florestas decíduas e semidecíduas (Seasonal dry tropical forests – SDTFs) é resultado da contração e fragmentação de uma distribuição passada mais ampla que existiu durante o último máximo glacial do Pleistoceno (Prado & Gibbs, 1993). A predição para esta hipótese é de que as espécies que hoje habitam cada um destes núcleos de SDTFs derivam de uma linhagem evolutiva única. Esta hipótese é corroborada por uma análise genética e de modelagem de nicho feita para uma espécie de planta (*Tabebuia impetiginosa*, Collevatti et al., 2012). Uma visão alternativa para a história da conexão das SDTFs, é de que esses núcleos sempre foram isolados e fragmentados e a conexão entre eles se deu por dispersão a longa distância (Mayle, 2004, 2006). Um exemplo com o linhagem de planta *Tetramerium* mostram por meio da estimativa do tempo de divergência deste grupo que a separação dos núcleos das SDTFs é anterior ao Pleistoceno, possivelmente datada do Neogeno (Cortes et al., 2015).

#### *Modelos de evolução da diagonal de formações de vegetações abertas da América do Sul*

A diagonal de formações abertas da América do Sul é composta pelas Florestas Tropicais Sazonalmente Secas (com a maior área no nordeste do Brasil: a Caatinga), o Cerrado (Brasil central), e o Chaco (nordeste da Argentina, Paraguai ocidental, e sudeste da Bolívia) (Vanzolini, 1963). Estes biomas formam um complexo mosaico de vegetação que evoluiu sob forte pressão dos períodos secos (Furley & Metcalfe, 2007). Apesar das semelhanças entre estes biomas, eles respondem de forma diferente ao clima e às mudanças ambientais e devem ser considerados separadamente em análises biogeográficas (Pennington et al., 2000).

Linhagens independentes associadas aos biomas vêm sendo identificadas em diversos grupos de animais, como por exemplo, em lagartos da espécie *Phyllopezus pollicaris*. Estudo filogeográfico mostrou uma estrutura populacional associado a cada um dos biomas pertencentes à diagonal (Werneck et al., 2012), assim como análise mitocondrial do gênero *Vanzosaura* indica a presença de três espécies distintas associadas aos biomas de áreas abertas

(Recoder et al., 2014) e em aranhas, a filogeografia da espécie *Nephila clavipes* sugerem linhagens associadas ao Chaco e ao Cerrado (Bartoletti et al., 2017).

No entanto, as hipóteses existentes sobre as relações entre os biomas não são concordantes. Estudos utilizando as ligações faunísticas (Ab'Saber, 1977; Andrade-Lima, 1982), assim como padrões morfológicos em aves, como exemplo, o padrão morfológico do complexo *Suiriri* (Hayes, 2001), sugerem que o Cerrado e a Caatinga sejam mais próximos entre si quando comparados ao Chaco. Já a filogenia morfológica do gênero *Caenomoda* sugere que o Cerrado é mais próximo do Chaco e a Caatinga ocupa uma posição basal desta relação (Zanella, 2002). Além disso, análises considerando distribuição e endemismos dos biomas sugerem que a herpetofauna do Cerrado compartilha uma história evolutiva mais recente com a herpetofauna do Chaco do que com qualquer outro bioma de área aberta (Colli, 2005), e que espécies de aves do Cerrado e do Chaco são mais próximas entre si do que quando comparadas às espécies da Caatinga (Porzecanski & Cracraft, 2005). Werneck (2011), aponta a importância dos testes dessas hipóteses do ponto de vista evolutivo já que podem representar apenas padrões de similaridade ecológica e os resultados podem ter um viés deste padrão. Como alternativa as hipóteses das relações entre as áreas devem ser testadas por métodos de filogenia/filogeografia molecular. Como exemplo desta abordagem, a filogeografia de *Phyllopezus pollicaris*, sugere o Cerrado mais próximo a Caatinga quando comparados ao Chaco (Werneck et al., 2012).

Tais hipóteses tem um grande potencial de serem testadas com uso de marcadores moleculares utilizando as abordagens filogeográficas existentes, além disso o uso de modelagens de nicho tem grande potencial de elucidar padrões biogeográficos, neste sentido apresentarei aqui uma breve discussão sobre o referencial teórico acerca da metodologia que será empregada nesta tese.

### *Filogeografia*

A filogeografia foi proposta por Avise e colaboradores (1987) com o objetivo de avaliar a distribuição da variabilidade genética no tempo e no espaço. São inúmeros os trabalhos avaliando filogeografia utilizando o DNA mitocondrial, e este fragmento foi escolhido por alguns motivos, dentre eles: o tamanho do DNA mitocondrial é pequeno quando comparado ao nuclear, ausência de íntrons, pseudogenes e DNA repetitivo, taxa de evolução rápida e herança matrilinear (Avise et al., 1987).

Estudos filogeográficos apresentam grande potencial para elucidar padrões de fluxo gênico, hibridação, expansão demográfica e especiação em diversas espécies de aves (Newton,

2003). E ainda, estes estudos podem proporcionar informações relevantes para melhor compreensão dos eventos históricos e recentes que contribuíram para a distribuição geográfica das variações fenotípicas e genéticas em diversas espécies de aves (Avise, 2000; Newton, 2003), e para a conservação da avifauna (Moritz, 2002; Frankham, 2010). Apesar da base conceitual que consolidou o campo da filogeografia não ter sido muito alterada desde a sua proposição por Avise (1987, 2000), os métodos filogeográficos sofreram modificações consideráveis em relação aos propostos no início deste campo da ciência, particularmente com o desenvolvimento da filogeografia estatística.

A filogeografia estatística que utiliza inferências demográficas e considera a estocasticidade dos processos genéticos, tem uma abordagem multidisciplinar, já que com o desenvolvimento de novas técnicas e ferramentas moleculares as análises são capazes de integrar áreas como ecologia de comunidades, climatologia, geologia e evolução em nível molecular e genômico (Knowles, 2009) Estas ferramentas em geral tem base em abordagens bayesianas e/ou verossimilhança e são capazes de estimativas populacionais do passado como tamanho populacional, taxas de fluxo gênico, tempo de divergência, processos de expansão, dentre outros (ver revisão em Excoffier & Heckel, 2006). Outra característica inovadora é o uso integrado do sistema de informação geográfica (GIS) que propiciam resultados mais acurados quanto a estrutura genética das populações, atribuição de indivíduos às populações e identificação de migrantes (Chen et al., 2007).

### *Sequenciamento de Nova Geração (NGS)*

As técnicas denominadas de sequenciamento de nova geração (NGS) tem permitido a obtenção de centenas de milhares de sequências a um custo muito baixo por base quanto comparado com o método de Sanger (Rothberg & Leamon, 2008; Shendure & Ji, 2008). Estas novas metodologias têm se mostrado muito promissoras para estudos nas áreas de genômica (Leoni et al., 2011), metagenômica (Chistoserdova, 2010) e transcriptoma (Costa et al., 2010). No entanto, ainda são pouco utilizadas em análises filogenéticas e filogeográficas devido à dificuldade de se desenvolver protocolos laboratoriais e computacionais para obter uma grande quantidade de fragmentos ortólogos de múltiplos indivíduos ou espécies que justifique o custo do sequenciamento em larga escala (Bybee et al., 2011; Faircloth et al., 2012). Como exemplo desta dificuldade, um trabalho com o sequenciamento do genoma completo de 48 espécies de aves representando todas as ordens e um sequenciamento de mais de 40 milhões de pares de base expressa a dificuldade computacional de lidar com essa quantidade de dados, no entanto os autores desenvolveram ferramentas e maneiras de analisar as hipótese de divergência das

espécies chegando a conclusão de uma diversificação rápida em um curto espaço de tempo (Jarvis et al., 2014). Uma saída para aproveitar os recursos do sequenciamento de nova geração escapando das dificuldades computacionais pode ser observado em um estudo taxonômico de duas espécies, *Rallus longirostris* e *Rallus elegans* onde o sequenciamento de nova geração foi utilizado para a identificação de regiões informativas e para o desenho de primers possibilitando o sequenciamento via Sanger (Maley & Brumfield, 2013).

Devido a estas limitações, novas abordagens foram implementadas como a utilização do ancoramento de elementos ultraconservados do genoma (Faircloth et al., 2012; Lemmon et al., 2012). As relações evolutivas das Neoaves foram estudadas pela análise destas regiões ultraconservadas do genoma com o sequenciamento de 1.541 *loci* de 32 espécies. Neste estudo, os autores afirmam que o aumento de loci estudados aumenta o suporte estatístico das relações entre os *taxa* dando maior resolução à árvore filogenética (McCormack et al., 2013). Outra abordagem adotada foi o enriquecimento de bibliotecas genômicas (Lemmon et al., 2012), que tem se mostrado muito eficientes para estudos evolutivos tornando o NGS extremamente promissor para estudos filogenéticos e filogeográficos (Lemmon & Lemmon, 2012). Como exemplo desta abordagem, uma análise filogenética foi realizada com 198 espécies de aves, com ampla representatividade das famílias e ordens, e cerca de 394 loci (Prum et al., 2015). Os resultados sugerem novos clados, assim como aumentam suporte de várias relações filogenéticas já propostas em estudos prévios. O trabalho ainda confere uma melhor resolução em ramos terminais das Neoaves, mostrando a eficiência desta técnica aplicada à filogenia (Prum et al., 2015).

Ainda que existam dificuldades laboratoriais e computacionais no uso desta metodologia para filogeografia alguns trabalhos demonstram a utilidade desta abordagem nas relações filogenéticas. Como exemplo, atualmente temos uma outra metodologia denominada RAD-Seq “*Restriction site associated DNA Sequence*” que foi utilizada para inferir a relação filogenética em um grupo de peixes e este estudo demonstra que o aumento do número de sequências utilizadas nas análises filogenéticas são capazes de produzir árvores com uma resolução sem precedentes o que demonstra o poder dos dados de nova geração para resolver até os mais difíceis desafios filogenéticos (Wagner et al., 2013). Ainda utilizando peixes como modelos de estudo, Rodríguez-Ezpeleta e colaboradores (2016) definiram a estrutura genética com suporte estatístico para *Scomber colaris* e sugerem que as estratégias de análises de dados de nova geração são diretamente aplicáveis a outras espécies. A análise de Rad-Seq combinada ao genoma de referência do *Zebrafinch* permite a identificação de SNPs “*Single nucleotide polymorphism*” ortólogos a genes conhecidos é uma possibilidade de descrever um contexto

de aves não modelos (Bourgeois et al., 2013) ampliando ainda mais o uso do sequenciamento de nova geração.

Uma abordagem de delimitação de espécies e análises demográficas de populações também vem se mostrando promissoras. Um estudo avaliando os limites das espécies definidas com base em fenótipos, genes e regiões genômicas que permaneçam diferentes ainda que em face de potenciais zonas de hibridização e introgressão sugerem que ferramentas moleculares permitem uma caracterização detalhada da diferenciação entre linhagens divergentes e definem padrões de variação nas zonas híbridas naturais (Harrison & Larson, 2014). Outro estudo integra sequenciamento de nova geração com análises morfológicas para avaliar divergências entre duas populações de salamandra (*Tribolonotus pseudoponceleti*) em ilhas do Arquipélago de Salomon e mostram como o sequenciamento de nova geração pode ser utilizado para testar cenários demográficos (Rittmeyer & Austin, 2015).

Esta nova abordagem de sequenciamento permite ainda a avaliação de locos sob seleção o que configura um grande potencial para o avanço do conhecimento dos processos de diversificação uma vez que avaliar as bases genéticas de características que afetam o valor adaptativo do indivíduo é a chave para se compreender a seleção natural (Ellegren & Sheldon, 2008).

Uma análise denominada “*approximate bayesian computation – ABC*” permite o teste de hipóteses usando complexos modelos demográficos evolutivos e dados filogenômicos (Hickerson et al., 2010). Esta análise consiste na simulação de um número de conjuntos de dados sob um modelo de interesse e o cálculo de um sumário estatístico para cada conjunto de dados simulados. O conjunto de dados simulados com o sumário estatístico que mais se aproxima ao sumário estatístico observado pode ser usado para estimar o suporte do modelo que melhor se adequa (Beaumont et al., 2002; Csilléry et al., 2010). Um exemplo da aplicação deste modelo associado a sequências de nova geração pode ser observado na análise demográfica de populações de salamandras avaliando cenários com diferentes níveis de migração (Rittmeyer & Austin, 2015).

#### *Uso dos modelos de distribuição de espécies associados à filogeografia*

A modelagem bioclimática prediz a distribuição potencial para as espécies a partir de pontos de distribuição conhecidas e projeta pontos por meio de uma interpolação considerando as superfícies climáticas de cada área (Nix, 1986). O uso da modelagem climática para prever locais de refúgios, bem como estimar a magnitude das mudanças das áreas apropriadas e testar

tais predições utilizando o DNA mitocondrial foi proposto por (Hugall e colaboradores (2002) para uma espécie de caracol endêmica da Austrália.

A MDE também pode ser aplicada de forma integrada com os diferentes métodos de análise de marcadores moleculares, assim como proposto por Hewitt (2004), segundo este autor, a associação de análises genéticas com reconstruções paleoclimáticas pode ser muito informativa sobre a evolução das espécies ao longo das oscilações climáticas do Quaternário. O autor ainda sugere que, populações em regiões onde as linhagens persistem ao longo de vários ciclos climáticos, podem acumular diferenças genéticas e possivelmente iniciar um processo de especiação (Hewitt, 2004). Buscando testar essa abordagem, um trabalho com lagartos, usou os modelos de distribuição de espécie para definir áreas estáveis para *Micrablepharus atticolus* e testar a hipótese de diferenciação genética entre áreas estáveis e não estáveis. Neste caso, não foi encontrada diferença entre áreas estáveis e não estáveis, sugerindo que a espécie tenha habilidade de se adaptar a mudanças ambientais durante as flutuações climáticas (Santos et al., 2014). Além disso, o tempo de divergência estimado para a maioria das populações é mais antigo do que os tempos disponíveis para a modelagem, o que sugere linhagens já diferenciadas que persistiram mesmo durante períodos de condições climáticas desfavoráveis (Santos et al., 2014). Este resultado é similar ao encontrado por (Werneck e colaboradores (2012) para *Phyllopezus pollicaris*, onde também não foi encontrado diferenças de variabilidade genética entre áreas estáveis e não estáveis sugerindo à capacidade da espécie de adaptação às mudanças climáticas.

Outros trabalhos usando os modelos de distribuição de espécies (MDE) tiveram como um dos objetivos testar hipóteses do efeito das variações climáticas na diversificação da biodiversidade. Para uma espécie de lagarto (*Norops meridionalis*), foi identificado um complexo de cinco linhagens, sendo que para duas destas linhagens foi testada a hipótese de conservantismo de nicho usando a modelagem de distribuição de espécie. Os dados encontrados suportam o conservantismo de nicho como principal processo de diversificação para essas duas linhagens. E ainda, os autores sugerem que o mecanismo responsável pela diversificação pode estar relacionado à inabilidade das populações de rápida adaptação a novas condições climáticas (Guarnizo et al., 2016).

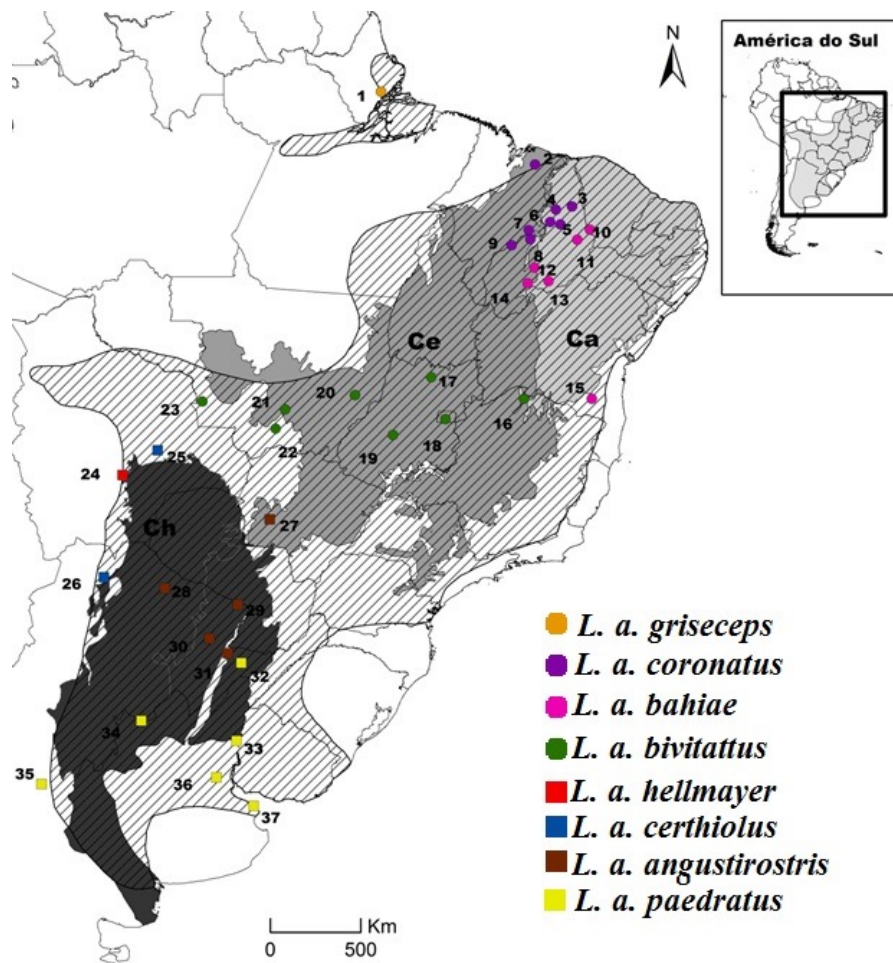
#### *Arapaçu-do-cerrado: a espécie modelo*

O arapaçu-do cerrado (*Lepidocolaptes angustirostris*, Passeriformes, Dendrocolaptidae) foi escolhido como modelo para testar hipóteses biogeográficas das áreas abertas por que possui distribuição coincidente em quase sua totalidade com a distribuição da



diagonal de formações de vegetação aberta da América do Sul, ocorrendo no Suriname, Uruguai, Argentina, Paraguai, Bolívia e Brasil (Figura 1) (Del Hoyo et al., 2003). No Brasil, esta espécie pode ser encontrada desde a Ilha do Marajó ao restante do país extra-amazônico (Sick, 1997). Tem por característica a locomoção entre árvores, voando do topo a base das árvores (Sick, 1997), tornando sua dispersão possivelmente dependente de ambientes com estrato arbóreo, o que pode ser um fator limitante para o fluxo gênico entre populações desta espécie.

Além de um bom modelo para o teste de hipótese das áreas abertas, o arapaçu-do-cerrado apresenta certa complexidade taxonômica ainda não explorada em nível genético. Atualmente, são reconhecidas oito subespécies de *L. angustirostris*, agrupadas em dois grupos com base no padrão de coloração da plumagem (Marantz et al., 2003): grupo *angustirostris*, sendo marrom no dorso e rajado no ventre (*L. a. angustirostris*, *L. a. praedatus*, *L. a. certhiolus* e *L. a. hellmayri*) e grupo *bivittatus*, sendo rufescente no dorso e sem raias no ventre (*L. a. bivittatus*, *L. a. griseiceps*, *L. a. coronatus* e *L. a. bahiae*) (Figura 1).



**Figura 1.** Distribuição geográfica de *Lepidocolaptes angustirostris* (área hachurada) com destaque para os locais de coleta das amostras de tecido que foram analisadas via sanger na presente tese. As cores representam as diferentes subespécies atualmente reconhecidas: círculos representam o grupo morfológico com ventre não estriado (*bivitattus*) e os quadrados representam o grupo morfológico com ventre estriado (*angustirostris*). Os biomas da diagonal de áreas abertas da América do Sul também aparecem em destaque: Caatinga (Ca), Cerrado (Ce) e Chaco (Ch).

Segundo filogenia proposta para o gênero *Lepidocolaptes* (Arbeláez-Cortés et al., 2012), *L. angustirostris* é a linhagem basal e apresenta baixa variabilidade intraespecífica para os genes mitocondriais citocromo c oxidase subunidade I e citocromo b. Além disso, Arbeláez-Cortés e colaboradores (2012) propõem hipótese de diversificação do gênero baseado em elementos da paisagem como a altitude (terras altas e baixas) e aos diferentes tipos de fitofisionomia (áreas abertas e florestais). Em revisão taxonômica de *L. angustirostris*, Bolívar-Leguizamón e Silveira (2015) avaliaram vários caracteres morfológicos (Figura 2) da espécie.

Após este estudo, concluiu-se que devido a ausência de caracter diagnóstico e a ausência de limites geográficos para as subespécies descritas apenas reconhece-se um táxon válido para esta linhagem. Portanto, considerando tais resultados, *L. angustirostris* é uma população contínua com variação morfológica clinal, sem distinção evolutiva fenotípica (Bolívar-Leguizamón & Silveira, 2015).



**Figura 2.** Padrão morfológico da espécie *Lepidocolaptes angustirostris*. **A.** Ilustração de quatro das oito subespécies destacando a variação do ventre exemplificando os grupos estriados e não estriados (Fonte: The Handbook Alive). **B:** Série de peles representando a variação morfológica do ventre, ressaltando a característica de variação clinal (Fonte: (Bolívar- Leguizamón & Silveira, 2015).

Espécies de arapaçus com várias subespécies descritas já vem sendo estudadas em áreas florestais. Em *Xiphorhynchus ocellatus*, as análises filogenéticas mostram várias incongruências com a atual taxonomia das subespécies, tendo em um dos casos, uma subespécie com diferenciação antiga a ponto de ser mais consistente com o nível de espécies do que de subespécie (Sousa-Neves et al., 2013). Ainda com o gênero *Xiphorhynchus*, três subespécies de *X. guttatus* não correspondem a clados diferentes, mostrando que não há

diferenciação genética suficiente para a separação em subespécies (Rocha et al., 2015).

Até o presente momento, nenhum estudo genético foi realizado buscando compreender quais fatores podem ter contribuído para o estabelecimento do atual padrão de distribuição das diferentes subespécies de *L. angustirostris* e se as unidades taxonomicamente reconhecidas podem ser realmente consideradas subespécies ou elevadas ao nível de espécie à luz do conceito unificado de espécie (unified species concept, De Queiroz, 2007). Este conceito de espécie sugere tratar linhagens que evoluíram de forma independente como espécies, quando evidências independentes estão disponíveis e demonstram isolamento genético recíproco entre estas linhagens.

Considerando os aspectos apresentados anteriormente esta tese está organizada em três partes, as quais são descritas abaixo:

- No capítulo II, utilizando um fragmento mitocondrial e um nuclear buscamos testar três hipóteses não mutuamente excludentes: (I) Hipótese do Arco Pleistocênico, (II) Áreas de estabilidade climática e (III) isolamento por distância. Por fim, avaliamos se a estrutura filogeográfica está relacionada a variação morfológica e as subespécies atualmente reconhecidas.
- No capítulo III, partimos de uma matriz de loci obtidos por sequenciamento de nova geração visamos testar também três hipóteses não excludentes: (I) cada bioma corresponde a uma unidade biogeográfica distinta; (II) os biomas respondem de forma diferente às variações climáticas e (III) considerando a ampla área dentro de cada bioma, é esperado encontrar isolamento por distância dentro de cada população.
- Por fim, apresento uma produção relacionada a esta tese, onde foram descritos os mitogenomas de duas espécies neotropicais (*Sicalis olivascens* e *Lepidocolaptes angustirostris*) com arranjos de genes diferentes. Ainda neste artigo, foi estimado o quanto frequente são os eventos de duplicação que ocorreram no genoma mitocondrial ao longo da evolução dos Passeriformes e se as duas cópias da região controladora do DNA mitocondrial demonstram um padrão de evolução em concerto (concerted evolution) na superfamília Sylvioidea. Este manuscrito será apresentado ao final da tese como "Produção relacionada".

## REFERENCIAS

- Ab'Saber, A.N. (1977) Espaços ocupados pela expansão dos climas secos na América do Sul, por ocasião dos períodos glaciais quaternários. *Paleoclimas*, **3**, 1–19.
- Aleixo, A. (2004) Historical diversification of a terra-firme forest bird superspecies: a phylogeographic perspective on the role of different hypotheses of Amazonian diversification. *Evolution*, **58**, 1303–1317.
- Amaral, F.R. do, Albers, P.K., Edwards, S. V., & Miyaki, C.Y. (2013) Multilocus tests of Pleistocene refugia and ancient divergence in a pair of Atlantic Forest antbirds (Myrmeciza). *Molecular Ecology*, **22**, 3996–4013.
- Andrade-Lima, D. (1982) Present-day forest refuges in northeastern Brazil. *Biological Diversification in the Tropics*. (ed. by G.T. Prance), pp. 245–251. Columbia University Press, New York.
- Arbeláez-Cortés, E., Navarro-Sigüenza, A.G., & García-Moreno, J. (2012) Phylogeny of woodcreepers of the genus *Lepidocolaptes* (Aves, Furnariidae), a widespread Neotropical taxon. *Zoologica Scripta*, **41**, 363–373.
- Avise, J.C. (2000) *Phylogeography*. Harvard University Press, Cambridge.
- Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., & Saunders, N.C. (1987) Intraspecific Phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.
- Bartoletti, L.F. de M., Peres, E.A., Sobral-Souza, T., Fontes, F. von H.M., Silva, M.J. da, & Solferini, V.N. (2017) Phylogeography of the dry vegetation endemic species *Nephila sexpunctata* (Araneae: Araneidae) suggests recent expansion of the Neotropical Dry Diagonal. *Journal of Biogeography*, 1–14.
- Batalha-Filho, H., Fjeldså, J., Fabre, P.H., & Miyaki, C.Y. (2013) Verbindungen zwischen atlantischen und amazonischen Waldvogelfaunen spiegeln distinkte historische Ereignisse wider. *Journal of Ornithology*, **154**, 41–50.
- Beaumont, M.A., Zhang, W., & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Bolívar-Leguizamón, S. & Silveira, L.S. (2015) Morphological variation and taxonomy of *Lepidocolaptes angustirostris* (Vieillot, 1818) (Passeriformes: Dendrocolaptidae). *Papéis avulsos de Zoologia*, **55**, 281–316.
- Bourgeois, Y.X.C., Lhuillier, E., Cézard, T., Bertrand, J.A.M., Delahaie, B., Cornuault, J.,

- Duval, T., Bouchez, O., Milá, B., & Thébaud, C. (2013) Mass production of SNP markers in a nonmodel passerine bird through RAD sequencing and contig mapping to the zebra finch genome. *Molecular Ecology Resources*, **13**, 899–907.
- Bybee, S.M., Bracken-Grissom, H.D., Hermansen, R.A., Clement, M.J., Crandall, K.A., & Felder, D.L. (2011) Directed next generation sequencing for phylogenetics: An example using *Decapoda* (Crustacea). *Zoologischer Anzeiger - A Journal of Comparative Zoology*, **250**, 497–506.
- Cabanne, G.S., D’Horta, F., Sari, E., Santos, F., & Miyaki, C. (2008) Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): Biogeography and systematics implications. *Molecular Phylogenetics and Evolution*, **49**, 760–773.
- Cabanne, G.S., Santos, F.R., & Miyaki, C.Y. (2007) Phylogeography of *Xiphorhynchus fuscus* (Passeriformes, Dendrocolaptidae): Vicariance and recent demographic expansion in southern Atlantic forest. *Biological Journal of the Linnean Society*, **91**, 73–84.
- Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T., & Moritz, C. (2009) Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. *Science*, **323**, 785–789.
- Chapman, F.M. (1917) The distribution of bird-life in Colombia: a contribution to a biological survey of South America. *Bulletin of the American Museum of Natural History*, **36**, 1–728.
- Chaves, J. a., Weir, J.T., & Smith, T.B. (2011) Diversification in *Adelomyia* hummingbirds follows Andean uplift. *Molecular Ecology*, **20**, 4564–4576.
- Chen, C., Durand, E., Forbes, F., & François, O. (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.
- Chistoserdova, L. (2010) Recent progress and new challenges in metagenomics for biotechnology. *Biotechnology Letters*, **32**, 1351–1359.
- Collevatti, R.G., Terribile, L.C., Lima-Ribeiro, M.S., Nabout, J.C., De Oliveira, G., Rangel, T.F., Rabelo, S.G., & Diniz-Filho, J.A.F. (2012) A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology*, **21**, 5845–5863.
- Collevatti, R.G., Terribile, L.C., Rabelo, S.G., & Lima-Ribeiro, M.S. (2015) Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a

- Neotropical savanna tree species in the deeper Quaternary. *Frontiers in Plant Science*, **6**, 1–15.
- Colli, G.R. (2005) As origens e a diversificação da herpetofauna do Cerrado. *Cerrado: Ecologia, Biodiversidade e Conservação*. (ed. by A. Scariot, J.C. Sousa-Silva, and J.M. Felfili), pp. 249–264. Ministério do Meio Ambiente, Brasília, Distrito Federal.
- Cortes, A.L.A., Rapini, A., & Daniel, T.F. (2015) The *Tetramerium* lineage (Acanthaceae: Justiceae) does not support the Pleistocene Arc hypothesis for South American seasonally dry forests. *American Journal of Botany*, **102**, 992–1007.
- Costa, V., Angelini, C., De Feis, I., & Ciccodicola, A. (2010) Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine and Biotechnology*, **2010**, .
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., & François, O. (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410–418.
- Ellegren, H. & Sheldon, B.C. (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169–75.
- Excoffier, L. & Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nature reviews. Genetics*, **7**, 745–758.
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.
- Frankham, R. (2010) Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, **143**, 1919–1927.
- Furley, P.A. & Metcalfe, S.E. (2007) Dynamic changes in savanna and seasonally dry vegetation through time. *Progress in Physical Geography*, **31**, 633–642.
- Gehara, M., Garda, A.A., Werneck, F.P., Oliveira, E.F., da Fonseca, E.M., Camurugi, F., Magalhães, F. de M., Lanna, F.M., Sites, J.W., Marques, R., Silveira-Filho, R., São Pedro, V.A., Colli, G.R., Costa, G.C., & Burbrink, F.T. (2017) Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. *Molecular Ecology*, **26**, 4756–4771.
- Guarnizo, C.E., Werneck, F.P., Giugliano, L.G., Santos, M.G., Fenker, J., Sousa, L., D'Angiolella, A.B., Santos, A.R., Strüssmann, C., Rodrigues, M.T., Dorado-Rodrigues, T.F., Gamble, T., & Colli, G.R. (2016) Cryptic lineages and diversification of an endemic anole lizard (Squamata, Dactyloidae) of the Cerrado hotspot. *Molecular Phylogenetics and Evolution*, **94**, 279–289.
- Haffer, J. (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.

- Haffer, J. (1997) Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Conservation*, **6**, 451–476.
- Harrison, R.G. & Larson, E.L. (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.
- Hayes, F.E. (2001) Geographic variation, hybridization, and the leapfrog pattern of evolution in the Suiriri flycatcher (*Suiriri suiriri*) complex. *The Auk*, **118**, 457–471.
- Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.
- Hickerson, M.J., Carstens, B.C., Cavender-Bares, J., Crandall, K.A., Graham, C.H., Johnson, J.B., Rissler, L., Victoriano, P.F., & Yoder, A.D. (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.
- Del Hoyo, J., Elliot, A., & Christie, D.A. (2003) *Handbook of the birds of the world*. Broadbills to Tapaculo. Lynx Edicions, Barcelona.
- Hugall, A., Moritz, C., Moussalli, A., & Stanisic, J. (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6112–6117.
- Jarvis, E.D., Mirarab, S., Aberer, A.J., et al. (2014) Whole-genome analyses resolve early branches in the tree of life of Modern Birds. *Science*, **346**, 1126–1138.
- Knowles, L.L. (2009) Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593–612.
- Ledo, R.M.D. & Colli, G.R. (2017) The historical connections between the Amazon and the Atlantic Forest revisited. *Journal of Biogeography*, .
- Lemmon, A.R., Emme, S., & Lemmon, E.C. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.
- Lemmon, A.R. & Lemmon, E.C. (2012) High-throughput development of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology*, **61**, 745–761.
- Leoni, C., Volpicella, M., De Leo, F., Gallerani, R., & Ceci, L.R. (2011) Genome walking in eukaryotes. *Febs Journal*, **278**, 3953–3977.
- Maley, J.M. & Brumfield, R.T. (2013) Mitochondrial and Next-Generation Sequence Data Used to Infer Phylogenetic Relationships and Species Limits in the Clapper/King Rail Complex. *The Condor*, **115**, 316–329.



- Marantz, C., Aleixo, A., Bevier, L., & Patten, M. (2003) Available at:  
<http://www.hbw.com/species/narrow-billed-woodcreeper-lepidocolaptes-angustirostris>.
- Mayle, F.E. (2004) Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. *Journal of Quaternary Science*, **19**, 713–720.
- Mayle, F.E. (2006) The Late Quaternary Biogeographical History of South American Seasonally Dry Tropical Forests: Insights from Palaeo-Ecological Data. *Neotropical Savannas and Seasonally Dry Forests Plant Diversity, Biogeography, and Conservation*, 395–416.
- McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., & Brumfield, R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.
- Moritz, C. (2002) Strategies to protect biological diversity and the processes that sustain it. *Systematic Biology*, **51**, 238–254.
- Moura, C.C. de M., Araujo, H.F.P. de, Aleixo, A., Wink, M., & Fernandes, A.M. (2018) Accepted Article The role of landscape change and paleoclimatic events in shaping the evolutionary history of the *Polioptila Gnatcatchers* (Passeriformes, Polioptilidae) with emphasis on species associated with open habitats. *Journal of Avian Biology*, **49**, .
- Muller, P. (1973) The dispersal centers of terrestrial vertebrates in the Neotropical realm: a study in the evolution of the Neotropical biota and its native landscapes. *Biogeographica*, **2**, 1–244.
- Nascimento, F.F., Lazar, A., Menezes, A.N., Durans, A.D.M., Moreira, J.C., Salazar-Bravo, J., D'Andrea, P.S., & Bonvicino, C.R. (2013) The role of historical barriers in the diversification processes in open vegetation formations during the Miocene/Pliocene using an ancient rodent lineage as a model. *PloS one*, **8**, e61924.
- Newton, I. (2003) *The Speciation and Biogeography of birds*. Academic Press, San Diego.
- Nix, H.A. (1986) A Biogeographic Analysis of Australian Elapid Snakes. *Atlas of Elapid Snakes of Australia*. (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Pennington, R.T., Prado, D.E., & Pedry, C.A. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261–273.
- Porzecanski, A.L. & Cracraft, J. (2005) Cladistic analysis of distribution and endemism (CADE): using raw distribution of birds to unravel the biogeography of the South American aridlands. *Journal of Biogeography*, **32**, 261–275.

- Prado, D.E. & Gibbs, P.E. (1993) Patterns of Species Distributions in the Dry Seasonal Forests of South America. *Annals of the Missouri Botanical Garden*, **80**, 902.
- Prates, I., Xue, A.T., Brown, J.L., Alvarado-Serrano, D.F., Rodrigues, M.T., Hickerson, M.J., & Carnaval, A.C. (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences*, **113**, 7978–7985.
- Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., & Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, .
- De Queiroz, K. (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.
- Quintero, E., Ribas, C.C., & Cracraft, J. (2012) The Andean Hapalopsittaca parrots (Psittacidae: Aves): an example of montane-tropical lowland vicariance. *Zoologica Scripta*, **42**, 28–43.
- Recoder, R.S., De Pinho Werneck, F., Teixeira, M., Colli, G.R., Sites, J.W., & Rodrigues, M.T. (2014) Geographic variation and systematic review of the lizard genus *Vanzosaura* (Squamata, Gymnophthalmidae), with the description of a new species. *Zoological Journal of the Linnean Society*, **171**, 206–225.
- Ribas, C.C., Aleixo, A., Nogueira, a. C.R., Miyaki, C.Y., & Cracraft, J. (2011) A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 681–689.
- Riccardi, A.C. & Rolleri, E.O. (1980) Cordillera Patagônia Austral. *Geologia Regional Argentina. Segundo Simpósio* pp. 1173–1306. Academic Nacional de Ciência, Córdoba Argentina.
- Rittmeyer, E.N. & Austin, C.C. (2015) Combined next-generation sequencing and morphology reveal fine-scale speciation in *Crocodile Skinks* (Squamata: Scincidae: Tribolonotus). *Molecular Ecology*, **24**, 466–483.
- Rocha, T.C., Sequeira, F., Aleixo, A., Rêgo, P.S., Sampaio, I., Schneider, H., & Vallinoto, M. (2015) Molecular phylogeny and diversification of a widespread Neotropical rainforest bird group : The Buff-throated Woodcreeper complex , *Xiphorhynchus guttatus / susurrans* ( Aves : Dendrocolaptidae). *Molecular Phylogenetics and Evolution*, **85**, 131–140.
- Rodríguez-Ezpeleta, N., Bradbury, I.R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X.

- (2016) Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: Effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, 991–1001.
- Rothberg, J.M. & Leamon, J.H. (2008) The development and impact of 454 sequencing. *Nature Biotechnology*, **26**, 1117–1124.
- Santos, M.G., Nogueira, C., Giugliano, L.G., & Colli, G.R. (2014) Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, **41**, 1506–1519.
- Sedano, R.E. & Burns, K.J. (2010) Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). *Journal of Biogeography*, **37**, 325–343.
- Shendure, J. & Ji, H. (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.
- Sick, H. (1967) Rios e enchentes na Amazônia como obstáculo para a avifauna. *Ata Simpósio sobre a biota Amazônica*, **5**, 495–520.
- Sick, H. (1997) *Ornitologia Brasileira*. Editora Nova Fronteira, Rio de Janeiro.
- Sousa-Neves, T., Aleixo, A., & Sequeira, F. (2013) Cryptic patterns of diversification of a widespread Amazonian woodcreeper species complex (Aves: Dendrocolaptidae) inferred from multilocus phylogenetic analysis: implications for historical biogeography and taxonomy. *Molecular phylogenetics and evolution*, **68**, 410–24.
- Toby Pennington, R., Prado, D.E., & Pendry, C.A. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261–273.
- Trujillo-Arias, N., Dantas, G.P.M., Arbeláez-Cortés, E., Naoki, K., Gómez, M.I., Santos, F.R., Miyaki, C.Y., Aleixo, A., Tubaro, P.L., & Cabanne, G.S. (2017) The niche and phylogeography of a passerine reveal the history of biological diversification between the Andean and the Atlantic forests. *Molecular Phylogenetics and Evolution*, **112**, 107–121.
- Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F., & Palma-Silva, C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.
- Vanzolini, P.E. (1963) Problemas faunísticos do Cerrado. *Ferri, M. (editor) Simpósio sobre o Cerrado. Editora da Universidade de São Paulo. São Paulo*
- Vanzolini, P.E. & Williams, E.E. (1970) South America anoles: geographic differentiation and evolution of the *Anolis chrysolepsis* species group (Sauria, Iguanidae). *Arquivos de*

- zoologia (São Paulo)*, **19**, 1–298.
- Vitorino, L.C., Lima-Ribeiro, M.S., Terribile, L.C., & Collevatti, R.G. (2018) Demographical expansion of *Handroanthus ochraceus* in the Cerrado during the Quaternary: Implications for the genetic diversity of Neotropical trees. *Biological Journal of the Linnean Society*, **123**, 561–577.
- Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., & Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
- Wallace, A.R. (1854) On the Monkeys of the Amazon. *Journal of Natural History*, **Series 2**, 451–454.
- Werneck, F.P. (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.
- Werneck, F.P., Gamble, T., Colli, G.R., Rodrigues, M.T., & Sites, J.W. (2012) Deep diversification and long-term persistence in the south american “dry diagonal”: Integrating continent-wide phylogeography and distribution modeling of geckos. *Evolution*, **66**, 3014–3034.
- Werneck, F.P., Leite, R.N., Geurgas, S.R., & Rodrigues, M.T. (2015) Biogeographic history and cryptic diversity of saxicolous Tropiduridae lizards endemic to the semiarid Caatinga. *BMC Evolutionary Biology*, **15**, 94.
- Zanella, F.C. (2002) Systematics and biogeography of the bee genus *Caenonomada* Asmead, 1899 (Hymenoptera: apidae: Tapinotaspidini). *Studies on Neotropical Fauna and Environment*, **37**, 249–261.

## CAPÍTULO II

# FILOGEOGRAFIA DO ARAPAÇU-DO-CERRADO (*LEPIDOCOLAPTES ANGUSTIROSTRIS*): INSIGHTS SOBRE A DIVERSIFICAÇÃO DA AVIFAUNA NEOTROPICAL ASSOCIADA ÀS FORMAÇÕES DE VEGETAÇÕES ABERTAS

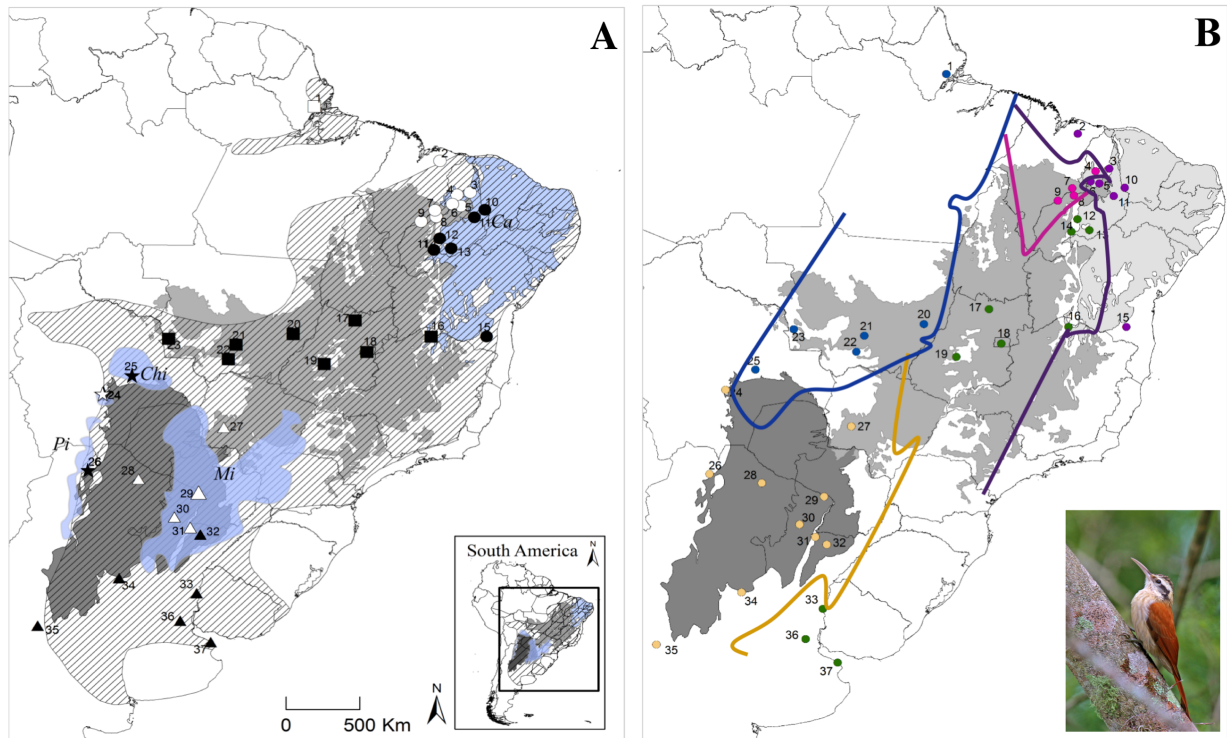
### RESUMO

Com base em análises filogeográficas e de modelagem de nicho de *Lepidocolaptes angustirostris*, nós avaliamos as predições de algumas hipóteses de diversificação relacionadas as vegetações de áreas abertas da América do Sul: (I) A hipótese do Arco Pleistocênico (PAH), (II) maior variabilidade genética em áreas climaticamente estáveis, e (III) correlação de distância genética e geográfica. Por fim, nós avaliamos se cada subespécie representa uma entidade evolutiva independente. Nós amostramos 63 indivíduos incluindo todas as subespécies descritas para *L. angustirostris*, e usamos sequências de fragmentos mitocondrial (ND2) e um nuclear (FIB5). Nós primeiramente identificamos as populações genéticas e então reconstruímos as relações filogenéticas e estimamos o tempo de divergência entre elas. Nós também realizamos testes demográficos para cada população. Além disso, nós estimamos o nicho ecológico de *L. angustirostris* ao longo do tempo sobrepondo as projeções de nicho provenientes de cada um dos quatro cenários climáticos (atual, Holoceno médio, Ultimo Máximo Glacial e Ultimo Interglacial) usando 105 pontos de ocorrência desta espécie. *Lepidocolaptes angustirostris* apresentou forte estrutura filogeográfica. Cada população dos principais tipos de formação de áreas abertas da América do Sul não são monofiléticas, ao invés disso, consistem em linhagens parafiléticas que aparentemente realizaram eventos de dispersão a curta distância entre biomas próximos. A diversidade genética estimada não foi maior em áreas climaticamente estáveis quando comparadas a áreas não estáveis. O Isolamento por distância por si só não é capaz de explicar a estrutura filogeográfica de *L. angustirostris*. Por fim, as linhagens genéticas encontradas neste estudo não correspondem as subespécies descritas para *L. angustirostris*. A diversificação em *L. angustirostris* deve ter sido influenciada primariamente pelos processos que moldaram individualmente cada tipo das principais formações de vegetação aberta da América do Sul. Com isso, nossos dados reforçam que os diferentes biomas da diagonal de formações de vegetação aberta devem ser considerados unidades biogeográficas independentes.

## INTRODUÇÃO

A região neotropical é a região mais biodiversa do mundo (Rull, 2008; IUCN, 2017). Muitos modelos e processos tem sido postulado para entender a evolução de tal alto nível de diversidade, mas muitos deles são focados apenas nas florestas tropicais, como os modelos de rios como barreiras (Aleixo, 2004; Cabanne et al., 2007, 2008; Ribas et al., 2012), o soerguimento da cordilheira dos Andes (Chapman, 1917; Sedano & Burns, 2010) e o efeito das alterações climáticas do Pleistoceno (Carnaval et al., 2009). O estudo dos processos de diversificação atuando nas matas secas e nas formações de vegetação abertas neotropicais continuam incipientes (Werneck, 2011; Turchetto-Zolet et al., 2013).

As formações de vegetação aberta (ou diagonal de formações abertas segundo Vanzolini, 1963) do Neotrópico incluem as Florestas secas semidecíduas (SDTFs daqui em diante), o Cerrado (Savana tropical do Brasil central) e o Chaco (Florestas secas do nordeste da Argentina, parte mais oeste do Paraguai e a parte sudeste da Bolívia; Fig 1a). As SDTFs apresentam uma distribuição altamente fragmentada e podem ser representadas por quatro núcleos principais: (I) Caatinga, a qual é a maior área localizada no nordeste do Brasil; (II) Misiones, localizada em uma área ao longo do sistema de rios Paraguai-Paraná.; (III) Piedmont, no sudoeste da Bolívia e noroeste da Argentina; e (IV) Chiquitano, na província Boliviana de Chiquitano (Pinsky et al., 2010). As formações de vegetação aberta formam um mosaico complexo de vegetação que aparentemente evoluíram sob forte pressão nos períodos secos (Furley & Metcalfe, 2007). Entretanto, eles parecem ter respondido de maneira diferente as mudanças climáticas e do ambiente e devem ser analisados separadamente em análises biogeográficas (Pennington et al., 2000).



**Figura 1A.** Distribuição geográfica das amostras de tecido mostrando as subespécies utilizadas neste estudo. Os quadrados brancos - *Lepidocolaptes angustirostris griseiceps*; círculos brancos - *Lepidocolaptes a coronatus*; círculos pretos - *L. a. bahiae*; quadrados pretos - *L. a. bivittatus*; estrelas pretas - *L. a. certhiolus*; estrelas brancas - *L. a. hellmayri*; triângulos brancos - *L. a. angustirostris*; triângulos pretos - *L. a. praedatus*. Os números correspondem as localidades descritas na Tabela 1. A área hachurada representa a distribuição de *Lepidocolaptes angustirostris* segundo a Birdlife (2015). Os biomas que compõem as formações de vegetação aberta da América do Sul são apresentados pelas cores: Cerrado - cinza e Chaco - cinza escuro. As áreas azuis representam a distribuição dos núcleos das SDTFs: Caatinga (Ca), Missiones (Mi), Chiquitanos (Ch) e Piedmont (Pi) (segundo Pennigton et al 2000). **B.** Populações genéticas de *L. angustirostris* (círculos coloridos) e as barreiras genéticas (linhas coloridas) identificadas utilizando o algoritmo de Monmonier com base no ND2. Círculos roxos indicam população MW\_Caatinga, círculos azuis indicam a população W\_Cerrado, círculos rosas indicam a população W\_Caatinga, círculos amarelos indicam a população do Chaco e os círculos verdes indicam a população E\_Cerrado.

A diversificação biológica das formações de vegetação aberta tem sido associada a uma combinação de flutuações climáticas (veja Werneck, 2011) e a eventos paleo geológicos do Terciário (ex. transgressões marinhas e o soerguimento do platô do Planalto Central). Em geral, as oscilações climáticas têm sido associadas a diferenciação interespecífica de plantas (Caetano et al., 2008; Ramos et al., 2009), insetos (Moraes et al., 2009) e aves (Silva & Bates, 2002), enquanto os eventos geológicos têm sido associados a diversificação em níveis taxonômicos mais altos, como gênero de lagartos (ex. Giugliano et al., 2007).

Prado e Gibbs (1993) postularam que a atual distribuição disjunta dos núcleos das SDTFs são resultado da contração e fragmentação de uma distribuição ampla das SDTFs no passado que existia durante o máximo glacial no Pleistoceno, a então chamada hipótese do Arco Pleistocênico (PAH). A PAH prediz que as espécies/populações compartilhadas por essas florestas secas devam ser derivadas de uma única linhagem evolutiva, e elas não devem apresentar estrutura genética, ou alternativamente devam apresentar apenas estrutura filogeográfica recente (originada no Pleistoceno) coincidente com cada um dos núcleos das matas secas. Por outro lado, outros estudos sugerem que os núcleos das matas secas sempre foram fragmentados e isolados, com similaridades faunísticas e florísticas como resultado de dispersão por longa distância (Mayle, 2004, 2006). Neste caso, é esperada uma única linhagem evolutiva com estrutura genética coincidente com cada núcleo, no entanto a divergência deve ter ocorrido antes do Pleistoceno. Por exemplo, um estudo com uma planta do gênero *Tetrameriu* sugere que os núcleos das SDTFs estão isolados desde o Neogeno (Cortes et al., 2015), enquanto um estudo com modelagem de nicho das SDTFs sugere que os núcleos têm estado isolados desde o Terciário superior (Werneck et al., 2011).

As oscilações climáticas do Quaternário também podem ter sido um importante mecanismo para explicar a evolução das populações nos outros biomas de formações de vegetação neotropical. Ainda que o modelo tenha sido primeiramente para explicar a diversificação no hemisfério norte (Hewitt, 2004), ele foi também considerado para a região Neotropical (Carnaval et al., 2009). De acordo com esse modelo, paisagens instáveis climaticamente podem ter moldado a demografia e o fluxo gênico, o que pode ter resultado na acumulação de diferenças genéticas, levando a uma alta diversidade genética em áreas estáveis comparada a áreas não estáveis (Hewitt, 2004; Carnaval et al., 2009). A longo prazo, isso pode levar as populações estáveis à especiação (Haffer, 1969; Hewitt, 2004). Regiões estáveis ao longo em cada um dos biomas possibilitam altos níveis de diversidade genética, enquanto áreas instáveis possibilitam a perda de variabilidade genética devido a deriva genética. Desta maneira, as populações mais diversas geneticamente dos organismos associados às formações



de vegetação aberta devam estar associadas as regiões mais estáveis climaticamente em cada um dos biomas.

Por fim, um outro fator que pode atuar na estrutura genética intraespecífica nas formações de vegetação aberta é o isolamento por distância. Este é um processo neutro relacionado a restrição da dispersão em longas distâncias que podem existir em uma paisagem aberta, onde a distância genética aumenta assim que a distância geográfica aumenta (Wright, 1943). Este processo pode ocorrer em populações subdivididas onde o fluxo gênico é dependente da distância de dispersão ou dentro de uma população com distribuição contínua onde a dispersão é espacialmente restrita (Hardy & Vekemans, 1999). A predição de isolamento por distância é a forte correlação entre a diferença genética e a distância geográfica, com baixa estrutura genética.

Sendo assim, nós estudamos o Arapaçu-do-Cerrado (*Lepidocolaptes angustirostris*), uma ave que ocorre em florestas secas, semidecíduas e bordas de matas de galeria no centro da América do Sul (Figura 1a.) para avaliar as predições das hipóteses anteriormente mencionadas sobre a diversificação da avifauna associada ad formações de vegetação aberta. O Arapaçu-do-Cerrado tem uma distribuição típica da “diagonal de formações abertas”, ocorrendo em formações de vegetação aberta e em algumas manchas de áreas abertas na parte mais a leste da Amazônia, o que representa um modelo ideal para investigar processos evolutivos nestas regiões. Existem oito subespécies neste grupo de Arapaçus agrupados em dois grupos com base no padrão de cor da plumagem: (I) grupo *angustirostris*, completamente marrom escuro em no dorso e muito estriado no ventre (*L. a. angustirostris*, *L. a. praedatus*, *L. a. certhiolus* and *L. a. hellmayri*) e (II) grupo *bivittatus*, mais rufiscente no dorso e predominantemente não estriado no ventre (*L. a. bivittatus*, *L. a. griseiceps*, *L. a. coronatus* and *L. a. bahiae*) (Marantz et al., 2003). Estas subespécies aparentemente não representam linhagens evolutivas independentes nem têm caracteres fenotípicos diagnósticos (Bolívar-Leguizamón & Silveira, 2015). No entanto, até agora nenhum estudo genético foi realizado para avaliar se as subespécies representam populações genéticas distintas ou unidades evolutivas independentes.

O objetivo principal deste trabalho foi realiza a filogeografia e a análise de modelagem de nicho de *Lepidocolaptes angustirostris* para avaliar as predições das hipóteses que possam ter atuado na diversificação das formações de vegetações abertas. Quanto a (I) Hipótese do arco pleistocênico (PAH) - nos esperamos uma única linhagem evolutiva ou espécies irmãs entre os núcleos das SDTFs com uma estrutura genética originada durante o Pleistoceno; quanto as (II) áreas de estabilidade climática, nós esperamos encontrar altos níveis de diversidade genética nas populações que ocorrem em áreas climaticamente estáveis quando

comparadas as áreas instáveis; e no que se refere ao (III) isolamento por distância, nos esperamos encontrar forte correlação entre a distância genética e geográfica dentro das populações. Por fim, nós avaliamos se a estrutura filogeográfica está relacionada à variação morfológica e às designações de subespécies.

## **MATERIAL E MÉTODOS**

### *Amostragem e dados genéticos*

Nós amostramos 63 indivíduos incluindo todas as subespécies descritas para *L. angustirostris* provenientes de 37 localidades distribuídas ao longo das formações de vegetação aberta da América do Sul (Figura 1a, Tabela 1). Nós amostramos de um a quatro indivíduos por cada localidade. Algumas amostras de tecido foram coletadas em expedições de campo pela equipe do Laboratório de genética e biodiversidade da Universidade de Brasília, e outras foram provenientes de empréstimos do Museu Argentino de Ciencias Naturales (MACN), Museu Paraense Emilio Goeldi (MPEG), Museu de Zoologia da Universidade de São Paulo (MZUSP) e Burke Museum of Natural History, and Culture (UWBM).

O DNA total foi extraído de tecidos de musculo e sangue utilizando o kit comercial Quiagen DNAeasy, seguindo as recomendações do fabricante. Nós amplificamos um fragmento mitocondrial da subunidade 2 do gene NADH desidrogenase (ND2) utilizando os iniciadores LMET (Groth, J. *apud* Ribas et al., 2005) e H6313 (Sorenson et al., 1999), e um fragmento nuclear do íntron 5 do gene beta-fibrinogênio (FIB5) utilizando os iniciadores FIB5 e FIB6 (Marini & Hackett, 2002). Os produtos de PCR foram purificados utilizando reação de EXO/SAP. As sequencias foram obtidas por empresas terceirizadas sendo elas MACROGEN Inc., Unidad de Genómica at Argentina e Centro de Genômica de Alto Desempenho na Universidade Católica de Brasília. Nós avaliamos as sequencias quanto a qualidade (clareza dos picos) e presença de ambiguidades com o auxilio do programa Geneious 6.1.6 (Biomatters). O alinhamento foi realizado pelo algoritmo Clustal W (Thompson et al., 1997). Para inferir a fase haplotípica do loco nuclear, nós utilizamos o algoritmo bayesiano PHASE (Stephens et al., 2001) implementado no programa DNAsp (Rozas et al., 2003). Nós escolhemos o melhor modelo de evolução para loco com base nos valores do critério de informação bayesiana (BIC) utilizando o programa MEGA 6.0 (Tamura et al., 2013).

**Table 1.** Amostras genéticas. LOC corresponde ao Número no mapa de amostragem genética (veja Figura 1). A palavra “origem” indica a origem do tecido: LGBUNB – Laboratório de Genética e Biodiversidade da Universidade de Brasília – Brasil; MPEG – Museu Paraneze Emílio Goeldi – Brasil; MUZUSP – Museu de Zoologia da Universidade de São Paulo – Brasil; UWBM - Burke Museum of Natural History, and Culture – USA; MACN – Museo Argentino de Ciencias Naturales – Argentina. As coordenadas geográficas estão apresentadas em graus decimais. “Clima” refere-se a estabilidade climática da localidade considerando a modelagem de nicho, onde S significa Área Estável e NS significa Área Não-estável. Genbank refere-se ao número de acesso da sequência no Genbank, o número superior indica a sequência do gene ND2 sequence e o número inferior à sequência do gene FIB5.

LOC	Amostra	Localidade	Cidade - Estado	País	Origem	LAT	LONG	Clima	Genbank
1	UFG 4011	Faz. Embrapa	Macapá - AP	Brasil	LGBUNB	0.4	-51.05	NS	MG459021 MG459086 MG459022
1	UFG 4038	Faz. Embrapa	Macapá - AP	Brasil	LGBUNB	0.4	-51.05	NS	MG459087 MG459023
1	UFG 4048	Faz. Embrapa	Macapá - AP	Brasil	LGBUNB	0.4	-51.05	NS	MG459088 MG459024
1	UFG 4094	Faz. Embrapa	Macapá - AP	Brasil	LGBUNB	0.4	-51.05	NS	MG459089 MG459025
2	UFG 2606	Fazenda Monte Carlo	Urbano Santos - MA	Brasil	MPEG	-3.21	-43.39	S	MG459090 MG459026
3	UFG 2600	Fazenda Bonito	Castelo de Piauí - PI	Brasil	MPEG	-5.31	-41.55	NS	MG459091 MG459027
4	UFG 2611	Santa Tereza	Beneditinos - PI	Brasil	MPEG	-5.46	-42.35	NS	MG459092 MG459028
5	UFG 2609	Fazenda Caluba	Elesbão Veloso - PI	Brasil	MPEG	-6.2	-42.12	NS	MG459093 MG459029
6	UFG 2608	Chapada dos Macedos	Jardim do Mulato - PI	Brasil	MPEG	-6.08	-42.63	NS	MG459094 MG459030
7	UFG 2601	Mancha verde	São João dos Patos - MA	Brasil	MPEG	-6.49	-43.7	S	MG459095

8	UFG 3913	Fazenda São Pedro	Guadalupe - PI	Brasil	MPEG	-6.93	-43.61	NS	MG459031 MG459096
9	UFG 2615	Vale do Rio Pratinha	Uruçuí - PI	Brasil	MPEG	-7.23	-44.54	S	MG459032 MG459097
10	UFG 3916	Fazenda Arsênio	Parambú - CE	Brasil	MPEG	-6.46	-40.65	S	MG459033 MG459098
11	UFG 3914	Comunidade Salseiro	Bocaina - PI	Brasil	MPEG	-6.96	-41.28	NS	MG459034 MG459099
12	UFG 3919	P.N. Serra das Confusões	Canto do Buriti - PI	Brasil	MPEG	-8.35	-43.4	NS	MG459035 MG459100
13	UFG 3917	P.N. Serra da Capivara	São Raimundo Nonato - PI	Brasil	MPEG	-9.02	-42.7	NS	MG459036 MG459101
14	UFG 3918	P.N. Serra das Confusões	Caracol - PI	Brasil	MPEG	-9.12	-43.75	NS	MG459037 MG459102
15	UFG 3227	Barra do Choça	Bahia	Brasil	MUZUSP	-14.8	-40.55	S	MG459038 MG459103
16	UFG 1052	Lagoa do Cajoeiro	Matias Cardoso - MG	Brasil	LGBU	-14.8	-43.93	S	MG459039 MG459104
16	UFG 1053	Lagoa do Cajoeiro	Matias Cardoso - MG	Brasil	LGBU	-14.8	-43.93	S	MG459040 MG459105
17	UFG 3228	Campinaçu	Goiás	Brasil	MUZUSP	-13.7	-48.56	S	MG459041 MG459106
18	UFG 3685	E. E. Águas Emendadas	Brasília - DF	Brasil	LGBUNB	-15.8	-47.84	S	MG459042 MG459107
18	UFG 3686	E. E. Águas Emendadas	Brasília - DF	Brasil	LGBUNB	-15.8	-47.84	S	MG459043 MG459108
18	UFG 2405	Reserva Ecológica do IBGE	Brasília - DF	Brasil	LGBUNB	-15.8	-47.84	S	MG459044 MG459109
19	UFG 0438	Fazenda Morro Branco	Aurilândia - GO	Brasil	LGBUNB	-16.6	-50.45	S	MG459045 MG459110
20	UFG 2978	Fazenda Ouro e Prata	Nova Xavantina - MT	Brasil	LGBUNB	-14.6	-52.35	NS	MG459046 MG459111



27	UFG 4307	Fazenda Pitangueiras	Bonito - MS	Brasil	LGBUNB	-20.8	-56.58	NS	MG459063 MG459128 MG459064
28	MACN 5399	Reserva Natural Formosa	Província de Formosa	Argentina	MACN	-24.3	-61.8	S	MG459129 MG459065
28	MACN 5434	Reserva Natural Formosa Parque Nacional	Província de Formosa	Argentina	MACN	-24.3	-61.8	S	MG459130 MG459066
29	MACN 3155	Pilcomayo Parque Nacional	Província de Formosa	Argentina	MACN	-25.1	-58.17	S	MG459131 MG459067
29	MACN 3191	Pilcomayo Parque Nacional	Província de Formosa	Argentina	MACN	-25.1	-58.17	S	MG459132 MG459068
29	MACN 3663	Pilcomayo	Província de Formosa	Argentina	MACN	-25.1	-58.17	S	MG459133 MG459069
30	MACN 4503	Parque Nacional Chaco	Província de Chaco	Argentina	MACN	-26.8	-59.6	S	MG459134 MG459070
30	MACN 4506	Parque Nacional Chaco	Província de Chaco	Argentina	MACN	-26.8	-59.6	S	MG459135 MG459071
30	MACN 4508	Parque Nacional Chaco	Província de Chaco	Argentina	MACN	-26.8	-59.6	S	MG459136 MG459072
31	MACN 1357	EBCO	Província de Corrientes	Argentina	MACN	-27.5	-58.68	S	MG459137 MG459073
31	MACN 1404	EBCO Parque Nacional	Província de Corrientes	Argentina	MACN	-27.5	-58.68	S	MG459138 MG459074
32	MACN 4991	Mburucuya Parque Nacional El	Província de Corrientes	Argentina	MACN	-28	-58.01	S	MG459139 MG459075
33	MACN 4339	Palmar	Província de Entre Rios	Argentina	MACN	-31.9	-58.24	NS	MG459140 MG459076
34	MACN 6134	La Para	Província de Córdoba	Argentina	MACN	-30.8	-62.98	S	MG459141 MG459077
34	MACN 6150	La Para	Província de Córdoba	Argentina	MACN	-30.8	-62.98	S	MG459142 MG459078
34	MACN 5314	La Para	Província de Córdoba	Argentina	MACN	-30.8	-62.98	S	MG459143

35	MACN 4077	Nacunan Santa Rosa	Provincia de Mendoza	Argentina	MACN	-34	-67.95	NS	MG459079
									MG459144
									MG459080
35	MACN 4086	Nacunan Santa Rosa	Provincia de Mendoza	Argentina	MACN	-34	-67.95	NS	MG459145
									MG459081
36	MACN 2117	Isla las Palmas Baradero	Provincia de Buenos Aires	Argentina	MACN	-33.7	-59.25	NS	MG459146
									MG459082
36	MACN 2136	Isla las Palmas Baradero	Provincia de Buenos Aires	Argentina	MACN	-33.7	-59.25	NS	MG459147
									MG459083
37	MACN 4252	Reserva El destino	Provincia de Buenos Aires	Argentina	MACN	-35.1	-57.39	NS	MG459148
									MG459084
OG	CVC 017	San Lorenzo	Departamento Santa Cruz	Bolivia	UWBM	-19.4	-63.47		MG459149
									MG459085
OG	CVC 019	San Lorenzo	Departamento Santa Cruz	Bolivia	UWBM	-19.4	-63.47		MG459150

### *Populações genéticas e estrutura genética*

Nós definimos as populações para as análises com base em barreiras genéticas estimadas pelo algoritmo de Monmonier e pelos valores de  $\Phi_{sc}$  como descrito por Rodríguez-Robles e colaboradores (2010). Brevemente, as barreiras genéticas foram identificadas com base nas distâncias genéticas do gene ND2 utilizando o algoritmo de Monmonier implementado no programa AIS 1.0 (Miller, 2005). As barreiras foram plotadas uma a uma até que o número máximo de barreiras ( $K_{max}$ ) foi atingido quando uma barreira isolou uma única localidade como população. Quando o número máximo de barreiras foi atingido, nós definimos os números possíveis de populações. Depois, nós realizamos uma análise de variância molecular (AMOVA, Excoffier et al., 1992) para estimar os índices de estrutura genética ( $\Phi_{sc}$ ,  $\Phi_{st}$  and  $\Phi_{ct}$ ) com base em uma matriz de distância utilizando o modelo de Tajima e Nei (1984) para o número de populações ( $k$ ) variando de dois ao valor máximo ( $k_{max}$ ) previamente definidos pela análise anterior. A significância estatística foi calculada com base em 1.000 permutações implementadas no programa Arlequin 3.0 (Excoffier et al., 2005). Por fim, nos escolhemos o melhor número de populações avaliando o menor valor positivo do índice  $\Phi_{sc}$ . Então, todas as populações genéticas foram caracterizadas com base no número de haplótipos, diversidade haplotípica, diversidade nucleotídica e as diferenças nucleotídicas para ambos os marcadores utilizando DNAsp. A estrutura genética de *L. angustirostris* foi avaliada com base no valor do  $\Phi_{st}$  global (Wright, 1951). Nós também estimamos a diferença genética par a par entre as populações a fim de identificar as populações mais divergentes entre si utilizando o programa Arlequin 3.0.

### *Análises filogeográficas*

A relação entre os haplótipos recuperados foi investigada por meio de uma rede de haplótipos utilizando o modelo *median-joining* (Bandelt et al., 1999) com o auxílio do programa POPART 1.7.2 (French et al., 2013). Nós também avaliamos a relação filogenética entre os haplótipos com base em uma árvore filogenética construída por uma inferência bayesiana (BI) utilizando uma matriz concatenada de ND2 (modelo de substituição de HKY) e FIB5 (modelo de substituição de JC69). A BI foi realizada sob um modelo de “yule process” no programa Beast 1.8 (Drummond & Rambaut, 2007) utilizando os códigos IUPAC, e considerando as ambiguidades como informativas para o loco FIB5. Nós também realizamos uma análise de “species tree” no programa STARBeast (Heled & Drummond, 2010) também sob o modelo de “yule process” e um “piecewise” modelo de tamanho de população constante utilizando a fase haplotípica para o FIB5. Em ambas análises, realizamos as análises de MCMC



com 100 milhões de gerações amostradas a cada 10 mil gerações. Nós avaliamos a qualidade das corridas considerando os valores de ESS (>200) utilizando o programa Tracer 1.6 (Rambaut et al., 2014). As árvores foram sumarizadas no programa TreeAnnotator considerando um *burn-in* de 10%. As topologias foram visualizadas e editadas no programa Figtree v. 1.4.2 (disponível em <http://tree.bio.ed.ac.uk/software/figtree/>).

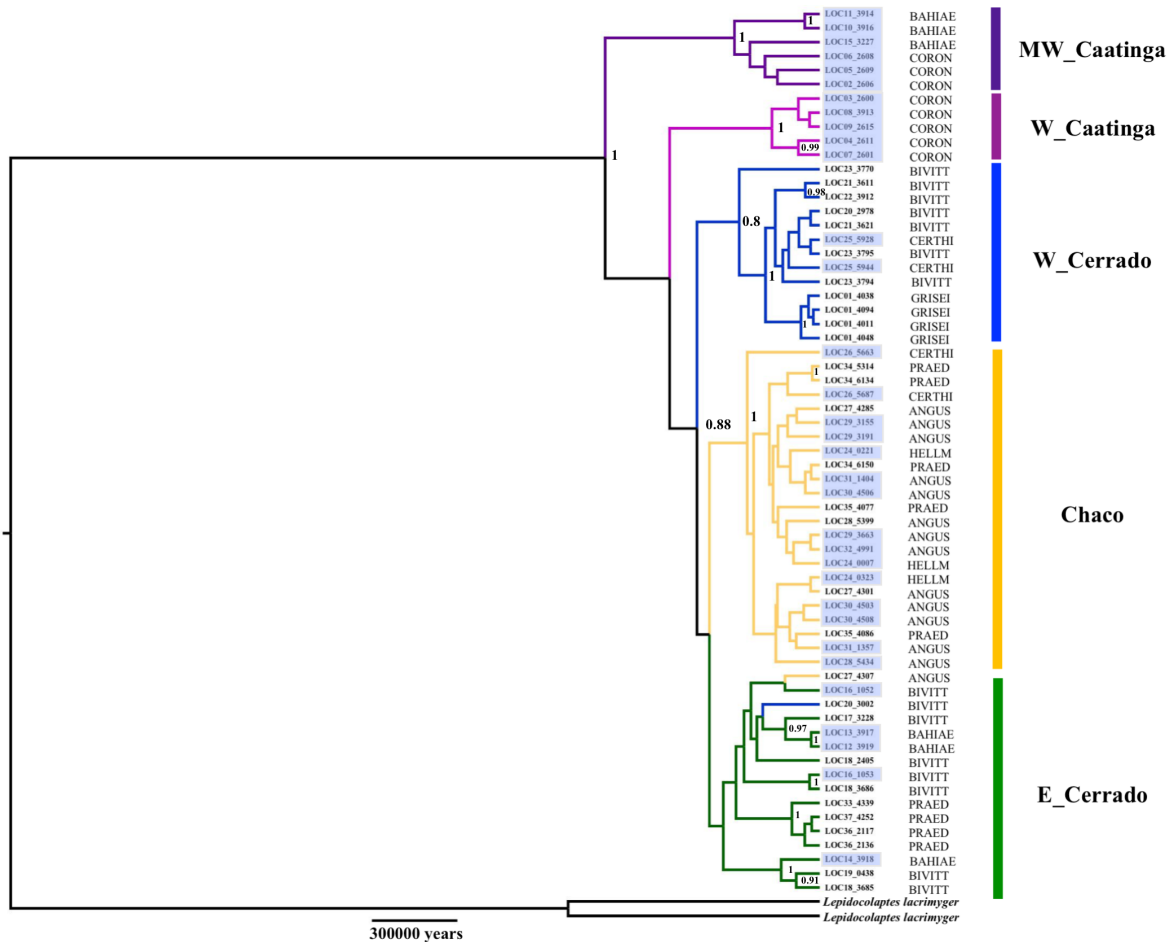
### *Demografia histórica*

Para avaliar a variação temporal no tamanho da população, nós realizamos os seguintes testes estatísticos: a estatística do  $R_2$  (Ramos-Onsins & Rozas, 2002), Tajima's D (Tajima, 1989) and Fu's FS (Fu, 1997) para cada população utilizando o programa DNAsp. Nós também avaliamos a dinâmica do tamanho efetivo populacional ao longo do tempo através da análise de *extended Bayesian skyline plots* (EBSP) realizadas para cada população (Heled & Drummond, 2008). Nós realizamos 100 mil corridas e os parâmetros foram salvos a cada mil corridas. Nós avaliamos a qualidade da corrida utilizando o programa Tracer 1.6 através dos valores de ESS (>200). O resultado gráfico é a correlação da mediana do tamanho efetivo populacional em função do tempo em anos e o intervalo de confiança foram os limites superior e inferior dos valores de 95% HPD de cada população. Para todas as análises nós utilizamos a taxa de 0.009 substituições/sítio/milhão de anos (Pacheco et al., 2011) e 0.0036 substituições/sítio/milhão de anos (Axelsson et al., 2004) como taxas de substituição para ND2 e FIB5 respectivamente.

### *Tempo de divergência*

Com o intuito de estimar o tempo de divergência entre as populações, nós realizamos uma análise de isolamento com migração (IM) assumindo a relação entre as populações obtidas pela topologia da inferência bayesiana (veja resultados, Figura 2), onde cada população ancestral foi considerada idêntica em tamanho a sua população descendente de maior tamanho, e apenas um único parâmetro de migração foi considerado para todos os pares de populações (com o intuito de simplificar o modelo) implementado no programa IMA2 (Hey & Nielsen, 2004). Nós utilizamos um tempo de corrida do tipo *floating-point* (as análises correram até 22 milhões de gerações), um milhão de gerações foram descartadas por *burn-in*, o valor do parâmetro de máximo tempo de separação das populações foi 15, o valor do parâmetro de migração foi 60, o valor do parâmetro de máximo tamanho populacional foi 50. Nós utilizamos um modelo de aquecimento geométrico das cadeias e 0.9 foi o parâmetro de aquecimento da

primeira e da segunda cadeia. Nós assumimos um ano como tempo de geração. Para estimar os valores em unidades demográficas nós utilizamos a média geométrica das taxas de mutação.

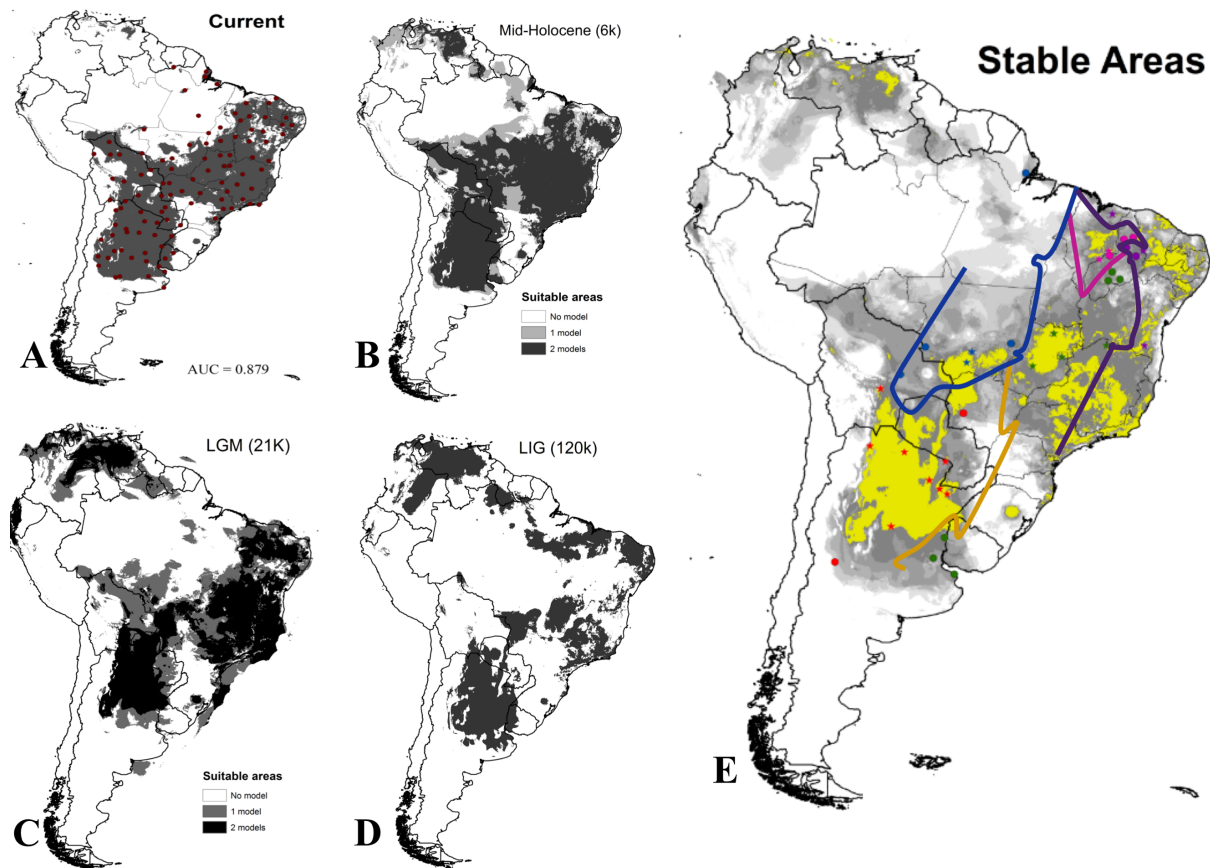


**Figura 2.** Árvore filogenética com base na inferência bayesiana mostrando as relações das linhagens utilizando os marcadores ND2 e FIB 5. Os valores de suporte em cada nó são apresentados sempre que as probabilidades a posteriori foram maiores que 0.8. As subespécies de cada amostra foram descritas como GRISEI – *Lepidocolaptes angustirostris griseiceps*; CORON – *L. a. coronatus*; BAHIAE – *L. a. bahiae*; BIVITT – *L. a. bivittatus*; CERTHI – *L. a. certhiolus*; HELLM – *L. a. hellmayri*; ANGUS – *L. a. angustirostris*; PRAE – *L. a. praedatus*. As cores correspondem as populações definidas com base na análise de estrutura genética: roxo: MW\_Caatinga; azul: W\_Cerrado; rosa: W\_Caatinga; amarelo: Chaco e verde: E\_Cerrado. As amostras dos núcleos das SDTFs foram destacadas em azul.

### *Modelagem de nicho*

Para propor hipóteses sobre a estabilidade climática do nicho ao longo do tempo, nós realizamos uma modelagem de nicho utilizando 105 pontos de ocorrência (veja resultados: Figura 3a, Tabela 2) obtidos por nossos trabalhos de campo, espécimes de museu e revisões da literatura, assim como do sítio Global Biodiversity Information Facility – GIBF (disponível em [www.gbif.org](http://www.gbif.org)).

As 19 variáveis bioclimáticas (Nix, 1986; Busby, 1991), com a resolução de 2.5 minutos, foram obtidas da base de dados do WorldClim (Hijmans et al., 2005). Nós exploramos três conjuntos de variáveis climáticas para realizar a modelagem de nicho para o período atual: utilizando todas as variáveis; um conjunto obtido após a exclusão das variáveis mais correlacionadas (correlação >0.9); um conjunto obtido com as variáveis mais importantes para o modelo (com base no percentual de contribuição de cada variável maior que 5%). As análises finais foram realizadas com todas as 19 variáveis bioclimáticas, utilizando o algoritmo MAXENT (Phillips et al., 2010) que implementa uma abordagem *machine learning*, onde o peso de cada variável está embasado na importância (Seeholzer & Brumfield, 2018) e desta maneira reduzir o número de variáveis não é obrigatório (Elith et al., 2011; Braunisch et al., 2013), assim como porque apenas quando usamos todas as variáveis o modelo foi capaz de recuperar com alta adequabilidade as áreas de distribuição isoladas da espécie modelo (ex. em Macapá, Brasil) com altos valores de AUC. Nos projetamos a distribuição potencial para quatro janelas temporais: Ultimo Interglacial (LIG, ~150 a 120 mil anos atrás), Ultimo Máximo Glacial (LGM, 21 mil anos atrás) e Holoceno médio (6 mil anos atrás). Para o LGM e Holoceno médio dois modelos de circulação global (CCSM4 e MIROC) foram utilizados e ambas as projeções foram combinadas utilizando a ferramenta *raster calculator* no ArcMap 10.1. Nós configuramos 10 mil como número máximo para pontos de *background*, *Crossvalidate* foi utilizado como tipo de replicas da corrida, utilizamos 500 iterações e *10-percentile training presence* como critério de limiar. Todos os outros parâmetros foram mantidos como padrão do programa.



**Figura 3.** Modelagem de nicho potencial com base nas variáveis climáticas do WorldClim aplicando o critério de limiar *10 percentile training presence*. **A.** projeção atual utilizando todas as variáveis; os pontos vermelhos indicam os pontos de ocorrência utilizados nesta modelagem (para mais informações, veja a Tabela 2); **B.** Projeção para o Holoceno médio (6 mil anos atrás); cinza claro representa áreas adequadas apenas para um modelo de circulação global e cinza escuro representa as áreas adequadas concordantes; **C.** Projeção para o Último Máximo Glacial (21 mil anos atrás); cinza claro representa áreas adequadas apenas para um modelo de circulação global e cinza escuro representa as áreas adequadas concordantes; **D.** Projeção para o Último Interglacial (~ 150 a 120 mil anos atrás); **E.** Áreas estáveis definidas como as áreas adequadas para todos os períodos indicados pelas áreas em amarelo. Áreas adequadas nas demais combinações de tempo são mostradas em tons de cinza. As localidades genéticas utilizadas para testar a hipótese de maior variabilidade genética são indicadas pelas estrelas coloridas (localidades estáveis) e círculos coloridos (localidades não estáveis). As populações são representadas pelas cores: roxo: MW\_Caatinga; azul W\_Cerrado; rosa: W\_Caatinga; amarelo: Chaco e verde: E\_Cerrado.

**Tabela 2.** Pontos de ocorrência utilizados na modelagem de nicho potencial. O mapa mostrando a distribuição destes pontos está disponível na figura 3a.

<b>Localidade</b>	<b>Estado</b>	<b>País</b>	<b>Lat</b>	<b>Long</b>	<b>Referência</b>
Buenos Aires	Buenos Aires	Argentina	-36,000	-60,000	Bolívar e Silveira, 2015
Reserva El Destino	Buenos Aires	Argentina	-35.133	-57.416	Cueto e Cosenave, 2000
Catamarca, Alrededores De La Ciudad	Catamarca	Argentina	-28.467	-65.783	Bolívar e Silveira, 2015
47 km Ne Castelli (C57), Guemes	Chaco	Argentina	-25.950	-60.620	Bolívar e Silveira, 2015
Belleville	Córdoba	Argentina	-32.617	-62.700	Bolívar e Silveira, 2015
Unquillo	Córdoba	Argentina	-31.233	-64.317	Bolívar e Silveira, 2015
Cuay Grande	Corrientes	Argentina	-28.667	-56.283	Bolívar e Silveira, 2015
Las Lomas, Capital	Corrientes	Argentina	-28.480	-58.980	Bolívar e Silveira, 2015
Paraná	Entre Ríos	Argentina	-31.733	-60.533	Bolívar e Silveira, 2015
Río Uruguay	Entre Ríos	Argentina	-34.200	-58.300	Bolívar e Silveira, 2015
4 km Ne De Laguna Yema, Bermejo	Formosa	Argentina	-24.25	-61.233	Bolívar e Silveira, 2015
Paraje Ñandhy Vera, Laishi	Formosa	Argentina	-26.233	-58.617	Bolívar e Silveira, 2015
Guerrero	Jujuy	Argentina	-24.217	-57.850	Bolívar e Silveira, 2015
Conhella	La Pampa	Argentina	-36.017	-64.600	Bolívar e Silveira, 2015
La Pampa, Loventuel	La Pampa	Argentina	-36.183	-65.25	Bolívar e Silveira, 2015
La Rioja, Sañogasta	La Rioja	Argentina	-29.300	-67.600	Bolívar e Silveira, 2015
Reserva de Telteca	Mendoza	Argentina	-32.350	-68.050	Blendinger, 2008
Nacuman, Santa Rosa	Mendoza	Argentina	-34.05	-67.967	Bolívar e Silveira, 2015
Aguaray	Salta	Argentina	-22.267	-63.733	Bolívar e Silveira, 2015
El Bananal, Urundel	Salta	Argentina	-23.55	-64.417	Bolívar e Silveira, 2015
Pozo Negro, General Belgrano	San Luis	Argentina	-32.65	-66.467	Bolívar e Silveira, 2015
Gato Colorado, El Tostado	Santa Fé	Argentina	-28.033	-61.500	Bolívar e Silveira, 2015
Villa Constitución, Islar Río Parana	Santa Fé	Argentina	-33.233	-60.333	Bolívar e Silveira, 2015
Suncho Corral	Santiago del Estero	Argentina	-27.933	-63.450	Bolívar e Silveira, 2015

Concepción	Tucumán	Argentina	-27.333	-63.583	Bolívar e Silveira, 2015
Tapia	Tucumán	Argentina	-26.600	-65.300	Bolívar e Silveira, 2015
Laguna Cayaca	Cordillera	Bolivia	-18.673	-63.917	UWBM
Yabaré	Chiquitos	Bolivia	-17.433	-61.167	Bolívar e Silveira, 2015
Localidade El Tuná, 300 m N do Rio Mercedes	Velasco	Bolivia	-16.550	-59.650	AMNH
Macapá, Campus Experimental da EMBRAPA	Amapá	Brasil	0.4500	-50.933	Bolívar e Silveira, 2015
Alto Salitre, Morro do chapéu	Bahia	Brasil	-11.408	-41.164	Silveira e Machado, 2012
Tartarugalzinho, Fazenda Casemiro	Bahia	Brasil	-13.830	-50.750	Bolívar e Silveira, 2015
Fazenda do Marcelo, Vicinal	Bahia	Brasil	-14.923	-40.727	Bolívar e Silveira, 2015
Palmas de Monte Alto, Fazenda Boa Vista	Bahia	Brasil	-14.283	-43.333	Bolívar e Silveira, 2015
Santa Rita de Cássia	Bahia	Brasil	-11.35	-43.867	Bolívar e Silveira, 2015
Pau-a-pique, Remanso, BA	Bahia	Brasil	-9.527	-41.646	Ross et al 2006
IBGE	Brasilia	Brasil	-15.883	-47.842	Este trabalho
Ipú, Serra do Ibiapaba	Ceara	Brasil	-4.333	-40.700	Bolívar e Silveira, 2015
Itapipoca	Ceara	Brasil	-3.500	-39.583	Bolívar e Silveira, 2015
Aurilândia	Goiás	Brasil	-16.679	-50.456	Este trabalho
Campinaçu	Goiás	Brasil	-13.915	-48.504	Bolívar e Silveira, 2015
Cana Brava	Goiás	Brasil	-13.850	-46.950	Bolívar e Silveira, 2015
Lagoa Formosa, Cabeceiras	Goiás	Brasil	-15.800	-46.980	Bolívar e Silveira, 2015
Grajaú, Fazenda Canto da Onça	Maranhão	Brasil	-6.217	-46.117	Bolívar e Silveira, 2015
Piritoró	Maranhão	Brasil	-4.383	-44.333	Fonoteca Neotropical
Arinos	Mato Grosso	Brasil	-14.45	-56.217	Bolívar e Silveira, 2015
Chavantina, Rio das Mortes	Mato Grosso	Brasil	-11.75	-50.733	Bolívar e Silveira, 2015
Dumbá	Mato Grosso	Brasil	-14.500	-51.000	Bolívar e Silveira, 2015
Nossa Senhora do Livramento	Mato Grosso	Brasil	-16.366	-56.299	MPEG
Tapirapoan	Mato Grosso	Brasil	-14.850	-57.750	Bolívar e Silveira, 2015
Campo Grande	Mato Grosso do Sul	Brasil	-20.45	-54.617	Bolívar e Silveira, 2015
Corumbá	Mato Grosso do Sul	Brasil	-19.017	-57.650	Bolívar e Silveira, 2015

Fazenda Beija-Flor, margem esquerda do Rio Sucuruí, Três Lagoas	Mato Grosso do Sul	Brasil	-21.130	-56.470	Bolívar e Silveira, 2015
margem direita do Rio Sucuruí, Três Lagoas	Mato Grosso do Sul	Brasil	-20.800	-51.717	Bolívar e Silveira, 2015 Coleção Zoológica de Referência da Seção de Vírus
Corumbá	Mato Grosso do Sul	Brasil	-18.975	-56.647	
Antonio Dias	Minas Gerais	Brasil	-16.500	-41.883	Cunha e Vasconcelos 2009
Baependi	Minas Gerais	Brasil	-21.950	-44.883	Bolívar e Silveira, 2015
Piraporã	Minas Gerais	Brasil	-17.350	-44.933	Bolívar e Silveira, 2015
Santa Barbara	Minas Gerais	Brasil	-19.983	-43.466	Cunha e Vasconcelos 2009
Uberlândia	Minas Gerais	Brasil	-18.876	-48.255	Valadão et al 2006
Reserva Brisa do Indaiá, Quartel General Oliveira, Sítio Jacaré	Minas Gerais	Brasil	-19.236	-45.803	Freitas et al 2009
Fazenda Fartura, Retiro 8	Pará	Brasil	-9.833	-50.367	Bolívar e Silveira, 2015
Monte Alegre, Serra do Ererê	Pará	Brasil	-2.017	-54.167	Bolívar e Silveira, 2015
Reserva Indígena Missão Tiriós	Pará	Brasil	-2.233	-55.95	Bolívar e Silveira, 2015
Riosinho, Posto Nilo-Peçanha	Pará	Brasil	-6.650	-51.983	Bolívar e Silveira, 2015
Coremas	Paraíba	Brasil	-7.017	-37.967	Bolívar e Silveira, 2015
Fazenda Campos Bons 38 km N. Floresta Arcoverde	Pernambuco	Brasil	-82.900	-38.580	Bolívar e Silveira, 2015
Bocaina, Comunidade Salseiro	Pernambuco	Brasil	-8.418	-37.053	Fonoteca Neotropical
Mato dos Jegues	Piauí	Brasil	-6.950	-41.283	Bolívar e Silveira, 2015
Parque Nacional da Serra das Confusões	Piauí	Brasil	-6.765	-43.820	Olmos e Brito 2007
Ribeiro Gonçalves	Piauí	Brasil	-9.226	-43.463	Bolívar e Silveira, 2015
EMBRAPA_Seropédica	Piauí	Brasil	-75.300	-45.243	Bolívar e Silveira, 2015
Franca	Rio de Janeiro	Brasil	-22.763	-43.688	Luz et al 2007
Santa Madalena	São Paulo	Brasil	-20.533	-47.400	Bolívar e Silveira, 2015
Parque Ibirapuera	São Paulo	Brasil	-22.667	-53.150	Bolívar e Silveira, 2015
	São Paulo	Brasil	-23.587	-46.658	Fonoteca Neotropical

Ribeirão Bonito	São Paulo	Brasil	-21.975	-48.279	Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodes - Banco de Aves
Itapetininga	São Paulo	Brasil	-23.666	-48.000	Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodes - Banco de Aves
Fazenda Roma, Santa Rosa de Tocantins	Tocantins	Brasil	-11.267	-48.450	Bolívar e Silveira, 2015
Mata do Rio Galhão	Tocantins	Brasil	-10.527	-46.106	Bolívar e Silveira, 2015
Puerto Guaraní	Alto Paraguay	Paraguai	-21.300	-57.917	Bolívar e Silveira, 2015
San Rafael	Itapúa	Paraguai	-26.667	-54.883	Bolívar e Silveira, 2015
Escobar	Paraguari	Paraguai	-25.650	-57.017	Bolívar e Silveira, 2015
Rio Negro	Presidente Hayes	Paraguai	-23.400	-57.333	Bolívar e Silveira, 2015
San Gregorio, Artigas	Artigas	Uruguai	-30.550	-57.867	Bolívar e Silveira, 2015
Rio Negro, R. O., Del Uruguay	Rio Negro	Uruguai	-32.750	-57.333	Bolívar e Silveira, 2015
Arari, Ilha do Marajó	Pará	Brasil	-0.90	-49.040	University of Michigan Museum of Zoology - GBIF ID:767307271
Guarai	Tocantins	Brasil	-8.820	-48.490	e-bird Observation dataset - GBIF ID: 944073679
Estrada para canudos	Bahia	Brasil	-10.060	-38.480	e-bird Observation dataset - GBIF ID: 1185937130
Area aripuana	Mato Grosso	Brasil	-9.160	-60.730	e-bird Observation dataset - GBIF ID: 829031155
Triangulo		Bolivia	-11.450	-66.300	e-bird Observation dataset - GBIF ID: 832320030
Flor de Oro		Bolivia	-13.540	-61.010	e-bird Observation dataset - GBIF ID: 197137708
Estancia Motacuzal	El Beni	Bolivia	-13.750	-64.630	e-bird Observation dataset - GBIF ID: 587915596



Barba Azul Nature Reserve		Bolivia	-13.760	-66.100	e-bird Observation dataset - GBIF ID: 117750213
Puerto Moscoso		Bolivia	-13.630	-68.730	e-bird Observation dataset - GBIF ID: 197136613
Rio Torotoro		Bolivia	-18.130	-65.760	e-bird Observation dataset - GBIF ID: 961183164
Parque Nacional das Emas	Goiás	Brasil	-18.260	-52.890	e-bird Observation dataset - GBIF ID: 1196944745
Enceso		Paraguai	-21.210	-61.640	e-bird Observation dataset - GBIF ID: 962857624
San Francisco, Calilegua, Valle Grande	Jujuy	Argentina	-22.030	-66.150	Fundacion Miguel Lillo Provider - GBIF ID:730444244
Ocloyas	Jujuy	Argentina	-23.940	-65.220	Macaulay Library Audio and Video Collection - GBIF ID:1147042609
Restinga de Massambaba	Rio de Janeiro	Brasil	-22.930	-42.300	e-bird Observation dataset - GBIF ID: 977183623
Bosque Alemão	Paraná	Brasil	-25.410	-49.290	e-bird Observation dataset - GBIF ID: 588202889
Tacuaremba		Uruguai	-31.740	-56.010	e-bird Observation dataset - GBIF ID: 976556876
Reserva Natural Chancani	Córdoba	Argentina	-31.380	-65.480	e-bird Observation dataset - GBIF ID: 1163523988
Reserva Ecológica Mar Del Plata	Buenos Aires	Argentina	-38.060	-57.540	e-bird Observation dataset - GBIF ID: 829291214

### *Hipótese das Áreas estáveis*

As áreas climaticamente estáveis foram definidas sobrepondo a distribuição potencial atual e as outras três projeções de paleo distribuição (LIG, LGM e Holoceno médio) usando a ferramenta *raster calculator* no ArcMap 10.1. Áreas que foram ecologicamente estáveis em todas as projeções foram definidas como as áreas estáveis para *L. angustirostris*. Nós categorizamos todas as amostras em grupos estáveis e não estáveis com base na localização geográfica das amostras genéticas sob o mapa de áreas estáveis (Figura 3e). Para avaliar se as áreas estáveis abrigam uma maior diversidade genética quando comparada as áreas não estáveis, nós comparamos as médias dos valores de diversidade nucleotídica, diversidade haplotípica e diferença nucleotídica para as áreas estáveis e não estáveis por meio de um teste de Kruskal Wallis utilizando o programa Past3 (Hammer et al., 2001). Nós utilizamos as populações como réplicas para este teste.

### *Isolamento por distância*

Para testar o isolamento por distância como um processo que guiou a diversificação de *L. angustirostris*, nós analisamos a estrutura genética e realizamos análises de correlação entre matrizes de distância genética ( $\Phi_{st}$ ) e geográfica (Km) por meio do teste de Mantel com 10 mil iterações implementado no programa Arlequin para todos os indivíduos. Como nós encontramos barreiras genéticas ao fluxo gênico (veja resultados), nós também realizamos esta análise para cada população. Nós realizamos uma análise de auto correlação espacial para identificar as classes de distância onde as correlações do teste de Mantel podem ser consideradas estatisticamente significativas (Sokal et al., 1997). Nós realizamos esta análise utilizando o programa *Alleles in Space* (AIS; (Miller, 2005)) com base na média da distância genética entre os pares de indivíduos pertencentes a uma mesma classe  $y$ , considerando 10 classes e mil permutações. Para evitar o viés proveniente da baixa variabilidade do gene FIB5, nós realizamos esta análise apenas com as sequencias de ND2.

## **RESULTADOS**

### *Estrutura genética da população*

A análise de variância molecular sugere que a espécie deva ser dividida em cinco populações, de acordo com o menor valor de  $\Phi_{sc}$  para ambos os marcadores analisados (Tabela 3). Nós identificamos duas populações na Caatinga, duas no Cerrado e uma no Chaco. As populações da Caatinga ocorrem na região centro-oeste (MW\_Caatinga) e mais oeste

(W\_Caatinga) do bioma, as populações do Cerrado nas regiões mais a leste (E\_Cerrado) e mais a oeste (W\_Cerrado) do bioma (Figura 1b). De acordo com a divisão espacial, a estrutura genética de *L. angustirostris* é forte (ND2  $\Phi_{st}$  = 0.76; FIB5  $\Phi_{st}$  = 0.62; Tabela 3).

**Tabela 3.** Identificação do melhor número de populações utilizando os índices de fixação da análise de variância molecular (AMOVA). O melhor número de populações (números destacados em negrito) foi escolhido pelo menor valor positivo de  $\Phi_{SC}$  (para mais detalhes veja a seção Populações genéticas e estrutura genética)

Gene	Grupos	$\Phi_{SC}$	$\Phi_{ST}$	$\Phi_{CT}$
ND2	2	0.670	0.780	0.320
	3	0.530	0.800	0.580
	4	0.450	0.790	0.620
	5	<b>0.320</b>	0.760	0.650
FIB5	2	0.620	0.610	-0.020
	3	0.610	0.620	0.010
	4	0.600	0.620	0.040
	5	<b>0.580</b>	0.620	0.090

Como esperado, a variabilidade genética do ND2 foi maior quando comparado ao FIB5 para todos os parâmetros analisados (Tabela 4). MW\_Caatinga apresentou maior valor de diversidade haplotípica para ambos os marcadores ( $H_d$  = 1.00 para ND2 e  $H_d$  = 0.79 para FIB5), enquanto o menor valor de diversidade haplotípica foi observado na população W\_Caatinga para o gene ND2 ( $H_d$  = 0.40) e na população W\_Cerrado para o gene FIB5 ( $H_d$  = 0.07). O maior valor de  $\Phi_{st}$  par a par foi encontrado entre MW\_Caatinga e Chaco, enquanto que o menor valor foi encontrado quando comparadas as populações E\_Cerrado e W\_Cerrado, e E\_Cerrado e Chaco (Tabela 5).

**Tabela 4.** Caracterização da variabilidade genética para todos os indivíduos e para cada população. Números entre parênteses indicam o número de indivíduos analisados. As medidas de variabilidade são Nh: número de haplótipos; Hd: diversidade haplotípica; Nd: diversidade nucleotídica; k: diferenças nucleotídicas.

	<b>All localities (63)</b>		<b>MW_Caatinga (6)</b>		<b>W_Cerrado (14)</b>		<b>W_Caatinga (5)</b>		<b>Chaco (24)</b>		<b>E_Cerrado (14)</b>	
	ND2	FIB5	ND2	FIB5	ND2	FIB5	ND2	FIB5	ND2	FIB5	ND2	FIB5
<b>Nh</b>	32	11	6	4	9	2	2	2	15	5	8	4
<b>Hd</b>	0.939	0.402	1.000	0.788	0.901	0.071	0.400	0.533	0.866	0.235	0.912	0.550
<b>Nd</b>	0.006	0.001	0.003	0.003	0.003	0.000	0.001	0.001	0.004	0.006	0.002	0.002
<b>k</b>	4.708	0.610	3.600	1.696	3.098	0.071	0.800	0.533	1.902	0.246	2.868	1.047

**Tabela 5.** Valores de  $\Phi_{ST}$  para a par entre as populações de *L. angustirostris*. Os valores da diagonal inferior representam a análise do gene mitocondrial (ND2) com base em uma matriz de distância K2P. Os valores da diagonal superior representam a análise do gene nuclear (FIB5) com base em uma matriz de distância p. Valores estatisticamente significativos são ressaltados em negrito.

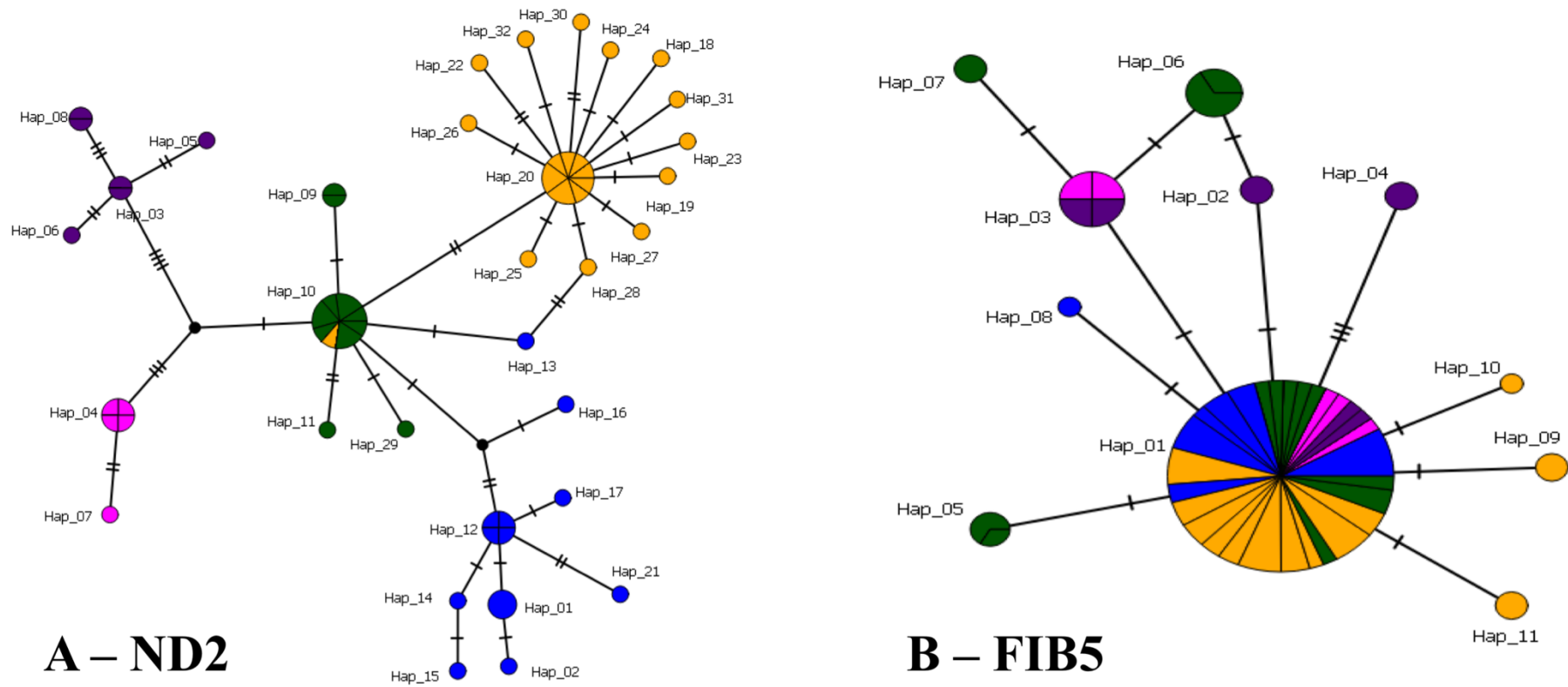
	<b>E_Caatinga</b>	<b>W_Cerrado</b>	<b>W_Caatinga</b>	<b>Chaco</b>	<b>E_Cerrado</b>
<b>E_Caatinga</b>	-	0.19673	<b>0.01183</b>	0.2133	<b>0.02176</b>
<b>W_Cerrado</b>	0.73727	-	0.5764	<b>0.0025</b>	0.17544
<b>W_Caatinga</b>	0.70665	0.69921	-	0.447	<b>0.02639</b>
<b>Chaco</b>	0.80343	0.64455	0.76276	-	0.18695
<b>E_Cerrado</b>	0.70862	0.48455	0.64574	0.4895	-

### *Análises filogeográficas*

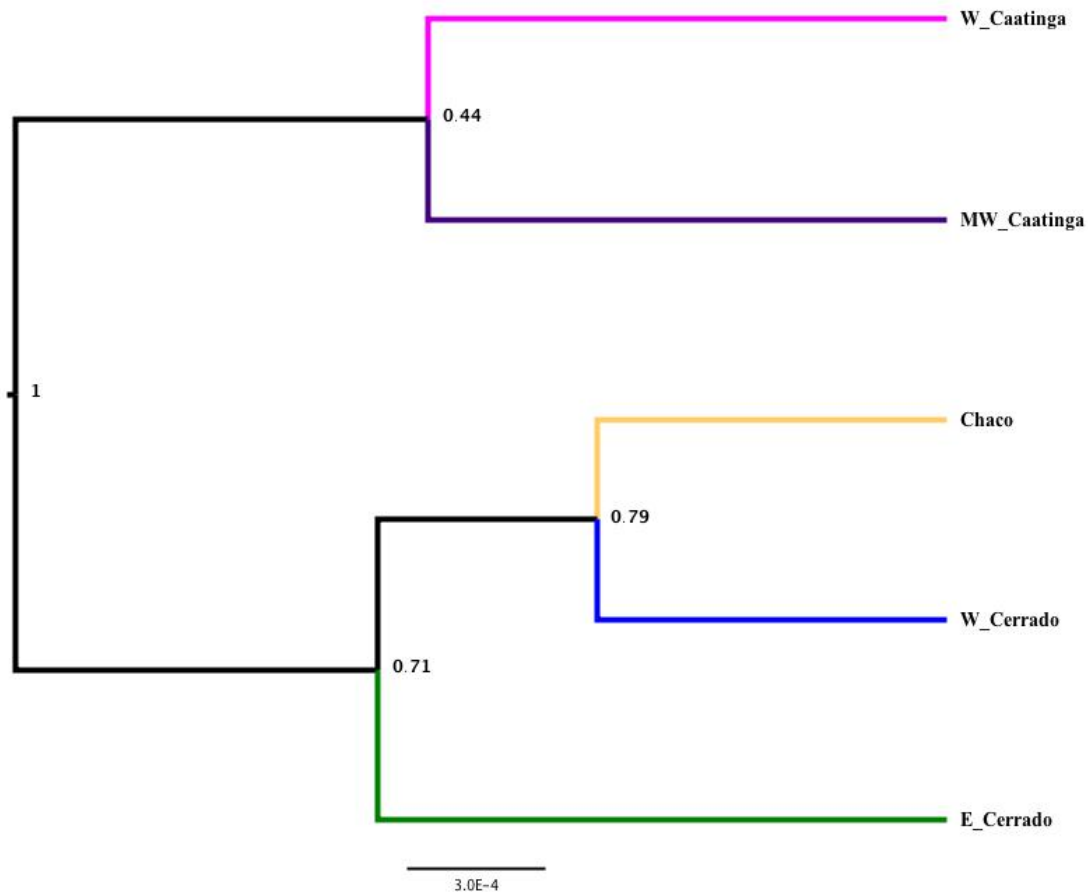
A rede de haplótipos do ND2 (Figura 4a) apresenta cinco haplogrupos concordantes com as análises de estrutura genética. É claro um padrão de estrela na população do Chaco sugerindo uma expansão recente. Por outro lado, a rede de haplótipo do FIB5 apresenta um haplótipo muito frequente e amplamente distribuído (H\_01, Figura 4b) e poucos haplótipos derivados menos frequentes.

A análise de inferência bayesiana (Figura 2) apresenta todas as populações como clados monofiléticos bem resolvidos, com exceção da população E\_Cerrado. Esta última possui um haplótipo proveniente da população W\_Cerrado (LOC20\_3002) e um proveniente da população do Chaco (LOC27\_4307). Todos os clados monofiléticos apresentaram altos valores de probabilidade posterior ( $pp = 1$ ), com exceção da população W\_Cerrado ( $pp = 0.8$ ). Quanto a relação entre as populações, a única relação bem suportada ( $pp = 1$ ) é a população MW\_Caatinga como a primeira população a divergir. A análise de *species tree* (Figura 5) mostram as populações da Caatinga formando um clado diferente do clado formado por Cerrado + Chaco com altos valores de probabilidade posterior ( $pp = 1$ ), enquanto que a relação dentro do clado Cerrado + Chaco continua incerta devido aos baixos valores de probabilidade posterior.

Nós não encontramos linhagens únicas para as SDTFs, nem encontramos que as linhagens pertencentes aos núcleos das STDFs sejam linhagens com relação filogenética próxima (Figura 2). Ao invés disso, o núcleo da Caatinga apresenta duas linhagens e possui algumas localidades inclusas na linhagem E\_Cerrado; o núcleo de Misiones e Piedmont pertencem a linhagem do Chaco; enquanto que o núcleo Chiquitano pertence a linhagem W\_Cerrado.



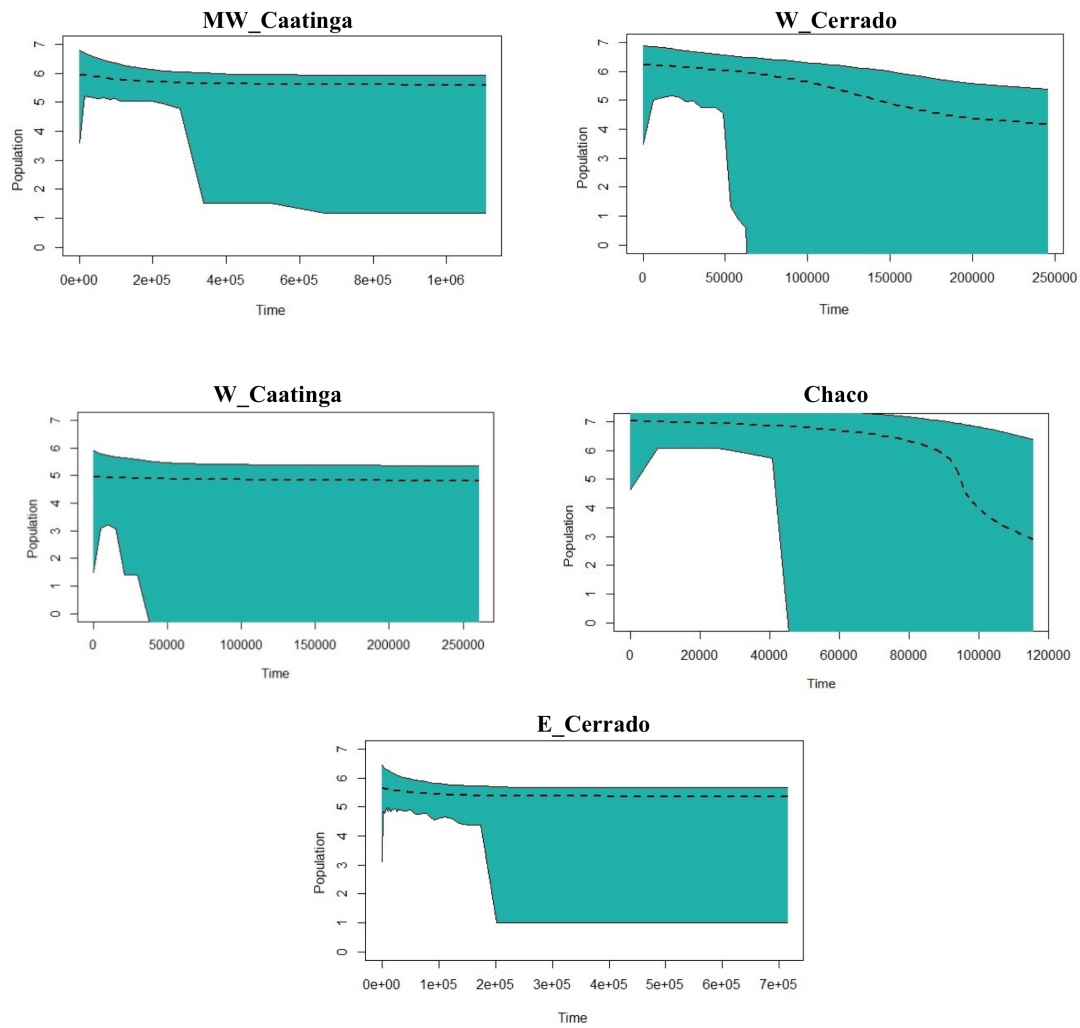
**Figura 4.** Rede de haplótipos mostrando a frequência e distribuição dos haplótipos. As cores correspondem as populações definidas pela análise de estrutura genética: roxo: MW\_Caatinga; azul: W\_Cerrado; rosa: W\_Caatinga; amarelo: Chaco e verde: E\_Cerrado. **A.** Rede de haplótipos do gene mitocondrial (ND2); **B.** Rede de haplótipos do gene nuclear (FIB5).



**Figura5.** Análise de *Species tree* utilizando modelo de *Yule process* e modelo de Tamanho populacional constante de *piecewise* para os genes ND2 e FIB5. Os valores de suporte são descritos para cada nó. As cores correspondem as populações definidas pela análise de estrutura genética: roxo: MW\_Caatinga; azul: W\_Cerrado; rosa: W\_Caatinga; amarelo: Chaco e verde: E\_Cerrado.

#### *Processos demográficos*

A história demográfica recuperada pelos testes  $R_2$ ,  $F_u$ 's  $F_s$ , e Tajima's  $D$  mostram um claro sinal de expansão recente apenas na população do Chaco, enquanto a população W\_Caatinga mostra sinais de expansão apenas no teste  $R_2$  para o marcador ND2 ( $R_2 = 0.0847$ ,  $p = 0.0045$ ) (Tabela 6). Um padrão similar foi observado na análise de EBSP onde a população do Chaco apresentou sinais de mudanças de tamanho populacional ao longo do tempo (Figura 6).



**Figura 6.** Extended Bayesian skyline plot mostrando a tendência do tamanho efetivo populacional ao longo do tempo. A linha tracejada indica a tendência do tamanho populacional ao longo do tempo nas diferentes populações de *L. angustirostris* identificadas neste estudo. O desvio padrão é representado pelos blocos verdes.

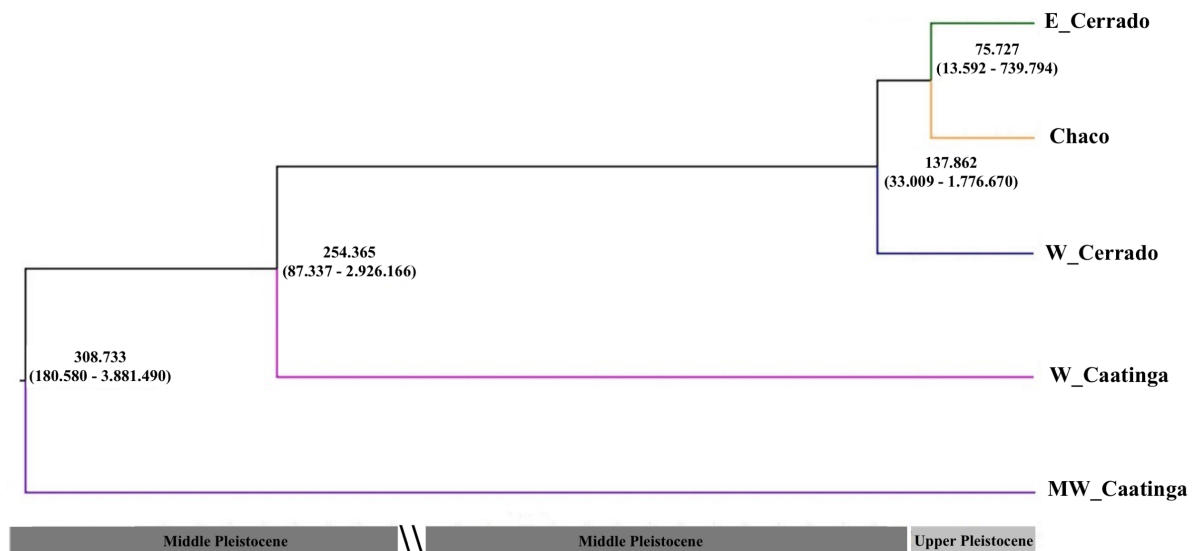


**Tabela 6.** Estimativas demográficas de expansão populacional e correlação da distancia genética e geográfica dentro de cada população. Valores estatisticamente significativos são destacados em negrito.

Gene	Population	Fu's FS	p	Tajima'D	p	R <sup>2</sup>	p	Mantel (r value)	p
<b>ND2</b>	MW_Caatinga	0.121	0.473	-11.007	0.167	0.177	0.323	NAN	1.000
	W_Cerrado	-30.189	0.037	-13.987	0.081	<b>0.084</b>	<b>0.004</b>	0.820	0.107
	W_Caatinga	10.986	0.738	-0.709	0.251	0.433	100.000	-0.538	1.000
	Chaco	<b>-12.551</b>	<b>0.000</b>	-24.242	0.000	<b>0.041</b>	<b>0.000</b>	0.088	0.431
	E_Cerrado	-20.800	0.097	-0.673	0.281	0.112	0.108	<b>0.323</b>	<b>0.009</b>
<b>FIB5</b>	MW_Caatinga	0.480	0.625	-0.171	0.476	0.149	0.364	<b>0.440</b>	<b>0.005</b>
	W_Cerrado	-11.545	0.338	-11.514	0.114	0.185	0.972	0.290	0.068
	W_Caatinga	0.966	0.850	14.442	0.920	0.285	0.696	0.152	0.338
	Chaco	<b>-41.630</b>	<b>0.000</b>	<b>-16.634</b>	<b>0.001</b>	0.054	0.083	0.102	0.316
	E_Cerrado	0.452	0.631	0.050	0.571	0.131	0.136	-0.1314	0.7390

### Tempo de divergência

Todos os eventos de diversificação intraespecífica de *L. angustirostris* ocorreram durante o Pleistoceno (Figura 5, Tabela 7). A primeira, separando a população MW\_Caatinga das demais ocorreu por volta de 308 mil anos atrás (variando de 180 a 3.880 mil anos atrás), seguida pela divergência da população W\_Caatinga por volta de 254 mil anos (variando de 87 a 2.920 mil anos atrás) e W\_Cerrado por volta de 137 mil anos atrás (variando de 33 a 1.77<sup>o</sup> mil anos atrás), todas elas no Pleistoceno médio. Por fim, a diversificação mais recente entre Chaco e E\_Cerrado é datada por volta de 75 mil anos atrás (variando de 13 a 740 mil anos atrás) durante o Pleistoceno superior.



**Figura 7.** Estimativa do tempo de divergência das populações. O desvio padrão é indicado em parêntesis abaixo das estimativas de tempo. Os valores foram estimados pelo modelo de Isolamento com migração utilizando os marcadores ND2 e FIB5. As cores correspondem as populações definidas pela análise de estrutura genética: roxo: MW\_Caatinga; azul: W\_Cerrado; rosa: W\_Caatinga; amarelo: Chaco e verde: E\_Cerrado.

**Tabela 7.** Tamanho efetivo populacional (Ne) e número de migrantes (M) considerando uma única taxa de migração para todos os pares de população. Números entre parêntesis indicam o limite superior e inferior dos valores de *95% highest posterior density*.

<b>Population</b>	<b>Ne</b>	<b>M</b>
MW_Caatinga	263,750 (134,302 – 493,520)	23,737 (12 – 44,416)
W_Cerrado	127,830 (30,744 – 422,323)	11,504 (2,766 – 38,009)
W_Caatinga	208,734 (59,870 – 668,274)	18,786 (5,388 – 60,144)
Chaco	66,342 (8,090 – 409,378)	5.970 (728 – 36,844)
E_Cerrado	315,529 (95,468 – 1,159,853)	28,397 (8,592 – 104,386)

### *Modelagem de nicho*

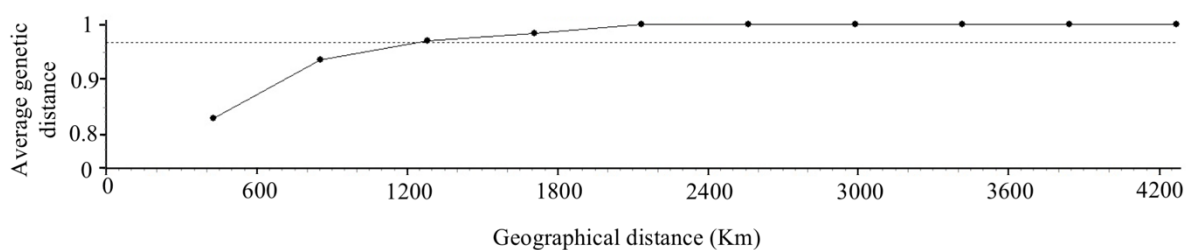
O melhor modelo para a distribuição potencial atual foi o modelo incluindo todas as variáveis bioclimáticas (AUC = 0.879). Este modelo também foi o modelo mais próximo a atual distribuição descrita para esta espécie (Figura 3a). A projeção para p LIG (~150 a 120 mil anos atrás; Figura 3d) sugere um período com as menores e mais fragmentadas áreas. No LIG, o Chaco foi a única mancha de área adequada próxima ou encontrado atualmente. No LGM (21 mil anos atrás; Figura 3c), foi possível identificar a expansão das áreas adequadas, e sinal de conexão dessas áreas. A expansão é mais evidente para mancha coincidente com o Cerrado. No Holoceno médio (6 mil anos atrás; 3b), as áreas adequadas têm um tamanho levemente maior do que o observado para o período atual.

### *Hipótese de áreas estáveis*

A maior área ecologicamente estável ao longo do tempo identificada para *L. angustirostris* foi dentro do que atualmente é a região que compreende o Chaco e os núcelos das SDTFs Misiones e Piedmont (Figura 3e). Nós também identificamos três manchas de áreas estáveis principais na parte mais oeste, no centro e na parte mais a leste do Cerrado e duas manchas na parte mais oeste e mais a leste da Caatinga (Figura 3e). A comparação dos índices de diversidade genética entre áreas estáveis e não estáveis não demonstram diferenças estatisticamente significativas (Tabela 8, Tabela 9). Porém, nós identificamos uma correspondência geográfica entre as áreas ecologicamente estáveis e as populações genéticas identificadas para *L. angustirostris*, exceto para a população MW\_Caatinga (Figura 3e).

### *Hipótese de Isolamento por distância*

Nós encontramos correlação entre distância genética e geográfica considerando todos os indivíduos ( $r = 0.22$   $p = 0.000$ ), enquanto nenhuma evidencia de correlação foi observada dentro das populações com exceção da população E\_Cerrado ( $r = 0.32$   $p = 0.009$ ) (Tabela 6). No entanto, a análise de auto correlação para ND2 mostra que o isolamento por distância apenas ocorre em distâncias menores que 1.400 km ( $p = 0.003$ ; Figura 8). Então, considerando todos os indivíduos e que algumas localidades amostradas na população E\_Cerrado estão distantes entre si mais de 1.400km uma das outras, os resultados significativos de Mantel para todos os indivíduos e dentro da população E\_Cerrado podem ser espúrias.



**Figure 8.** Análise de auto correlação especial com base na distância genética média do gene ND2 entre pares de indivíduos dentro de uma mesma classe de distância y com 1000 permutações. A linha pontilhada corresponde a hipótese nula de nenhuma diferença de correlação devido ao acaso. Pontos abaixo da linha pontilhada representam as classes com auto correlação especial enquanto os pontos acima da linha pontilhada indicam classes onde não é observado auto correlação .

**Tabela 8.** Resultado do teste de maior variabilidade genética nos grupos estáveis quando comparados aos grupos não estáveis. Números entre parêntesis indicam o número de indivíduos analisados. As medidas de variabilidade são Nh: número de haplótipos; Hd: diversidade haplotípica; Nd: diversidade nucleotídica; k: diferenças nucleotídicas. A média de cada medida e o desvio padrão são apresentados por gene

Gene	Diversidade	Área estável (34)		Área instável(29)		Kruskal-Wallis	
		média	sd	média	sd	$\chi^2$	p
ND2	Hd	0,93740	0,086248	0,74020	0,421462	0,5345	0,4354
	Nd	0,00219	0,000637	0,00285	0,001964	0,8836	0,3472
	k	212460	0,632582	274180	1915568	0,5345	0,4647
FIB5	Hd	0,48300	0,359458	0,302600	0,247669	1091	0,2933
	Nd	0,00217	0,002341	0,000642	0,000526	1091	0,2933
	k	102470	1101454	0,302630	0,247818	1091	0,2933

**Tabela 9.** Índices genéticos de variabilidade nos grupos de áreas climaticamente estáveis (S) e áreas climaticamente não-estáveis (NS) por população. As medidas de variabilidade são: N – número de indivíduos; *Hd* – diversidade haplotípica; *Nd* – diversidade nucleotídica; *k* – diferenças nucleotídicas. As medias e desvio padrão de cada medida são apresentadas por gene por população.

Gene	População	Diversity Index											
		N		Hd				Nd				k	
		S	NS	S	NS	S	NS	S	NS	S	NS		
				média	sd	média	sd	média	sd	média	sd	média	média
ND2	MW_Caatinga	3	3	1.000	0.272	1.000	0.272	0.002	0.001	0.005	0.001	266	533
	W_Cerrado	3	11	1.000	0.272	0.891	0.074	0.001	0.000	0.003	0.000	133	330
	W_Caatinga	2	3	1.000	0.500	0.000	0.000	0.002	0.001	0.000	0.000	200	0.00
	Chaco	19	24	0.830	0.085	1.000	0.126	0.001	0.000	0.002	0.000	176	240
	E_Cerrado	7	7	0.857	0.137	0.810	0.130	0.002	0.000	0.002	0.000	285	266
FIB5	MW_Caatinga	3	3	0.800	0.122	0.533	0.172	0.005	0.001	0.001	0.000	266	0.533
	W_Cerrado	3	11	0.000	0.000	0.091	0.081	0.000	0.000	0.000	0.000	0.00	0.090
	W_Caatinga	2	3	0.667	0.204	0.533	0.172	0.001	0.000	0.001	0.000	0.66	0.533
	Chaco	19	24	0.201	0.085	0.356	0.159	0.000	0.000	0.000	0.000	0.20	0.355
	E_Cerrado	7	7	0.747	0.078	0.000	0.000	0.003	0.000	0.000	0.000	158	0.000

## DISCUSSÃO

Nossos resultados indicam que *L. angustirostris* apresenta uma forte estrutura genética populacional, provavelmente influenciado por eventos que ocorreram durante o Pleistoceno. As amostras provenientes dos diferentes núcleos SDTFs não formam grupos monofiléticos, contrário do que é esperado pela hipótese do arco Pleistocênico. Ao invés disso, a presença de linhagens parafiléticas associadas as florestas secas podem ser melhor explicada pelo produto de dispersão a curta distância para os biomas vizinhos. A diversidade genética não foi significativamente maior em regiões climaticamente estáveis quando comparadas às regiões não estáveis, sugerindo que as flutuações climáticas do Pleistoceno não representam uma força evolutiva primária para esta espécie. Similarmente, o processo de isolamento por distância não foi um fator importante para explicar a estrutura populacional desta espécie. Abaixo nós discutiremos estas afirmações com base nos nossos principais resultados.

### *Hipótese do Arco Pleistocênico*

A Hipótese do Arco Pleistocênico não foi suportada pelos nossos resultados porque as amostras dos núcleos das SDTFs não mostram uma estrutura filogeográfica coincidente com linhagens únicas das SDTFs (Figura 2). Além disso, as modelagens de paleo distribuição sugere que habitat potencial das populações de *L. angustirostris* ocupando as SDTFs foi muito mais disjunto durante o LIG do que nos dias atuais (Figura 3d), e mostram uma expansão progressiva no sentido norte e oeste para uma escala continental desde o LGM (Figura 3c). Este sentido de expansão foi contrário ao observado Werneck e colaboradores (2011) que identificaram um sentido de expansão ao sul e ao leste das áreas dos núcleos das SDTFs desde o LGM.

Nossos resultados não suportam a ideia de dispersão a longa distância proposta por Mayle (2004) para explicar as similaridades biogeográfica entre os núcleos das SDTFs. Ao invés disso, nossos resultados indicam que as regiões dos núcleos de Chiquitano, Piedmont e Misiones foram recentemente colonizadas por indivíduos que vieram de biomas vizinhos (Chaco e W\_Cerrado), possivelmente favorecida pela conectividade estabelecida entre esses diferentes biomas. Por exemplo, a localidade 25 amostrada no núcleo de Chiquitano foi incluído na população W\_Cerrado, enquanto as localidades 29-32 amostradas no núcleo de Misiones e as localidades 24 e 26 amostras nos núcleo Piedmont foi incluída na população do Chaco (Figura 2). Nossos modelos de paleo distribuição concordam com a ligação genética entre as populações do Arapaçu do Chaco e do núcleo de Misiones porque sugerem que essas regiões estão conectadas desde 120 mil anos atrás. Além disso, a população do Chaco mostra

sinal consistente de expansão da distribuição (com base nos testes de  $R_2$ ,  $F_u$ 's  $F_s$ , Tajima's  $D$ , rede de haplótipos e EBSP), enquanto  $W_{\text{Cerrado}}$  mostram sinais de expansão da distribuição nas análises de EBSP. Outra evidência que suporta a conectividade entre a região de Chiquitano e do Cerrado são os registros de dados de pólen indicando que o estabelecimento final da vegetação do Cerrado só aconteceu no Holoceno médio (Ledru, 2002), e a potencial expansão das regiões de SDTFs durante esse período são indicadas por paleo modelagem (Werneck et al., 2011).

### *Hipótese de Áreas estáveis*

Nossos resultados indicam que as populações de *L. angustirostris* de áreas estáveis não apresentem diversidade genética maior quando comparadas as áreas instáveis. Este resultado discorda de estudos prévios com espécies florestais, ex: sapos da mata atlântica, que mostram maior diversidade genética em áreas estáveis (Carnaval et al., 2009). Por outro lado, os resultados de *L. angustirostris* concordam com outros estudos de espécies de biomas abertos (Werneck et al., 2012b; Santos et al., 2014; Lima, 2016). Por exemplo, uma ave endêmica do Cerrado *Neothraupis fasciata* também mostra variabilidade genética similar entre áreas estáveis e não estáveis (Lima, 2016). Além disso, uma espécie de lagarto que ocorre nas formações de vegetação aberta da América do Sul mostra maior diversidade genética em áreas instáveis quando comparadas as áreas estáveis (Werneck et al., 2012a). Com isso, espécies de formações de vegetações abertas da América do Sul podem apresentar respostas diferentes às oscilações do Quaternário quando comparadas às espécies florestais, possivelmente porque elas podem ser mais tolerantes às variações climáticas ou terem uma capacidade adaptativa maior às essas oscilações, especialmente espécies amplamente distribuídas como *L. angustirostris*.

Ainda que as populações de *L. angustirostris* não apresentem maior variabilidade genética nas áreas estáveis, a influencia de refúgios potenciais do Quaternário na diversificação desta espécie não pode ser completamente descartado porque áreas com menor estabilidade podem ser recolonizadas por indivíduos provenientes de várias populações “estáveis”. Foi possível identificar pelo menos uma área estável geograficamente coincidente com todas as populações de *L. angustirostris* (Figura 3e), exceto a população  $MW_{\text{Caatinga}}$ , onde uma potencial barreira geográfica (Serra de Ibiapaba) separa esta população da potencial área estável no leste da Caatinga. Particularmente no caso das populações  $W_{\text{Cerrado}}$ ,  $E_{\text{Cerrado}}$  e Chaco o tempo de divergência estimado indica uma restrição do fluxo gênico entre essas populações durante o Pleistoceno superior (Figura 7), esta restrição pode ser explicada



possivelmente pelas flutuações climáticas sem ter afetado drasticamente o tamanho populacional.

Por outro lado, o tempo de divergência nas populações da parte mais nordeste (MW\_Caatinga e W\_Caatinga) parece ser bem mais antigo que a escala de tempo coberta pela paleo modelagem desta espécie (Figura 7) então a influência de potenciais refúgios não pode ser completamente atribuída a diversificação desta espécie. Além disso, um estudo recente encontrou uma forte estrutura genética em uma espécie de aranha endêmica das SDTFs (particularmente do núcleo da Caatinga; Magalhaes et al., 2014), com algumas populações genéticas mostrando distribuição geográfica similar as populações MW\_Caatinga e W\_Caatinga. Porém, os eventos geológicos ou climáticos que foram responsáveis pela divergência destas populações continua obscuro e deve ser ainda melhor investigado.

Por fim, nós reconhecemos que mancha de área estável na parte mais a leste do Cerrado pode ser maior no nosso modelo do que a área real em virtude do uso de alguns pontos de ocorrência da Mata atlântica que foram utilizados na modelagem de nicho. Entretanto, estes pontos são possivelmente explicados pela invasão das áreas desmatadas da Mata Atlântica e nos acreditamos que este viés não invalida nossos paleo modelos.

#### *Hipótese de Isolamento por distância*

Nós não encontramos evidência de isolamento por distância nas populações de *L. angustirostris*. Este resultado concorda com outra espécie de uma ave de área aberta (*Zonotrichia capensis*, Campagna et al., 2014) onde um sinal estatisticamente significativo foi observado considerando todas as localidades, porém não identificado em nível populacional. Apesar da significativa correlação entre a distância genética e geográfica para todos os indivíduos, a presença de barreiras genéticas ao longo da área de distribuição da espécie quebra a premissa de que a distância geográfica é a causa primária da diferenciação, o que pode explicar os padrões encontrados para *Lepidocolaptes angustirostris* e *Zonotrichia capensis*. Quanto a população E\_Cerrado que também apresentou uma correlação estatisticamente significativa entre a distância genética e geográfica, a distância entre as localidades foi maior que auto correlação espacial entre elas.

A ausência de isolamento por distância nas populações de *Lepidocolaptes angustirostris* pode ter algumas razões não mutuamente excludentes, como: (1) a capacidade de dispersão da espécie pode ser maior que a esperada; (2) o tempo para o equilíbrio entre o fluxo gênico e a deriva genética pode não ter sido alcançado devido a recente colonização das populações (Castric & Bernatchez, 2003; Pinsky et al., 2010); e (3) o tamanho amostral

pequeno de cada população pode dificultar o controle de outras forças adicionais de variação (Castric & Bernatchez, 2003). Além disso, a explicação para cada população pode ser diferente entre elas e devem ser melhor investigadas dentro de cada bioma.

#### *Evolução da plumagem e implicações taxonômicas*

Nossos resultados indicam que nenhuma das subespécies de *Lepidocolaptes angustirostris* foi recuperada como monofiléticas (Figura 2). Ademais, nenhuma população geneticamente delimitada é do ponto de vista morfológico completamente diagnosticável, e estas populações não coincidem com as subespécies válidas, exceto pela população W\_Caatinga que é formada exclusivamente por indivíduos da subespécie *coronatus*. No entanto, alguns indivíduos da subespécie *coronatus* foram encontrados na população MW\_Caatinga, então W\_Caatinga não pode ser considerada uma unidade taxonômica distinta. Mesmo considerando os grupos descritos com base no padrão de cor de plumagem descrito por Marantz e colaboradores (2003; veja introdução), estes grupos também foram recuperados como parafiléticos. Assim, nossos resultados suportam que *L. angustirostris* representa uma espécie única, concordando com a revisão da variação de plumagem para esta espécie (Bolívar-Leguizamón & Silveira, 2015), contudo ressaltamos a forte estrutura genética e uma distribuição não clinal (ausência de isolamento por distância). De qualquer maneira, a variação na cor da plumagem desta espécie pode ser fruto da heterogeneidade de nicho que pode promover uma seleção disruptiva, favorecendo os indivíduos dos extremos em uma população com distribuição normal (veja Galeotti et al., 2003) onde a presença/intensidade do padrão estriado aumenta gradualmente em direção a latitudes maiores.

Uma análise de som pode ser útil para investigar a diferença entre as populações descritas neste estudo, assim como para avaliar se essas populações representam divergência de organismos em níveis maiores que uma estrutura genética neutra. Como exemplo, um estudo molecular com o gênero *Zimmerius* mostra um padrão de plumagem como uma característica evolutivamente conversada e o som como um sinal melhor de divergência (Rheindt et al., 2008). Sendo assim, este mesmo padrão pode ser válido para *L. angustirostris*.

## **CONCLUSÕES**

As flutuações climáticas do Pleistoceno têm sido evocadas como um dos principais agentes responsável pela diversificação a nível intraespecífico. No entanto, a estrutura filogeográfica de *L. angustirostris* não parece ter sido afetado de maneira primária por tais flutuações climáticas do Pleistoceno. Na verdade, a maioria das populações identificadas neste

trabalho são concordantes com os limites geográficos do Chaco, Cerrado e Caatinga, indicando que a diversificação dentro das linhagens pode ter sido influenciada primariamente pelos processos que moldaram estes biomas (para revisão, veja Werneck, 2011). De forma secundária, o contato entre estes biomas, possivelmente favorecido pela expansão em períodos de condições climáticas favoráveis, pode ter facilitado a troca de indivíduos entre populações diferentes que evoluíram em cada bioma. Assim, nossos dados reforçam a ideia inicialmente proposta por Pennington e colaboradores (2000), corroborada por outros pesquisadores (ex: Werneck et al., 2012a; Fouquet et al., 2014), que os diferentes biomas da diagonal de áreas abertas devem ser considerados unidades independentes do ponto de vista biogeográfico. Estudos futuros com foco em espécies de ampla distribuição neste diagonal devem considerar os processos intrínsecos de cada bioma.

## REFERENCIAS

- Aleixo, A. (2004) Historical diversification of a terra-firme forest bird superspecies: a phylogeographic perspective on the role of different hypotheses of Amazonian diversification. *Evolution*, **58**, 1303–1317.
- Axelsson, E., Smith, N.G.C., Sundström, H., Berlin, S., & Ellegren, H. (2004) Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Molecular Biology and Evolution*, **21**, 1538–1547.
- Bandelt, H.J., Forster, P., & Röhl, A. (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology*, **16**, 37–48.
- Bolívar-Leguizamón, S. & Silveira, L.S. (2015) Morphological variation and taxonomy of *Lepidocolaptes angustirostris* (Vieillot, 1818) (Passeriformes: Dendrocolaptidae). *Papéis avulsos de Zoologia*, **55**, 281–316.
- Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., & Bollmann, K. (2013) Selecting from correlated climate variables: A major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**, 971–983.
- Busby, J.R. (1991) BIOCLIM A Bioclimate Analysis and Prediction System. *Plant Protection Quarterly*, **6**, 8–9.
- Cabanne, G.S., D’Horta, F., Sari, E., Santos, F., & Miyaki, C. (2008) Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): Biogeography and systematics implications. *Molecular Phylogenetics and Evolution*, **49**, 760–773.
- Cabanne, G.S., Santos, F.R., & Miyaki, C.Y. (2007) Phylogeography of *Xiphorhynchus*

- fuscus* (Passeriformes, Dendrocolaptidae): Vicariance and recent demographic expansion in southern Atlantic forest. *Biological Journal of the Linnean Society*, **91**, 73–84.
- Caetano, S., Prado, D., Pennington, R.T., Beck, S., Oliveira-Filho, A., Spichiger, R., & Naciri, Y. (2008) The history of Seasonally Dry Tropical Forests in eastern South America: Inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology*, **17**, 3147–3159.
- Campagna, L., Kopuchian, C., Tubaro, P.L., & Loughheed, S.C. (2014) Secondary contact followed by gene flow between divergent mitochondrial lineages of a widespread Neotropical songbird (*Zonotrichia capensis*). *Biological Journal of the Linnean Society*, **111**, 863–868.
- Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T., & Moritz, C. (2009) Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. *Science*, **323**, 785–789.
- Castric, V. & Bernatchez, L. (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchell). *Genetics*, **163**, 983–996.
- Chapman, F.M. (1917) The distribution of bird-life in Colombia: a contribution to a biological survey of South America. *Bulletin of the American Museum of Natural History*, **36**, 1–728.
- Cortes, A.L.A., Rapini, A., & Daniel, T.F. (2015) The *Tetramerium* lineage (Acanthaceae: Justiceae) does not support the Pleistocene Arc hypothesis for South American seasonally dry forests. *American Journal of Botany*, **102**, 992–1007.
- Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, **7**, 214.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.
- Excoffier, L., Laval, G., & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*, **1**, 47–50.
- Excoffier, L., Smouse, P.E., & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Fouquet, A., Santana Cassini, C., Fernando Baptista Haddad, C., Pech, N., & Trefaut Rodrigues, M. (2014) Species delimitation, patterns of diversification and historical

- biogeography of the Neotropical frog genus *Adenomera* (Anura, Leptodactylidae). *Journal of Biogeography*, **41**, 855–870.
- French, N., Yu, S., Biggs, P., Holland, B., Fearnhead, P., Binney, B., Fox, A., Grove-White, D., Leigh, J., Miller, W., Muellner, P., & Carter, P. (2013) Evolution of *Campylobacter* species in New Zealand. *Campylobacter Ecology and Evolution* (ed. by S. Sheppard and G. Méric), pp. 221–240. Norfolk.
- Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.
- Furley, P.A. & Metcalfe, S.E. (2007) Dynamic changes in savanna and seasonally dry vegetation through time. *Progress in Physical Geography*, **31**, 633–642.
- Galeotti, P., Rubolini, D., Dunn, P.O., & Fasola, M. (2003) Colour polymorphism in birds: Causes and functions. *Journal of Evolutionary Biology*, **16**, 635–646.
- Giugliano, L.G., Collevatti, R.G., & Colli, G.R. (2007) Molecular dating and phylogenetic relationships among Teiidae (Squamata) inferred by molecular and morphological data. *Molecular Phylogenetics and Evolution*, **45**, 168–179.
- Haffer, J. (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.
- Hammer, Ø., Harper, D.A.T. a. T., & Ryan, P.D. (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, **4(1)**, 1–9.
- Hardy, O.J. & Vekemans, X. (1999) Isolation by distance in a continuous population: Reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.
- Heled, J. & Drummond, A.J. (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.
- Heled, J. & Drummond, A.J. (2010) Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*, **27**, 570–580.
- Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.
- Hey, J. & Nielsen, R. (2004) Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

- IUCN (2017) Available at: [www.iucnredlist.org](http://www.iucnredlist.org).
- Ledru, M. (2002) Late Quaternary history and evolution of the cerrados as revealed by palynological records. *Radiocarbon* *The cerrados of Brazil*, 33–50.
- Lima, C.A. (2016) Influência de eventos históricos e recentes na filogeografia e estrutura genética de *Neothraupis fasciata* ( Aves : Thraupidae ), uma espécie endêmica do Cerrado Cássia Alves Lima Influência de eventos históricos e recentes na filogeografia e estrutura gen. *Tese de doutorado apresentado ao programa de pós Graduação em Biologia Animal da Universidade de Brasília* pp. 146.
- Magalhaes, I.L.F., Oliveira, U., Santos, F.R., Vidigal, T.H.D.A., Brescovit, A.D., & Santos, A.J. (2014) Strong spatial structure, Pliocene diversification and cryptic diversity in the Neotropical dry forest spider *Sicarius cariri*. *Molecular Ecology*, **23**, 5323–5336.
- Marantz, C., Aleixo, A., Bevier, L., & Patten, M. (2003) Available at: <http://www.hbw.com/species/narrow-billed-woodcreeper-lepidocolaptes-angustirostris>.
- Marini, M.A. & Hackett, S. (2002) A Multifaceted Approach to the Characterization of an Intergeneric Hybrid Manakin (Pipridae) from Brazil. *The Auk*, **119**, 1114–1120.
- Mayle, F.E. (2004) Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. *Journal of Quaternary Science*, **19**, 713–720.
- Mayle, F.E. (2006) The Late Quaternary Biogeographical History of South American Seasonally Dry Tropical Forests: Insights from Palaeo-Ecological Data. *Neotropical Savannas and Seasonally Dry Forests Plant Diversity, Biogeography, and Conservation*, 395–416.
- Miller, M.P. (2005) Alleles In Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**, 722–724.
- Moraes, E.M., Yotoko, K.S.C., Manfrin, M.H., Solferini, V.N., & Sene, F.M. (2009) Phylogeography of the cactophilic species *Drosophila gouveai*: Demographic events and divergence timing in dry vegetation enclaves in eastern Brazil. *Journal of Biogeography*, **36**, 2136–2147.
- Nix, H.A. (1986) A Biogeographic Analysis of Australian Elapid Snakes. *Atlas of Elapid Snakes of Australia*. (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Pacheco, M.A., Battistuzzi, F.U., Lentino, M., Aguilar, R.F., Kumar, S., & Escalante, A.A. (2011) Evolution of modern birds revealed by mitogenomics: Timing the radiation and origin of major orders. *Molecular Biology and Evolution*, **28**, 1927–1942.

- Pennington, R.T., Prado, D.E., & Pedry, C.A. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261–273.
- Phillips, S., Dudik, M., & Schapire, R. (2010) Available at: <https://www.cs.princeton.edu/~schapire/maxent/>.
- Pinsky, M., Humberto, L., Montes Jr, R., & Palumbi, S.R. (2010) Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution*, **64**, 2688–2700.
- Prado, D.E. & Gibbs, P.E. (1993) Patterns of Species Distributions in the Dry Seasonal Forests of South America. *Annals of the Missouri Botanical Garden*, **80**, 902.
- Rambaut, A., Suchard, M.A., Xie, D., & Drummond, A.J. (2014) Available at: <http://beast.bio.ed.ac.uk/Tracer>.
- Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution*, **19**, 2092–2100.
- Ramos, A.C.S., De Lemos-Filho, J.P., & Lovato, M.B. (2009) Phylogeographical structure of the neotropical forest tree *Hymenaea courbaril* (Leguminosae: Caesalpinioideae) and its relationship with the vicariant *Hymenaea stigonocarpa* from Cerrado. *Journal of Heredity*, **100**, 206–216.
- Rheindt, F.E., Norman, J.A., & Christidis, L. (2008) DNA evidence shows vocalizations to be a better indicator of taxonomic limits than plumage patterns in *Zimmerius tyrant-flycatchers*. *Molecular Phylogenetics and Evolution*, **48**, 150–156.
- Ribas, C.C., Gaban-Lima, R., Miyaki, C.Y., & Cracraft, J. (2005) Historical biogeography and diversification within the Neotropical parrot genus *Pionopsitta* (Aves: Psittacidae). *Journal of Biogeography*, **32**, 1409–1427.
- Ribas, C.C., Maldonado-Coelho, M., Smith, B.T., Cabanne, G.S., D’horta, F.M., & Naka, L.N. (2012) Towards an Integrated Historical Biogeography of the Neotropical Lowland Avifauna: Combining Diver-Sification Analysis and Landscape Evolution. *Ornitologia Neotropical*, **23**, 187–206.
- Rodríguez-Robles, J.A., Jezkova, T., & Leal, M. (2010) Climatic stability and genetic divergence in the tropical insular lizard *Anolis krugi*, the Puerto Rican “Lagartijo Jardinero de la Montaña.” *Molecular Ecology*, **19**, 1860–1876.
- Rozas, J., Sánchez-De, I., Barrio, J.C., Messeguer, X., & Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.
- Rull, V. (2008) Speciation timing and neotropical biodiversity: The Tertiary-Quaternary

- debate in the light of molecular phylogenetic evidence. *Molecular Ecology*, **17**, 2722–2729.
- Santos, M.G., Nogueira, C., Giugliano, L.G., & Colli, G.R. (2014) Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, **41**, 1506–1519.
- Sedano, R.E. & Burns, K.J. (2010) Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). *Journal of Biogeography*, **37**, 325–343.
- Seddon, N. & Tobias, J.A. (2007) Song divergence at the edge of Amazonia : an empirical test of the peripatric speciation model. 173–188.
- Seeholzer, G.F. & Brumfield, R.T. (2018) Isolation by distance, not incipient ecological speciation, explains genetic differentiation in an Andean songbird (Aves: Furnariidae: *Cranioleuca antisiensis*, Line-cheeked Spinetail) despite near threefold body size change across an environmental gradient. *Molecular Ecology*, **27**, 279–296.
- Silva, J.M.C. Da & Bates, J.M. (2002) Biogeographic Patterns and Conservation in the South American Cerrado: A Tropical Savanna Hotspot. *BioScience*, **52**, 225.
- Sokal, R.R., Oden, N.L., & Thomson, B.A. (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. *Biological Journal of the Linnean Society*, **60**, 73–93.
- Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., & Mindell, D.P. (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular phylogenetics and evolution*, **12**, 105–114.
- Stephens, M., Smith, N.J., & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978.
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Tajima, F. & Nei, M. (1984) Estimation of Evolutionary Distance between Nucleotide Sequences. *Molecular biology and evolution*, **1**, 269–285.
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., & Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.



- Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F., & Palma-Silva, C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.
- Vanzolini, P.E. (1963) Problemas faunísticos do Cerrado. *Ferri, M. (editor) Simpósio sobre o Cerrado. Editora da Universidade de São Paulo. São Paulo*
- Werneck, F.P. (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.
- Werneck, F.P., Costa, G.C., Colli, G.R., Prado, D.E., & Sites, J.W. (2011) Revisiting the historical distribution of Seasonally Dry Tropical Forests: New insights based on palaeodistribution modelling and palynological evidence. *Global Ecology and Biogeography*, **20**, 272–288.
- Werneck, F.P., Gamble, T., Colli, G.R., Rodrigues, M.T., & Sites, J.W. (2012a) Deep diversification and long-term persistence in the south american “dry diagonal”: Integrating continent-wide phylogeography and distribution modeling of geckos. *Evolution*, **66**, 3014–3034.
- Werneck, F.P., Nogueira, C., Colli, G.R., Sites, J.W., & Costa, G.C. (2012b) Climatic stability in the Brazilian Cerrado: Implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography*, **39**, 1695–1706.
- Wright, S. (1943) Isolation by Distance. *Genetics*, **28**, 114–138.
- Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

**CAPÍTULO III**

**ANÁLISE GENÔMICA DA ESTRUTURA POPULACIONAL DE  
*LEPIDOCOLAPTES ANGUSTIROSTRIS*: A INFLUÊNCIA DAS  
FLUTUAÇÕES CLIMÁTICAS E DOS BIOMAS SOB A AVIFAUNA DAS  
ÁREAS ABERTAS DA AMÉRICA DO SUL.**

**RESUMO**

Apesar da alta biodiversidade das áreas abertas da América do Sul sua história biogeográfica é pouco elucidada. Neste sentido, afim de contribuir com o conhecimento da história biogeográfica da avifauna desta região serão testadas três hipóteses não mutuamente excludentes: (I) cada bioma corresponde a uma unidade biogeográfica distinta.; (II) os biomas respondem de forma diferente às variações climáticas; (III) presença de isolamento por distância dentro das populações. Para isso, foram utilizados 379 locos nucleares, e 382 SNPs provenientes da técnica de sequenciamento de nova geração “*anchored loci*”. Foram realizadas uma série de análises a fim de avaliar estrutura populacional, estimativas de tamanho efetivo populacional, tempo de divergência entre linhagens e taxas de migração, diferenças climáticas entre os biomas, teste de independência entre os nichos climáticos dos biomas, um de teste de Mantel buscando indícios de Isolamento por distância, além do teste de três possíveis cenários que fossem capaz de explicar a diversificação da espécie: (I) Isolamento; (II) Isolamento com migração e (III) Contato Secundário. Foram encontradas três populações concordantes geograficamente com os biomas com algum nível de *admixture* entre as populações, no entanto o atual fluxo gênico é negligenciável. O tempo de separação entre as populações é recente datado do Pleistoceno. Os biomas são climaticamente diferentes e possuem nichos climáticos independentes. Indícios de isolamento por distância foram encontrados apenas na população do Chaco. As análises dos modelos rejeitam os cenários de Isolamento com Migração e Conato Secundário, sendo dentre os cenários avaliados o Isolamento com ausência de Migração o cenário que melhor explica a diversificação em *L. agustirostris*. Concluimos, portanto, que a independência dos biomas por si só é importante na diversificação da avifauna, e que as mudanças climáticas atuam no nível inter e intra populacional. Por fim, sugerimos a revisão taxonômica propondo apenas três subespécies para esta espécie.

## INTRODUÇÃO

As formações de vegetações abertas da América do Sul abrigam uma alta biodiversidade, no entanto possui uma história biogeográfica pouco conhecida quando comparada às áreas florestais (Turchetto-Zolet et al., 2013). A diagonal de áreas abertas da América do Sul é composta pelas Florestas Tropicais Sazonalmente Secas (com a maior área no nordeste do Brasil: a Caatinga), o Cerrado (Brasil central), e o Chaco (nordeste da Argentina, Paraguai ocidental, e sudeste da Bolívia) (Vanzolini, 1963). Estes biomas formam um complexo mosaico de vegetação que evolui sob forte pressão dos períodos secos (Furley & Metcalfe, 2007). Apesar das semelhanças entre estes biomas, estes devem ser considerados unidades biogeográficas distintas (Pennington et al., 2000). No entanto, tais relações não são ainda bem elucidadas.

As ligações faunísticas (AbSaber, 1977; Andrade-Lima, 1982) e os padrões morfológicos em aves (Hayes, 2001) sugerem que o Cerrado e a Caatinga sejam mais próximos entre si quando comparados ao Chaco. Por outro lado, existem evidências de que o Cerrado seja mais próximo ao Chaco quando comparados à Caatinga, como: a filogenia morfológica do gênero *Caenomoda* (Zanella, 2002); análises considerando distribuição e endemismos da herpetofauna (Colli, 2005) e das as avifauna dos biomas (Porzecanski & Cracraft, 2005). Contudo, estas hipóteses precisam ainda ser testadas já que podem representar apenas padrões de similaridade ecológica e os resultados podem ter um viés deste padrão (Werneck, 2011).

As flutuações climáticas já vem sendo evocado na explicação da diversificação da biodiversidade da América do Sul (Turchetto-Zolet et al., 2013) e mais especificamente nas áreas abertas já se tem registros dessa influência na diversificação de diversos táxons, como plantas (Collevatti et al., 2015; Vitorino et al., 2018) aranhas (Bartoleti et al., 2017) e lagartos (Werneck et al., 2012). Outra evidência do efeito das mudanças climáticas foi favorecer o intercâmbio da avifauna florestal via áreas abertas (Batalha-Filho et al., 2013). Os corredores florestais ligando a Mata Atlântica e a Amazônia vem sendo cada vez mais elucidados, como exemplo, as conexões via Cerrado estabelecidas mais de uma vez ao longo do tempo, e que aconteceram tanto no LGM (*Last Glacial Maximum*) quanto no LIG (*Last Interglacial*), e a conexão via Caatinga estabelecida no LGM (Ledo & Colli, 2017). Além disso, já foi reportada a conexão entre Mata Atlântica e as Florestas dos Andes que também se deram via Cerrado e que também aconteceram no período do LGM (Trujillo-Arias et al., 2017; Cabanne et al., *in review*). Vale ressaltar, que ainda não foram reportadas conexões florestais passando pelo

Chaco. Estas conexões podem ter isolado as populações de *L. angustirostris* atuando na diversificação da espécie.

Considerando a grande amplitude de área dos três biomas, outro fator que pode ter influenciado a diversificação dentro dos biomas é o isolamento por distância. Sendo este o caso, é esperado o aumento da distância genética em função do aumento da distância geográfica (Wright, 1943). Esta relação é influenciada pela capacidade de dispersão da espécie, se estabelecendo entre áreas geograficamente mais próximas, e em casos de distribuições amplas os extremos desta distribuição tendem a ser mais diferentes em virtude deste processo.

O arapaçu-do cerrado (*Lepidocolaptes angustirostris*) foi escolhido como modelo para identificar processos de diversificação nas áreas abertas por que possui distribuição coincidente em quase sua totalidade com a distribuição da diagonal de biomas de vegetação aberta da América do Sul (Figura 1). Além disso, a análise do DNA mitocondrial desta espécie (Capítulo I) sugere que os biomas tenham tido forte influência na diversificação desta espécie, uma vez que encontramos linhagens geograficamente coincidentes com cada um dos biomas.

Esta espécie possui também incongruências entre a taxonomia atual (Marantz et al., 2003) e as evidências morfológicas (Bolívar-Leguizamón & Silveira, 2015) e genéticas (Capítulo I). Atualmente são reconhecidas oito subespécies, sendo elas: *L. a. angustirostris*, *L. a. praedatus*, *L. a. certhiolus*, *L. a. hellmayri*, *L. a. bivitattus*, *L. a. griseiceps*, *L. a. coronatus* e *L. a. bahiae*. Contudo, estudo com base na variação de plumagem sugere uma única espécie (Bolívar-Leguizamón & Silveira, 2015), enquanto a análise do DNA mitocondrial corrobora uma única espécie, no entanto com subdivisões que não condizem geograficamente com as atuais subespécies (Capítulo I).

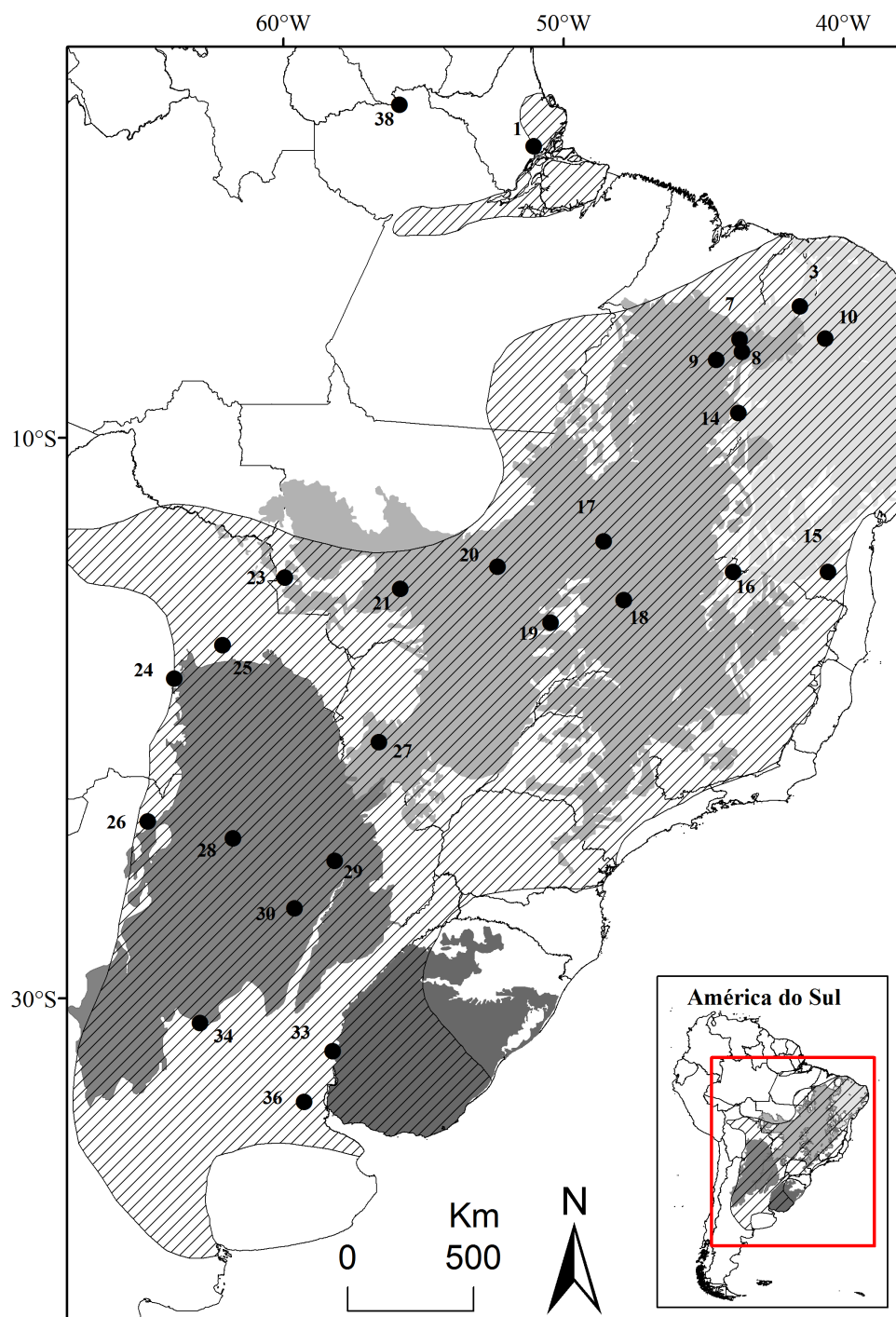
Por tanto, neste trabalho serão testadas três hipóteses não mutuamente excludentes: (I) cada bioma corresponde a uma unidade biogeográfica distinta: se cada bioma é independente é esperado que cada bioma tenha uma população independente, evidenciado pela Com isso, espera-se encontrar estrutura populacional coincidente com os biomas, altos valores de FST entre as populações de cada bioma além de baixas taxas de migração entre as populações; (II) os biomas respondem de forma diferente às variações climáticas do Pleistoceno. Se as flutuações climáticas foram importantes para cada bioma é esperado que o clima seja um fator diferente entre os biomas, então é cada bioma seja independente climaticamente e que os tempos de diversificação das linhagens sejam coincidentes com o Pleistoceno; (III) considerando a ampla área dentro de cada bioma e a baixa vagilidade da espécie, é esperado encontrar isolamento por distância dentro de cada população, para tanto espera-se encontrar

uma correlação positiva entre a distância genética e a distância geográfica dentro de cada bioma.

## **MATERIAL E MÉTODOS**

### *Amostragem*

Foram amostrados 27 indivíduos proveniente de 25 localidades distribuídos ao longo de toda distribuição da diagonal de áreas abertas (Figura 1, Tabela 1). As amostras foram coletadas durante expedições de campo realizadas pela equipe do Laboratório de Genética e Biodiversidade da Universidade de Brasília, além de gentilmente cedidas por duas instituições nacionais: Museu Paraense Emílio Goeldi (MPEG) e Museu de Zoologia da Universidade de São Paulo (MZUSP); e duas instituições internacionais: Museo Argentino de Ciencias Naturales Bernardino Rivadavia (MACN) e Burke Museum of Natural History, and Culture – USA (UWBM). A localidade Sipaliwini (LOC 38) é proveniente de um trabalho anterior (Prum et al 2015) utilizando a mesma metodologia de sequenciamento, no entanto esta localidade foi incluída apenas nas análises com dados de Sequências de polimorfismo único (SNP) em virtude do formato dos dados disponibilizados por Prum e colaboradores (2015)



**Figura 1.** Mapa das localidades amostradas genética, os detalhes das localidades estão disponíveis na Tabela 1. A área hachurada representa a distribuição da espécie segundo a Birdlife internacional. Os tons de cinza correspondem aos biomas: Cinza-claro: Caatinga, Cinza-médio: Cerrado; Cinza-escuro: Chaco e Savanas do Uruguai. Os números das localidades seguem a numeração definida no Capítulo II.

**Tabela 1.** Descrição das localidades utilizadas na obtenção dos dados genéticos. Os números LOC fazem referencia a figura 1. A origem refere-se a procedência do material biológico, as siglas são descritas na seção *Amostragem*. <sup>a</sup> - Dados brutos gentilmente cedida por Jacob Berv proveniente do trabalho Prum et al 2015

LOC	Amostra	Localidade	Cidade-Estado	País	Origem	LAT	LONG
1	UFG 4011	Faz. Embrapa	Macapá - AP	Brasil	LGBUNB	0.4	-51.05
3	UFG 2600	Fazenda Bonito	Castelo de Piauí - PI	Brasil	MPEG	-5.31	-41.55
7	UFG 2601	Mancha verde	São João dos Patos - MA	Brasil	MPEG	-6.49	-43.7
8	UFG 3913	Fazenda São Pedro	Guadalupe - PI	Brasil	MPEG	-6.93	-43.61
9	UFG 2615	Vale do Rio Pratinha	Uruçuí - PI	Brasil	MPEG	-7.23	-44.54
10	UFG 3916	Fazenda Arsênio	Parambú - CE	Brasil	MPEG	-6.46	-40.65
14	UFG 3918	P.N. Serra das Confusões	Caracol - PI	Brasil	MPEG	-9.12	-43.75
15	UFG 3227	Barra do Choça	Bahia	Brasil	MUZUSP	-14.8	-40.55
16	UFG 1053	Lagoa do Cajoeiro	Matias Cardoso - MG	Brasil	LGBU	-14.8	-43.93
17	UFG 3228	Campinaçu	Goiás	Brasil	MUZUSP	-13.7	-48.56
18	UFG 3685	E. E. Águas Emendadas	Brasília - DF	Brasil	LGBUNB	-15.8	-47.84
18	UFG 2405	Reserva Ecológica do IBGE	Brasília - DF	Brasil	LGBUNB	-15.8	-47.84
19	UFG 0465	Fazenda Morro Branco	Aurilândia - GO	Brasil	LGBUNB	-16.6	-50.45
20	UFG 3002	Fazenda Ouro e Prata	Nova Xavantina - MT	Brasil	LGBUNB	-14.6	-52.35
21	UFG 3621	P.N. Chapada dos Guimarães	Chapada dos Guimarães - MT	Brasil	LGBUNB	-15.4	-55.82
23	UFG 3770	Vila Bela da Santíssima Trindade	Vila Bela - MT	Brasil	MUZUSP	-15	-59.94
24	RIS 107	Laguna Caucaya	Departamento de Santa Cruz	Bolivia	UWBM	-18.6	-63.91
25	MACN 5928	Centro de Producción Yabaré	Chiquitos - Dpto Santa Cruz	Bolivia	MACN	-17.4	-62.17

25	MACN 5944	Centro de Producción Yabaré	Chiquitos - Dpto Santa Cruz	Bolivia	MACN	-17.4	-62.17
26	MACN 5663	Parque Nacional Calilegua	Provincia de Jujuy	Argentina	MACN	-23.7	-64.85
27	UFG 4284	Fazenda Pitangueiras	Bonito - MS	Brasil	LGBUNB	-20.8	-56.58
28	MACN 5434	Reserva Natural Formosa	Provincia de Formosa	Argentina	MACN	-24.3	-61.8
29	MACN 3191	Parque Nacional Pilcomayo	Provincia de Formosa	Argentina	MACN	-25.1	-58.17
30	MACN 4506	Parque Nacional Chaco	Provincia de Chaco	Argentina	MACN	-26.8	-59.6
33	MACN 4339	Parque Nacional El Palmar	Provincia de Entre Rios	Argentina	MACN	-31.9	-58.24
34	MACN 6150	La Para	Provincia de Córdoba	Argentina	MACN	-30.8	-62.98
36	MACN 2117	Isla las Palmas Baradero	Provincia de Buenos Aires	Argentina	MACN	-33.7	-59.25
38 <sup>a</sup>	YPM 2273	Sipalwini Savannah	Sipalwini	Suriname	MACN	1.88	-55.85

---



### *Obtenção dos dados genômicos*

#### *Extração de DNA*

O DNA total foi extraído individualmente utilizando o kit comercial DNAeasy (Quiagen), segundo recomendações do fabricante.

#### *MSN-NGS: construção da biblioteca genômica, enriquecimento e sequenciamento*

Esta etapa do trabalho foi desenvolvida integralmente no *Center for Anchored Phylogenomics* na *Florida State University* em colaboração com os Drs Emily M. Lemmon e Alan R. Lemmon. A metodologia empregada para a obtenção destes dados é sintetizada abaixo.

O DNA genômico de cada indivíduo foi fragmentado em um ultrasonicador Covaris E220, gerando fragmento entre 150 e 880 pares de bases. Aos fragmentos obtidos foram inseridos por ligação seqüências sinais (*tags*) e adaptadores de acordo com (Meyer & Kircher, 2010) e com as modificações propostas por Lemmon e colaboradores (2012), sendo que todas estas etapas foram realizadas na *workstation* automática *Biomek FXP* (Beckman-Coulter, Inc.). Alíquotas equimolares das bibliotecas genômicas obtidas foram misturadas em dois *pools* de 16 indivíduos e, em seguida, estes últimos foram enriquecidos utilizando o *kit SureSelect XT2 Custom DNA/RNA* (Agilent Technologies) e o *pool* de sondas (403 no total) projetadas por Lemmon e colaboradores (2012), segundo recomendações do fabricante. O sequenciamento de nova geração foi realizado de modo *paired-end 250 base pairs* no sistema Illumina HiSeq 2500 no *Translational Science Laboratory in the College of Medicine na Florida State University*.

#### *Processamento das leituras (reads) e montagem dos contigs*

As leituras foram processadas utilizando o pacote de programas Casava (Illumina, Inc.), de acordo com as seguintes etapas principais: (1) filtragem por qualidade de acordo com os parâmetros recomendados pela Illumina e (2) separação das leituras por indivíduos. As leituras pareadas de cada indivíduo foram sobrepostas (*merged*) segundo método descrito em (Rokyta et al., 2012) As leituras que não foram sobrepostas na etapa anterior foram separadas e junto com as leituras sobrepostas utilizadas no processo de *assembly*. As leituras (sobrepostas e não sobrepostas) de cada indivíduo foram utilizadas no processo de *assembly* com as seqüências das sondas utilizadas no enriquecimento da biblioteca com o algoritmo *assemble.class* elaborado por A. R. Lemmon (disponível para solicitação ao autor). Destaca-se que na ausência de contaminação no processo de preparação da biblioteca, o método de *assembly* aplicado anteriormente deverá produzir apenas um *contig* para cada uma das seqüências de referência (ou *loco*) para cada indivíduo analisado. Em seguida, as seqüências consenso de cada *contig* foram extraídas para serem utilizadas nas análises subsequentes. As ambiguidades presentes

nas sequências consenso em decorrência dos sítios polimórficos foram mantidas e codificadas (p.ex. C ou T = Y) de acordo com o código do DNA definido pela *International Union of Pure and Applied Chemistry* (IUPAC). *Contigs* com baixa cobertura (< de 10 leituras) foram excluídas das análises.

#### *Alinhamento das sequências consensos*

As sequências consensos de boa qualidade obtidas para todos os indivíduos foram agrupadas em conjuntos de acordo com o *loco* correspondente. A ortologia entre as sequências consenso foi determinada para cada *loco* individualmente com base na distância genética com os algoritmos *GetPairwiseDistanceMeasures.class* e *allOrthoScripts.sh* elaborados por A. R. Lemmon (disponíveis para solicitação ao autor). Em seguida, as fases alélicas das sequências consenso com códigos de ambiguidade em cada *loco* foram definidas com o programa Phaser 3 (Stephens et al., 2001; Stephens & Scheet, 2005). Sequências consensos sem ambiguidades foram duplicadas para assegurar dois alelos por indivíduo por *loco*. O alinhamento das sequências consensos foi feito para cada *loco* individualmente com o programa MAFFT v7.023b (Kato & Standley, 2013). Em seguida, cada sítio do alinhamento de cada *loco* obtido foi investigado quanto à qualidade da informação com base nas seguintes etapas: primeiro os sítios foram considerados “informativos” quando o mesmo nucleotídeo estava presente em mais de 62% das sequências; segmentos de 20pb de cada sequência/alinhamento que continham < 10 sítios informativos foram identificados e “marcados”; os locais com menos de quatro bases não “marcadas” foram removidos do alinhamento. Este processo de seleção de sítios “informativos” foi feito com o algoritmo *TrimAndMaskRawAlignments2.class* elaborado por A. R. Lemmon (disponível para solicitação ao autor). Por fim, todos os sítios contendo *missing data* foram excluídos dos alinhamentos. Após a seleção de sítios, apenas os *loci* que apresentaram variação entre as sequências consensos foram utilizados nas análises filogenéticas. O programa PYTHON (run\_mraic.py) incluindo uma modificação do MrAIC (Nylander, 2004) foi utilizado para estimar o melhor modelo de substituição pelo critério de Akaike (Akaike, 1974) para cada alinhamento obtido em paralelo.

#### *Obtenção dos Polimorfismos de nucleotídeos únicos (SNPs)*

Para a obtenção dos *SNPs* foi utilizado um protocolo desenvolvido por Michael G. Harvey disponível em [www.github.com/mgharvey/seq\\_pop](http://www.github.com/mgharvey/seq_pop) utilizando o pacote Phyluce (Faircloth, 2015). Em síntese, as sequências de um indivíduo escolhido arbitrariamente (UFG 4011) foram utilizadas como referências para o mapeamento dos *reads* dos demais indivíduos analisados. Depois disso, os *reads* foram cortados ao fim da sequência de referência e

posteriormente agrupados em um único alinhamento. Foi utilizado seguindo o default do script, um mínimo de sete *reads* por loco como corte para eliminação de *indels*. OS *SNPs* foram escolhidos considerando um indivíduo diplóide, seguido por um processo de anotação e posterior restrição aos *SNPs* de alta qualidade seguindo os parâmetros default do script. O arquivo com extensão *.vcf* phaseado foi utilizado nas análises subsequentes deste trabalho.

### *Análise de estrutura genética*

Para testar a hipótese de populações independentes em cada bioma foi feita a inferência da estrutura populacional de *Lepidocolaptes angustirostris* por meio da atribuição de indivíduos às populações utilizando um método de *clustering* com base nas frequências alélicas implementado no programa STRUCTURE 2.3.3 (Pritchard et al., 2000). Nesta análise foram utilizados 382 *SNPs*, obtidos por meio de um arquivo de base gerado utilizando o script *structure\_from\_vcf.py* desenvolvido por Michael G. Harvey e disponível em [www.github.com/mgharvey/seq\\_pop](http://www.github.com/mgharvey/seq_pop). Os *SNPs* foram escolhidos manualmente, de forma aleatória, sendo um *SNP* por loco evitando *singletons* e mantendo uma matriz completa, isto é, nenhum dado faltante. Foi utilizado um modelo de *admixture* com 100.000 réplicas após 10.000 réplicas de *burn-in*, avaliando 15 iterações, com um número de populações variando de um a cinco ( $k = 1 - 5$ ). O valor mais provável do número de populações foi obtido com base na média e desvio padrão de *log-likelihood*, já que há indícios de que o critério empregando  $\Delta k$  de Evanno (Evanno et al., 2005) possa ter um viés que favorece a escolha de duas populações (Janes et al., 2017). O resultado gráfico foi obtido pelo software *online CLUMpak* disponível em <http://clumpak.tau.ac.il>. Já a análise do melhor valor de  $k$  foi avaliada pelo software *online Structure Harvester* disponível em <http://taylor0.biology.ucla.edu/structureHarvester/>.

A fim de avaliar a formação de grupos nós realizamos uma análise de componentes principais (PCA) com os 382 *SNPs* utilizando a função *dudi.pca* do pacote *ade4*. Depois disso, para avaliar o nível da estrutura genética desta espécie, calculamos o valor de  $F_{ST}$  global com base na estimativa de Weir e Cockerham utilizando a função *wc* do pacote *hierfstat*. O valor de  $F_{ST}$  par a par entre as populações foi estimado também com o intuito de avaliar populações geneticamente mais distintas utilizamos a função *genet.dist* com método “Nei87” do pacote *hierfstat*. As duas últimas análises descritas foram realizadas no programa R.

Para avaliar a historia demográfica da espécie, nós estimamos o tamanho efetivo populacional das populações atuais e ancestrais, a taxa de migração entre as populações e o tempo de separação entre as linhagens considerando a relação filogenética obtidas pelo DNA mitocondrial (Figura 2 do Anexo I). Nos obtivemos estas estimativas utilizando o programa G-

phocs (Gronau et al., 2011) com 500 milhões de gerações amostradas a cada 500 gerações. A distribuição gamma foi utilizada nesta análise com o intuito de standardizar a distribuição da taxa de mutação (veja Smith et al., 2014). Para esta análise utilizamos 1 e 2000 como distribuição gamma  $\alpha$  e  $\beta$ , respectivamente. Nós avaliamos a qualidade das corridas pelos valores de ESS (>200) utilizando o programa Tracer (Rambaut et al., 2014).

### *Diferença climática e de nicho entre os biomas*

Para avaliar se há diferença climática entre os biomas, foi realizada uma Análise de componentes principais (PCA) considerando uma matriz de variância-covariância e um *bootstrap* de 1000 réplicas das 19 variáveis bioclimáticas (Nix, 1986; Busby, 1991) obtida na base de dados WorldClim (Hijmans et al., 2005). Os valores para cada localidade foram obtidos utilizando a opção *extract value by point* no programa ArcMap 10.1 considerando o *raster* de clima atual na resolução 30s. Para esta análise utilizamos duas abordagens: (I) uma mais ampla considerando os 105 pontos de ocorrência utilizados para a modelagem de nicho potencial (Capítulo II), sendo que a atribuição dos pontos a um bioma foi feita com base na localização desses pontos seguindo a definição das eco regiões terrestres (disponível em [https://worldmap.harvard.edu/data/geonode:wwf\\_terr\\_ecos\\_oRn](https://worldmap.harvard.edu/data/geonode:wwf_terr_ecos_oRn)), e (II) uma mais restrita considerando apenas os pontos de amostragem das amostras de tecido analisadas neste trabalho. Neste último caso, os pontos foram atribuídos aos biomas com base na população genética a qual pertencem. A abordagem mais ampla foi realizada a fim de construir um panorama de diferença climática entre os biomas evitando o viés das lacunas de nossa amostragem genética; já a abordagem mais restrita visa caracterizar climaticamente as populações genéticas identificadas a fim de avaliar diferenças climáticas entre elas.

Para avaliar se os biomas possuem nichos diferentes, utilizamos os pontos concordantes entre a localização dos biomas e os grupos formados na pca na abordagem ampla (ver resultados) em um teste de identidade de nicho, utilizando as métricas I e D para rejeitar a hipótese de equivalência de nichos (Warren et al., 2008). Os valores de significância de cada teste (valores de p) foram calculados com base em 101 permutações no pacote “*dismo*”, que considera a soma das distribuições das populações par a par, e extrai pontos aleatórios para uma subsequente comparação de quanto o observado desvia do acaso. Para essa análise de nicho foram utilizadas todas as variáveis climáticas (BIO 1 – BIO 19) do *Wordclim*.

### *Isolamento por distância*

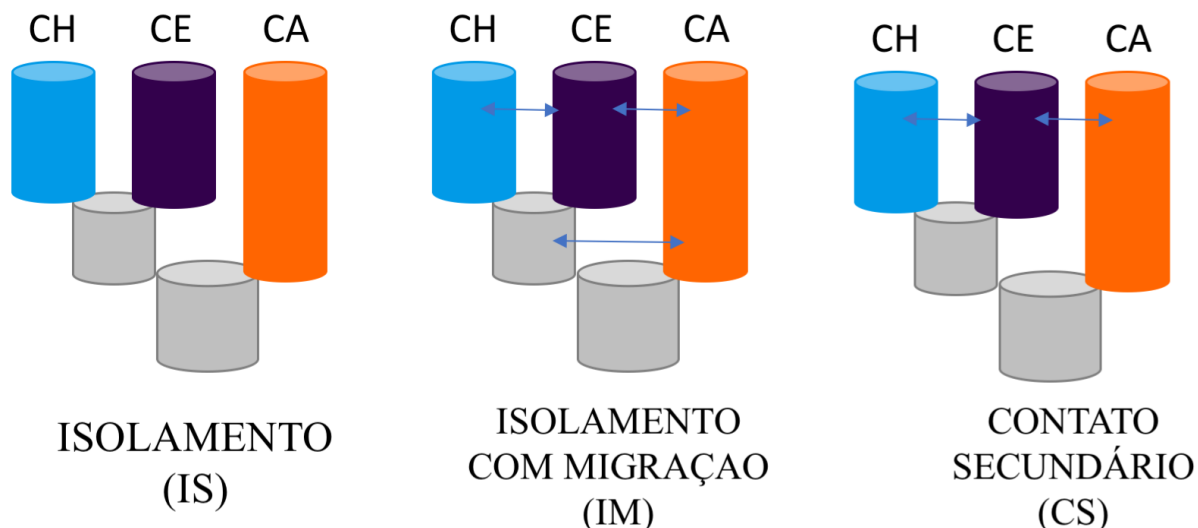
Para testar a hipótese de isolamento por distância dentro de cada bioma, nós realizamos o teste de Mantel utilizando a função *mantel.randtest* do pacote *ade4*. Esta correlação foi realizada entre uma matriz geográfica (km) e uma matriz de distância genética calculada com base na fração de identidade por estado de 382 SNPs por meio da função *snpGdsIBS* do pacote *SNPRelate*.

### *Teste dos possíveis cenários de diversificação*

Para testar três possíveis processos de diversificação das populações de cada bioma, nós propomos os possíveis cenários (Figura 2): isolamento com completa ausência de fluxo gênico entre as populações ancestrais e atuais (IS); isolamento com migração entre as populações ancestrais e as populações atuais (IM); ausência de fluxo gênico entre as populações ancestrais com fluxo gênico reestabelecido entre as populações atuais. Vale ressaltar que consideramos o fluxo gênico ancestral bidirecional e mesma intensidade entre a população da Caatinga e a população Cerrado+Chaco; já o fluxo gênico atual foi considerado também bidirecional e de mesma intensidade entre as populações do Cerrado/Chaco e Cerrado/Caatinga. Ressaltamos que o fluxo gênico entre Caatinga e Chaco foi desconsiderado por não haver nenhum resultado prévio que sustentasse essa possibilidade.

Nós construímos os modelos e realizamos os sumários estatísticos simulados de 379 locos utilizando a função *sim.msABC.sumstat* pacote *PipeMaster* (disponível em <https://github.com/gehara/PipeMaster>) com base em 100 mil simulações. Os modelos foram construídos utilizando os limites inferiores e superiores das estimativas geradas pela análise do G-phocs (ver resultados, Tabela 2), os parâmetros gerados a partir dos *priors* foram transformados para uma escala coalescente, as simulações dos locos foram feitas pelo algoritmo *ms*, a conversão do output do *ms* foi feita para o formato *Popgenome* e então é gerado um arquivo com os sumários estatísticos simulados. O sumário estatístico observado foi obtido pela função *observed.sumstat* do programa *PipeMaster*.

Depois dos sumários estatísticos simulados, realizamos uma *confusion matrix* afim de avaliar o quanto os modelos são diferentes entre si por meio de uma *crossvalidation* onde pseudoréplicas são atribuídas aos modelos e então avalia-se a porcentagem de atribuição correta aos modelos utilizando a função *cv4postpr* do pacote *abc* considerando um algoritmo de rejeição, para este trabalho testamos quatro valores de tolerância: 0,05;0,1;0,2;0,34 e 0,5 afim de avaliar se a estabilidade do resultado.



**Figura 2.** Representação gráfica dos três possíveis modelos testados para explicar a diversificação em *Lepidocolaptes angustirostris*.

A escolha do melhor modelo dentre os testados foi realizada por meio da função *postpr* do pacote *abc* utilizando um algoritmo de rejeição e utilizando as mesmos dois valores de tolerâncias a fim de avaliar a robustez do resultado. Após isso, realizamos uma análise de *goodness-of-fit*, que é um teste estatístico com a mediana do sumário estatístico observado e com a mediana do sumário estatístico simulado não rejeitado utilizando a função *gfit* do pacote *abc*.

## RESULTADOS

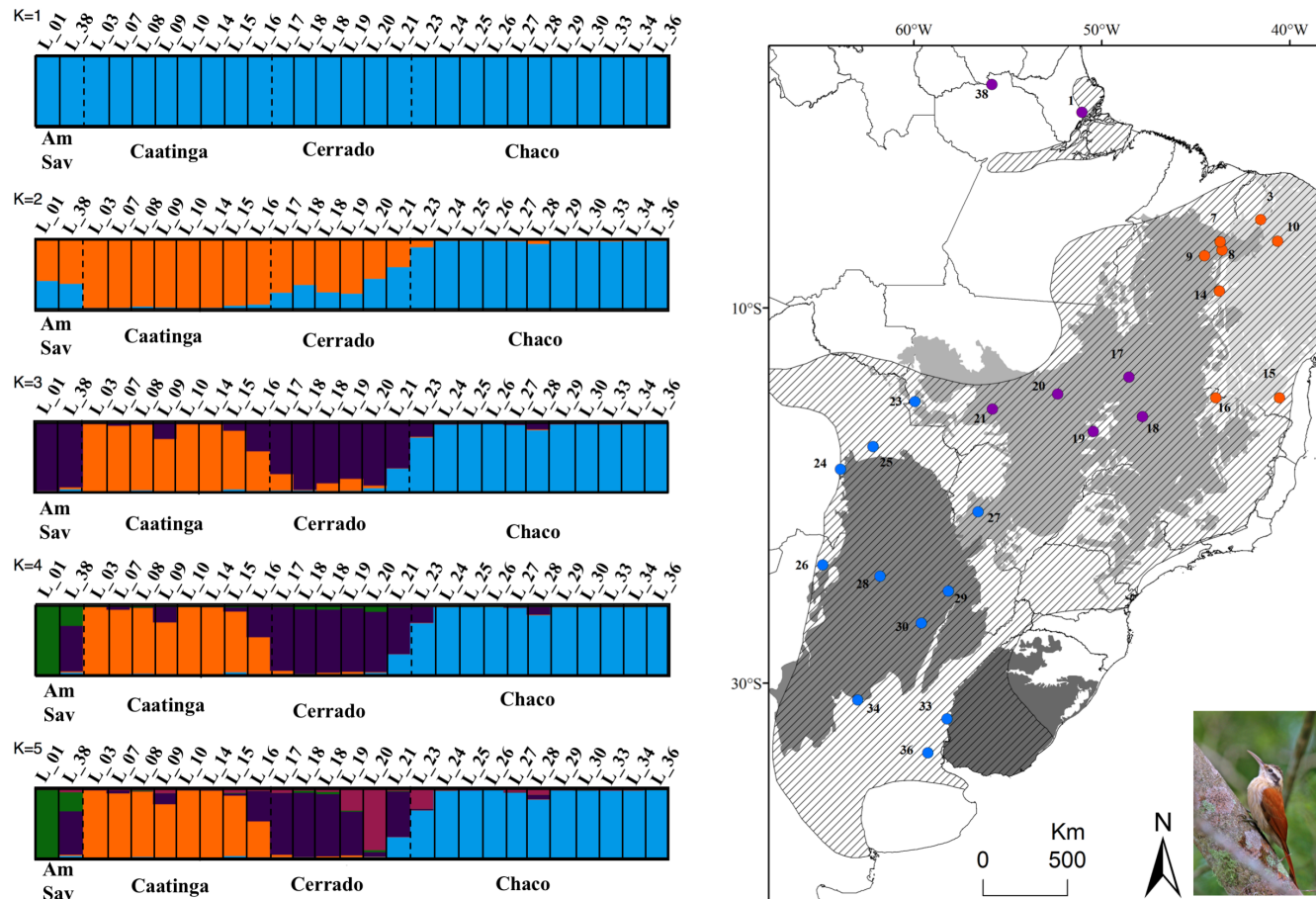
Foram capturados 401 *anchored loci* após a construção das sequências consensus, desses 379 foram selecionados utilizando os critérios de polimorfismo e ausência de dados faltantes. O tamanho médio dos locos foi de 1.524 bases (variando de 345 a 2.579 bases). Foram identificados 401 locos com 4% de dados faltantes, para as análises selecionamos arbitrariamente 382 SNPs considerando apenas um SNP por loco, nenhum loco com dados faltantes e evitando sempre que possível a escolha de *singletons*.

### *Análise de estrutura genética*

Na análise de estrutura populacional, em todas as possibilidades de número de populações ( $k=2$  a  $k=5$ ) foi possível identificar grupos distintos coincidentes geograficamente com a Caatinga e o Chaco. Por outro lado, os indivíduos geograficamente localizados no bioma

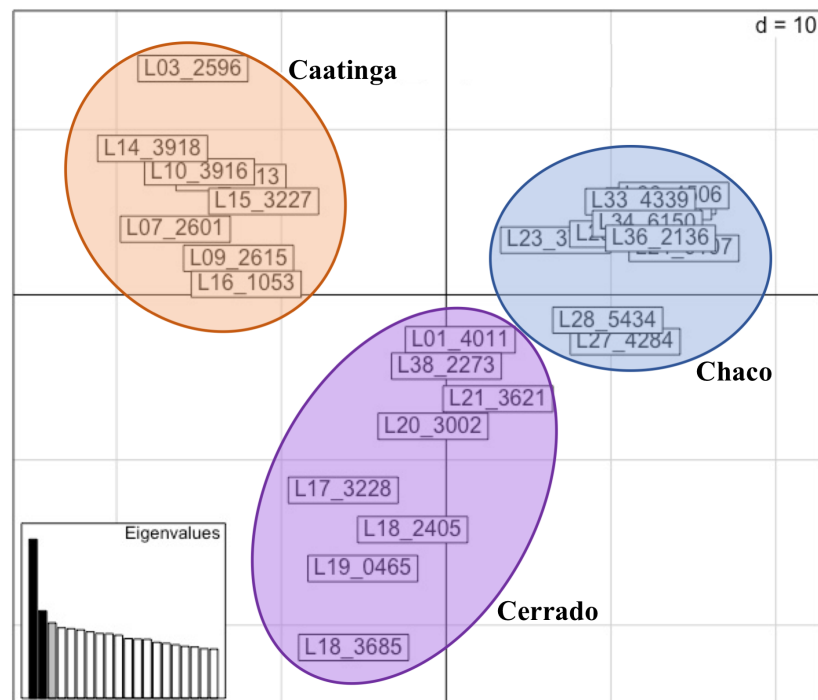
Cerrado são atribuídos à diferentes populações de acordo com o aumento do número de populações (Figura 3). O melhor valor de  $k$  segundo  $\Delta k$  de Evanno foi  $k=2$ . Porém, considerando que esta métrica pode subestimar a estruturação estudada (Janes et al 2016), optamos por trabalhar com a incerteza quanto ao melhor valor de  $k$ . Para todos os casos, os clusters podem ser claramente associados aos biomas, com isso mesmo assumindo que pode haver mais estruturação dentro do Cerrado, assumimos um número mínimo de três populações ( $\text{LnP}(k) = 8514.07 \pm 2.54$ ) para as análises subsequentes. É possível identificar níveis maiores de *admixture* nas localidades de ecótono entre os biomas, exceto pela localidade 28 situada na região core do Chaco que possui um nível de *admixture* com o Cerrado. Neste último caso, podendo explicar por retenção de polimorfismo ancestral e/ou separação incompleta das linhagens.

Os indivíduos localizados nas savanas do norte da Amazônia apresentam altos níveis de *admixture* com a população do Cerrado, acompanhando a estruturação do Cerrado com o aumento dos valores de  $k$ . Além disso, a análise da PCA (Figura 4) apresenta resultados coincidentes com a análise do STRUCTURE com três grupos coincidentes com os biomas, sendo o Chaco e a Caatinga separados pelo eixo  $y$ , enquanto o Cerrado foi separado do Chaco e da Caatinga pelo eixo  $x$ . Apenas dois indivíduos (a localidade 28 coincidente com o resultado anterior, e neste caso a localidade 27 (Bonito – MS) que apesar de não ter sido identificado nenhum nível de *admixture* no STRUCTURE é uma região localizada nos limites geográficos do bioma Cerrado, mas geneticamente atribuída a população do Chaco (STRUCTURE e Capítulo II). Há indícios de algumas áreas definidas geograficamente como Cerrado possuem uma vegetação mais próxima ao do Chaco (Prado, 1993) sugerindo nesse caso dois possíveis cenários não mutuamente excludentes: (I) uma possível zona de contato secundário entre as populações do Cerrado e do Chaco; (II) uma influência da separação incompleta de linhagens e/ou retenção de polimorfismo ancestral. O valor de  $F_{ST}$  global da espécie indica uma moderada e significativa estruturação ( $F_{ST} = 0,152$ ; Frankham et al., 2002; Hartl & Clark, 2007), já as comparações par a par (Figura 4) indicam uma alta diferenciação entre Caatinga e Chaco ( $F_{ST} = 0,209$ ) e uma moderada diferenciação similar entre Cerrado e os demais biomas (Cerrado e Chaco  $F_{ST} = 0,112$ ; Cerrado e Caatinga  $F_{ST} = 0,111$ ).



**Figura 3.** Resultado da análise de estrutura populacional do programa STRUCTURE. São apresentadas nesta figura a estrutura da população considerando de uma a cinco populações. As cores no mapa correspondem aos três principais clados: laranja: Caatinga, Roxo: Cerrado e Azul: Chaco





**Figura 4.** Análise de componentes principais de 382 SNPs. Indicando três grupos coincidentes com os três biomas da diagonal de áreas abertas e representados nesta figura pelas cores laranja: Caatinga, roxo: Cerrado e Azul: Chaco.

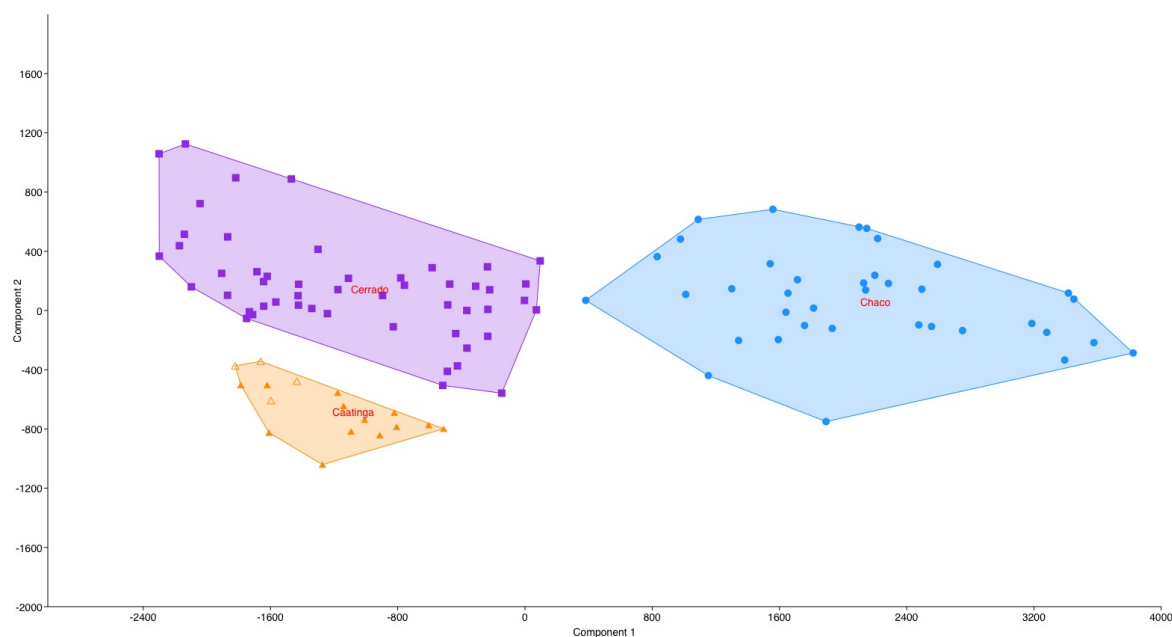
As estimativas de tamanho efetivo populacional (Tabela 2) sugerem uma população ancestral (Caatinga + Cerrado + Chaco) de tamanho muito similar ao tamanho efetivo da população da Caatinga, sendo estimado um tamanho de cerca de 1.617.000 indivíduos para a população ancestral e cerca de 1.633.000 para a população atual da Caatinga. O tamanho efetivo populacional da linhagem que compreende a população Cerrado + Chaco é a maior população estimada em cerca de 5.768.000 sugerindo que esta população possa representar o momento de colonização destes biomas seguido pela diminuição do tamanho efetivo atual na população do Cerrado para cerca de 1.132.000 bem como para a população do Chaco com cerca de 1.875.000 indivíduos. As taxas de migração par a par (Tabela 2) entre os biomas indicam um fluxo gênico mínimo, podendo ser negligenciável, uma vez que os índices sugerem menos de um migrante por geração para todas as comparações. As estimativas de tempo de divergência (Tabela 2) são todas para o pleistoceno, sendo a separação da Caatinga da população ancestral de Cerrado + Chaco datada de cerca de 370 mil anos (364.285 – 385.714) e a separação das linhagens do Cerrado e do Chaco datada de cerca de 250 mil anos (235.714 – 264.285).

**Tabela 2.** Resultado das estimativas de tamanho populacional, tempo de separação e taxas de migração com base em 379 locos utilizando o programa G-phocs. 95 L indica o limite superior, e 95H indica o limite superior da estimativa 95HPD.

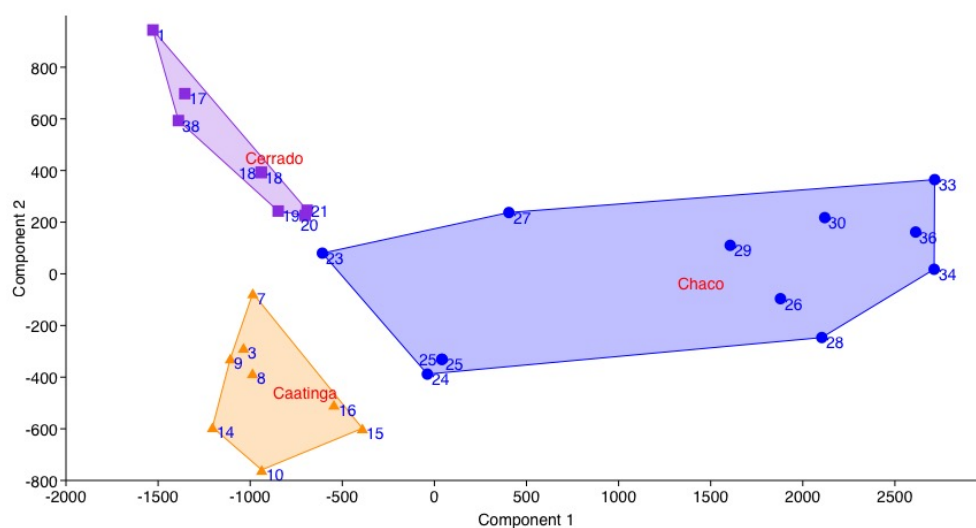
<b>Tamanho efetivo</b>	<b>theta</b>	<b>Indivíduos (Ne)</b>	<b>theta 95L</b>	<b>theta 95H</b>	<b>Ne 95L</b>	<b>Ne 95H</b>
Caatinga	0,009	1.632.982	0,009	0,009	1.571.429	1.692.857
Cerrado	6,33E-03	1.130.411	0,006	0,007	1.067.857	1.185.714
Chaco	0,011	1.875.000	0,010	0,011	1.803.571	1.946.429
Cerrado + Chaco	0,032	5.767.857	0,028	0,037	4.946.429	6.607.143
Caatinga + Cerrado + Chaco	0,009	<b>1.616.054</b>	0,009	0,009	1.539.286	1.683.929
<b>Tempo de separação</b>	<b>thau</b>	<b>Tempo</b>	<b>thau 95L</b>	<b>thau 95H</b>	<b>Tempo 95L</b>	<b>Tempo 95H</b>
Cerrado - Chaco	0,00035	249.750	0,00033	0,00037	235.714	264.286
Caatinga - Cerrado + Chaco	0,00052	373.479	0,00051	0,00054	364.286	385.714
<b>Migração</b>	<b>m</b>	<b>Migrantes</b>	<b>m 95L</b>	<b>m 95H</b>	<b>migrantes 95L</b>	<b>migrantes 95H</b>
Caatinga - Cerrado	0,209	2,92E-07	0,000	0,584	1,40E-14	8,18E-06
Cerrado - Caatinga	0,224	3,14E-06	0,000	0,662	1,40E-14	9,27E-06
Cerrado - Chaco	0,100	1,40E-06	0,000	0,299	1,40E-14	4,18E-06
Chaco - Cerrado	0,109	1,52E-06	0,000	0,323	1,40E-14	4,52E-06
Caatinga - Chaco	0,105	1,47E-06	0,000	0,315	1,40E-14	4,41E-06
Chaco - Caatinga	0,118	1,65E-07	0,000	0,359	1,40E-14	5,02E-06

### Diferença climática e de nicho entre os biomas

Foram identificados três grupos climáticos diferentes, os quais são coincidentes com os biomas em ambas abordagens: ampla (Figura 5) e na restrita (Figura 6). Os dois primeiros componentes principais explicaram cerca de 98% de toda variação nos dois casos, apresentando também resultados concordantes quanto as variáveis mais associadas a cada componente. O componente principal um explica na abordagem mais ampla cerca de 92% da variação e na mais restrita cerca de 90% da variação, e tem a variável de sazonalidade de temperatura (Bio4) como a mais correlacionada a este eixo. O componente principal dois explica na abordagem mais ampla cerca de 6% da variação e na mais restrita cerca de 8% da variação e possui principalmente variáveis de precipitação (BIO 12, BIO 16, BIO 19, BIO13 e BIO 18) associadas a este eixo, sendo a variável de precipitação anual (BIO 12) a mais significativa dentre elas.



**Figura 5.** Análise de componentes principais ampla das variáveis climáticas (BIO1 – BIO19) de 105 pontos utilizados na modelagem de nicho potencial de *Lepidocolaptes angustirostris* (Capítulo II). Indicando três grupos distintos condizentes com os biomas Caatinga (laranja), Cerrado (roxos) e Chaco (Azul).

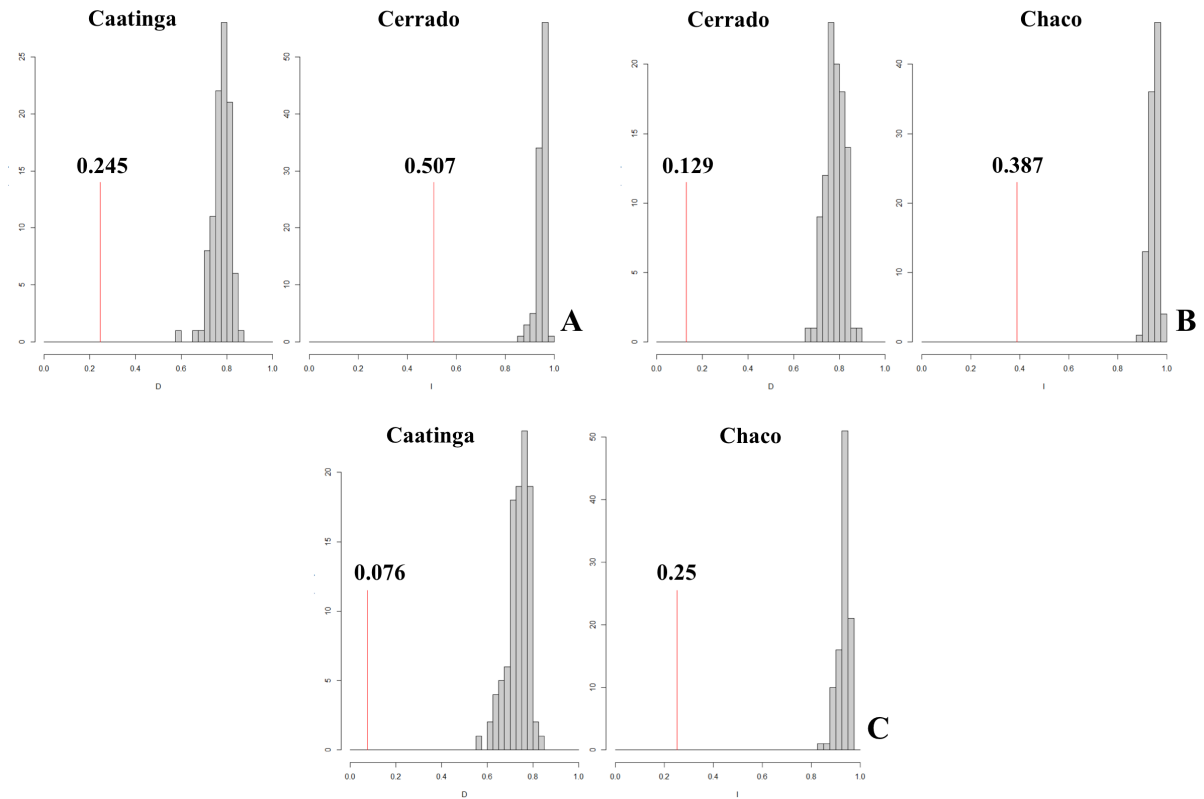


**Figura 6.** Análise de componentes principais restrita das variáveis climáticas (BIO1 – BIO19) de 27 pontos utilizados na análise de estrutura populacional de *Lepidocolaptes angustirostris*. Indicando três grupos distintos condizentes com os biomas Caatinga (laranja), Cerrado (roxo) e Chaco (Azul).

Além disso, não apenas os biomas são climaticamente diferentes (Figura 7), quanto há diferença entre os nichos de cada população de *L. angustirostris* associadas a cada bioma. As comparações par a par são significativas entre todos os biomas (Tabela 3), rejeitando a hipótese de que os nichos são idênticos ou equivalentes, sustentando assim a identidade de cada bioma.

**Tabela 3.** Métricas D e I utilizadas no teste de equivalência de nicho entre os pares de biomas. Os valores de p indicam que as comparações em negrito são estatisticamente significativas. As métricas D e I estão plotadas comparadas as médias de cada nicho na Figura 7.

Comparação	D	p	I	p
Caatinga e Cerrado	0,245	<b>6,24e<sup>-37</sup></b>	0,507	<b>1,53e<sup>-97</sup></b>
Cerrado e Chaco	0,129	<b>1,00e<sup>-54</sup></b>	0,387	<b>2,06e<sup>-196</sup></b>
Caatinga e Chaco	0,076	<b>2,85e<sup>-42</sup></b>	0,250	<b>7,25e<sup>-18 5</sup></b>

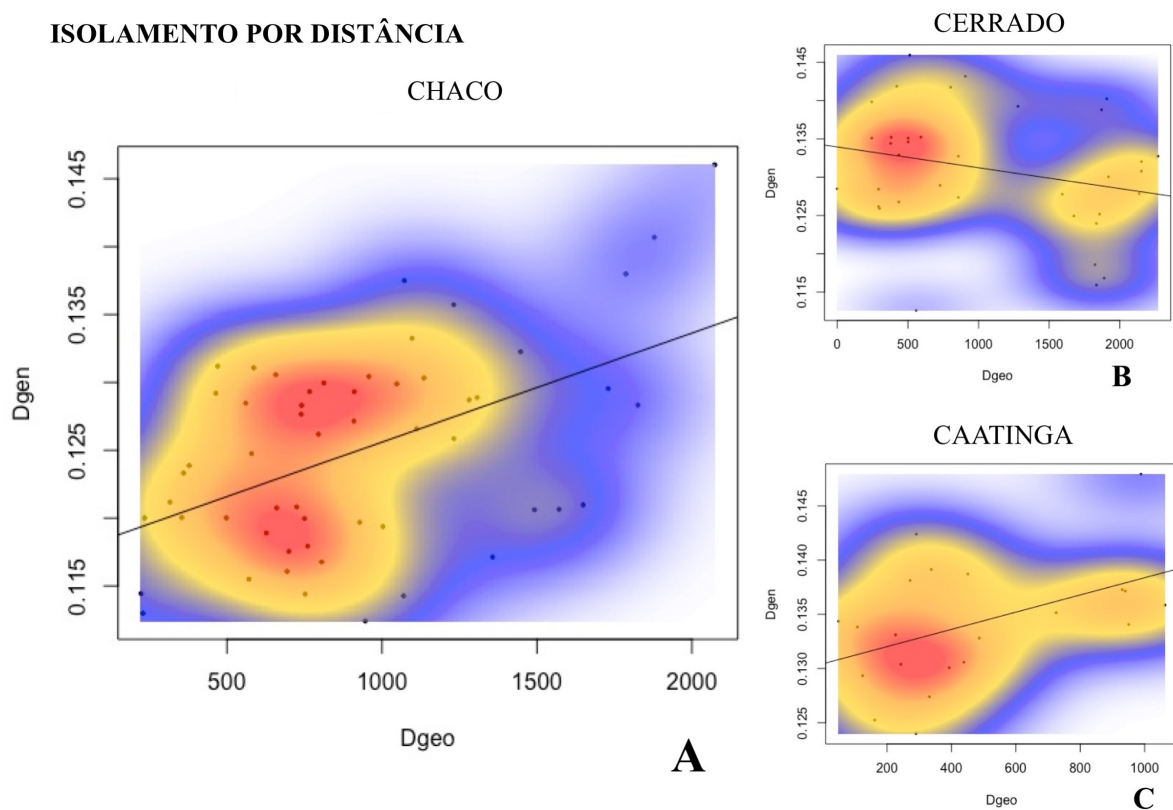


**Figura 7.** Resultado da análise de identidade de nicho. A linha vermelha indica a similaridade de um nicho em relação ao comparado. Todas as comparações rejeitaram a hipótese de equivalência de nichos ( $p < 0.00005$ ). Os valores correspondem as métricas D e I. **A.** Comparação entre as populações da Caatinga e do Cerrado; **B.** Comparação entre as populações do Cerrado e do Chaco; **C.** Comparação entre as populações da Caatinga e do Chaco.

#### *Isolamento por distância*

Nós identificamos uma correlação significativa entre a distância genética e geográfica apenas na população do Chaco ( $r = 0.49$ ,  $p = 0.001$ ; Figura 8a). Já nas outras duas populações, Cerrado ( $r = -0.25$ ,  $p = 0.81$ , Figura 8b) e Caatinga ( $r = 0.46$ ,  $p = 0.11$ , Figura 8c), a diferença genética não está correlacionada à distância geográfica sendo possível rejeitar a hipótese de isolamento por distância nestas populações.

## ISOLAMENTO POR DISTÂNCIA



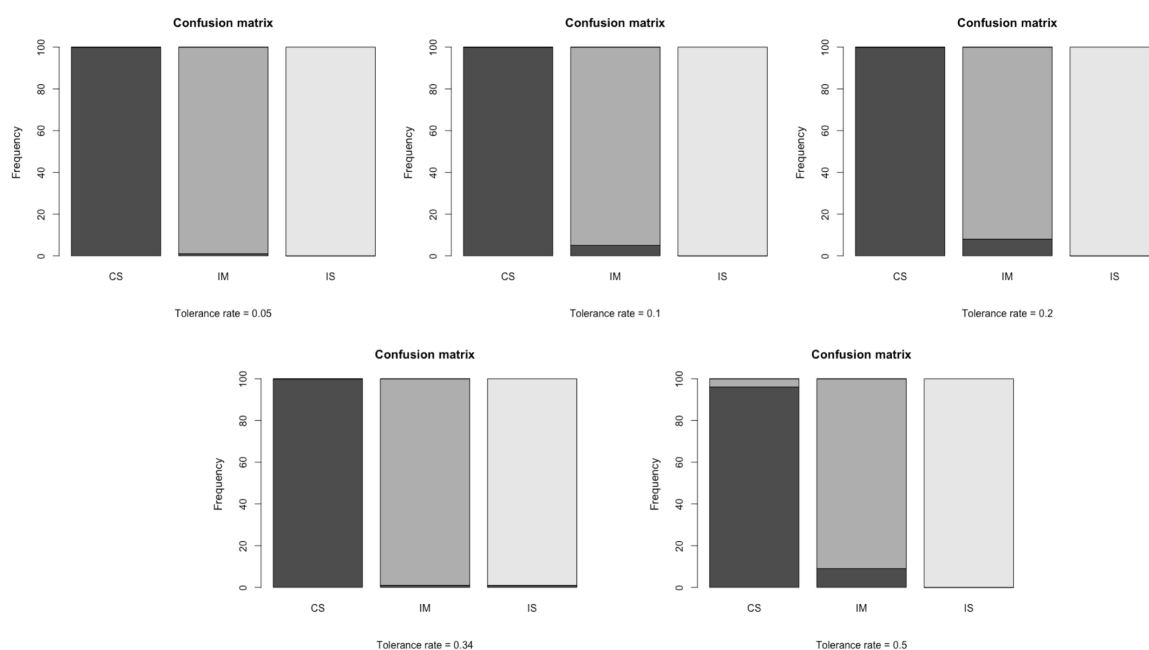
**Figura 8.** Teste de Mantel avaliando a correlação entre a distância genética e geográfica. As cores do gráfico indicam a densidade local com base na estimativa bidimensional da densidade de kernel (kde2d). **A.** Correlação positiva e significativa ( $p=0.001$ ) entre distância genética e geográfica no Chaco sugerindo a presença de Isolamento por distância; **B.** correlação não significativa ( $p=0.81$ ) entre distancia genética e geográfica no Cerrado; **C.** correlação não significativa ( $p=0.11$ ) entre distancia genética e geográfica na Caatinga.

#### *Teste dos possíveis cenários de diversificação*

A análise que determina o quanto os modelos simulados produzem sumários estatísticos diferentes entre si mostra que os modelos são distinguíveis entre si (Figura 9; Tabela4). Em todas as tolerâncias os modelos são distinguíveis entre si com uma taxa de atribuição ao modelo correto em todos os casos maior que 92%. Em todas as tolerâncias, o modelo Isolamento (IS) foi atribuído corretamente com taxa de 100%. Portanto, todos os casos de atribuição errada de modelo são entre os modelos de Isolamento com Migração (IM) e Contato Secundário(CS)

**Tabela 4.** Porcentagem de atribuição correta das pseudoréplicas dos modelos que explicam a diversificação em *L. angustirostris* da análise de *crossvalidation* sob várias taxas de tolerância.

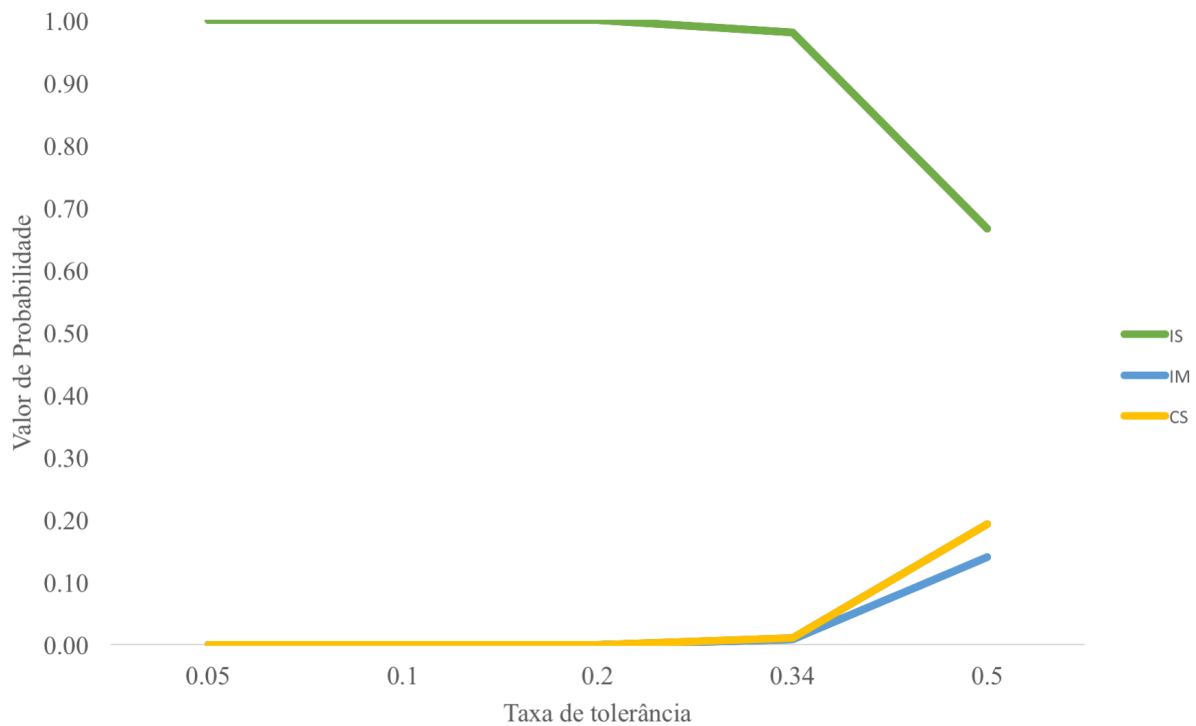
Tolerância	Isolamento (IS)	Isolamento com Migração (IM)	Contato Secundário (CS)
0,05	100	99	100
0,10	100	95	100
0,20	100	92	100
0,34	100	99	100
0,50	100	91	96



**Figura 9.** Confusion matrix avaliando o quanto os modelos são diferentes entre si por meio de uma *crossvalidation* onde pseudoréplicas são atribuídas aos modelos e então avalia-se a porcentagem de atribuição correta aos modelos. CS corresponde ao modelo de Contato Secundário, IM corresponde ao modelo de Isolamento com Migração e IS ao modelo de Isolamento (A representação grafica dos modelos está disponível na Figura 2). As *tolerance rates* indicam o quando de diferença é tolerada para aceitar um modelo.

Considerando ambas as tolerancias os modelos de Isolamento com Migração (IM) e Contato Secundário (CS) foram rejeitados. Dentre os modelos testados o modelo de Isolamento foi o melhor modelo considerando todas as tolerâncias (Figura10). As tolerâncias menores que 0,33 apresentam IS como o melhor modelo com um valor de  $p=1$ . Para tolerância de 0,34 o modelo IS apresenta um valor de  $p=0.98$  e valores similares para os outros modelos  $p=0,1$  para IM e CS. E na maior tolerância 0,5 o modelo IS apresenta um valor de  $p=0,67$  enquanto o

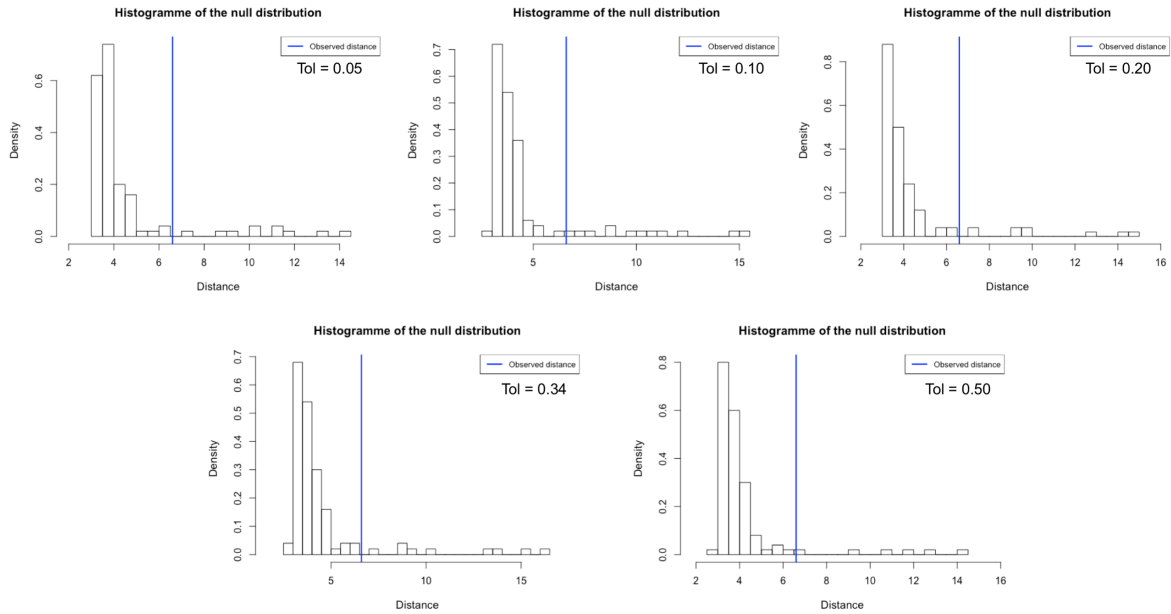
modelo CS apresenta um valor de  $p=0,19$  um pouco maior que o valor do modelo IM, onde  $p=0,19$ .



**Figura 10.** Valores de  $p$  na escolha do melhor sumário estatístico simulado quando comparado ao sumário estatístico observado, considerando os três cenários propostos: Isolamento (IS); Isolamento com Migração (IM) e Contato Secundário (CS) sob cinco taxas de tolerância (0,005;0,10;0,20;0,34 3 0,50) a fim de avaliar a estabilidade da escolha do modelo.

Quando comparamos as medianas pela análise de *goodness-of-fit* (Figura 11) a mediana do sumário estatístico observado não foi diferente do sumário estatístico simulado para o cenário de Isolamento (IS) em todas as tolerâncias sugerindo que este é o melhor modelo dentre os analisados.





**Figura 11.** Análise de goodness-of-fit indicando o valor da mediana observada em função do sumário estatístico simulado para o cenário de Isolamento (IS). As barras indicam os sumários estatísticos simulados

## DISCUSSÃO

Nossos resultados mostram independência biogeográfica entre os biomas uma vez que as populações identificadas são condizentes geograficamente com os biomas, atualmente possuem fluxo gênico negligenciável entre eles, e os cenários sugerindo Isolamento com Migração e Contato secundários foram rejeitados. Além disso, há indícios de que as flutuações climáticas tenham tido papel importante na diversificação desta espécie, já que as estimativas de tempo de separação das populações são condizentes com os períodos de flutuação climática do Pleistoceno, os biomas possuem condições climáticas diferentes e o nicho climático das populações são independentes nos três biomas. Avaliando as populações de cada bioma, apenas a população do Chaco apresenta indícios de isolamento por distância, enquanto o Cerrado apresenta uma sub estruturação e a Caatinga é a população mais homogênea desta espécie. Considerando todos esses resultados, sugerimos uma revisão taxonômica das atuais subespécies, propondo três subespécies, com base na diferença genética e de nicho das três populações descritas neste trabalho.

### *Independência dos biomas*

As populações independentes de *Lepidocolaptes angustirostris* em cada bioma (Figura 3), o fluxo gênico negligenciável entre as populações, como sugerem as taxas de migração estimadas pela análise demográfica deste trabalho (Tabela 2) e a diferença entre as condições climáticas (Figura 5 e Figura 6) os nichos climáticos de cada população (Figura 7) sustentam a hipótese de que os biomas que compõem a diagonal de áreas abertas são unidades biogeográficas independentes como proposto por Pennington e colaboradores (2000). Este padrão de populações independentes é confirmado por outros táxons, como exemplo, lagartos do gênero *Vanzosaura* (Recoder et al., 2014), e da espécie *Phyllopezus pollicaris* (Werneck et al., 2012) que possuem populações independentes no Chaco, Cerrado e na Caatinga. Assim como, em insetos do gênero *Zoniopoda* (Pocco et al., 2018) e aranhas (*Nephila sexpunctata*, Bartoletti et al., 2017) que possuem populações independentes no Chaco e no Cerrado.

A rejeição dos cenários incluindo fluxo gênico atual e/ou passado sugerem que as pressões seletivas nos biomas podem ser fatores muito fortes impedindo o fluxo gênico desde a colonização destes biomas. No entanto, reconhecemos que os modelos ainda precisam ser melhor explorando sendo possível avaliar outros cenários com maior potencial de proximidade ao sumário estatístico observado.

Quando nos referimos aos biomas, é interessante ressaltar que os indivíduos de *L. angustirostris* localizados geograficamente nos núcleos das Matas Secas (Seasonal Dry Tropical Forests – SDTFs) em Misiones (LOC 29 e LOC 30), Piedmont (LOC 24 e LOC 26) e Chiquitanos (LOC 25) pertencem à população do Chaco e não do núcleo da Caatinga (Figura 2). Além disso, a localidade de Bonito-MS (LOC 27) assim como nas análises com DNA mitocondrial (Capítulo II) pertence também a população do Chaco. Com isso, existe um suporte genético para a região do Gran Chaco proposto com base na análise da vegetação desta região (Prado, 1993). Assim como nossos dados, este mesmo padrão de indivíduos localizados geograficamente na região geográfica do Gran Chaco pertencem as populações genéticas do Chaco em outros táxons, como exemplo o gênero de insetos *Zoniopoda* (Pocco et al., 2018) e o gênero de lagartos *Vanzosaura* (Recoder et al., 2014).

A relação entre os biomas não parece estar relacionada à formação deles, e sim ao sentido de expansão de cada organismo, uma vez que as análises do DNA mitocondrial de *L. angustirostris* (Capítulo II) sugerem a Caatinga como população ancestral, sendo Chaco e Cerrado mais próximos entre si, e as análises demográficas deste trabalho sugerem que esta relação é apenas reflexo do sentido da expansão/colonização do Cerrado seguido do Chaco partindo da Caatinga. Já com lagartos (*Phyllopezus pollicaris*) onde a origem da espécie

aconteceu no Chaco, a expansão/colonização se deu no sentido Cerrado e Caatinga a relação de proximidade entre os biomas é diferente sendo Cerrado mais próximo da Caatinga (Werneck et al., 2012). Portanto, a discussão acerca da relação biogeográfica entre Caatinga, Cerrado e Chaco com base em padrões fitogeográficos (AbSaber, 1977; Andrade-Lima, 1982) e análises cladísticas de distribuição e endemismo (Colli, 2005; Porzecanski & Cracraft, 2005) podem não revelar a real relação biogeográfica entre eles.

### *Influência das flutuações climáticas*

As estimativas de tempo de separação das linhagens são todas coincidentes com as flutuações climáticas do Pleistoceno. Considerando os índices de “*marine isotope stages – MIS*” como indicadores de períodos glaciais e interglaciais (Kukla, 2005), a estimativa de separação da Caatinga (~370 mil anos atrás) coincide com um período glacial similar ao LGM (MIS 10). Já a separação das populações do Chaco e do Cerrado (~250 mil anos atrás) coincide com um período interglacial similar ao Holoceno (MIS 8). Este período de 250 mil anos atrás é coincidente com começo de mudanças na umidade da América do Sul associada aos níveis precipitação (Cheng et al., 2013). Considerando esse períodos de mudanças climáticas, a existência de diferenças climáticas entre os biomas (Figura 5 e Figura 6) e entre os nichos climáticos das populações (Figura 6), podemos propor que as flutuações climáticas do Pleistoceno foram uma força importante na diversificação de *L. angustirostris*. O elevado tamanho efetivo populacional da população ancestral de Cerrado+Chaco (Tabela 2) sugere um movimento de expansão do grupo partindo da Caatinga em direção ao Cerrado e Chaco em um momento de condições climáticas favoráveis a esta espécie, este padrão de expansão e aumento populacional em épocas climáticas favoráveis já foi elucidado em ambientes florestais, como exemplo a colonização de espécies de lagarto partindo da Amazônia em direção à Mata Atlântica (Prates et al., 2016). A subsequente diminuição do tamanho populacional no Cerrado e no Chaco podem ser explicados pela separação das populações como resposta diferentes às condições climáticas de cada bioma (Muller, 1973), além disso, a formação de corredores florestais no Cerrado ligando a Mata Atlântica aos Andes podem ter isolado as populações do Cerrado e Chaco, uma vez que esta conexão se estabeleceu no Pleistoceno (Cabanne et al *in review*). A influência das flutuações climáticas do Pleistoceno nos biomas de áreas abertas já vem sendo relatadas como importantes, em plantas o sentido de dispersão das espécies foi diferente em períodos glaciais e interglaciais (Collevatti et al., 2015; Vitorino et al., 2018), processos de bottleneck e expansão em decorrência de mudanças climáticas em aranhas

(Bartoleti et al., 2017) e sapos onde a expansão e diversificação deste grupo ocorreu no Pleistoceno em decorrência das mudanças climáticas (Prado et al., 2012).

É interessante ressaltar que há indícios de uma subestruturação dentro da população do Cerrado como parte das respostas a essas flutuações climáticas, considerando que há indícios da diminuição da vegetação do Cerrado no LGM (Ledru et al., 1996) e posterior expansão dessa vegetação em períodos mais quentes e úmidos como ocorreu no Holoceno (Ledru, 2002), podemos propor que as mudanças na dinâmica do Cerrado, possam ter ocorrido de forma cíclica assim como os períodos glaciais e interglaciais foram cíclicos (Kukla, 2005), explicando esta subestruturação. As evidências da análise de estrutura populacional (Figura 3) sugerem que esta subestruturação esteja organizada em leste e oeste do Cerrado, já que avaliando a possibilidade de cinco populações ( $k=5$ ) é possível identificar dois grupos, um mais a oeste da distribuição (LOC 19, LOC 20 e LOC 21) e um mais a leste da distribuição (LOC 18 e LOC 19). Este padrão de separação do leste e oeste do Cerrado já foi observado para outros grupos como exemplo, plantas (Collevatti et al., 2015; Vitorino et al., 2018), lagartos (Recoder et al., 2014), felinos (Santos et al., 2018) e sapos (Prado et al., 2012) sugerindo que deva existir algum evento biogeográfico atuando na diversificação dentro do bioma. Uma das hipóteses que pode explicar este processo podem ser diferenças climáticas citada acima, como também outros eventos climáticos que não apenas retração e expansão de vegetação. Tendo como pressuposto o exemplo da Amazônia, onde é sabido que a parte mais oeste permaneceu estável enquanto a parte mais leste tornou-se significativamente mais seca nos períodos glaciais (Cheng et al., 2013) podemos propor que uma porção do Cerrado possa ter sido mais estável climaticamente que a outra, ou ainda outra hipótese não excludente às anteriores, de que em virtude de corredores florestais que propiciaram a conexão entre a Mata Atlântica e a Amazônia (Ledo & Colli, 2017) e a Mata Atlântica e os Andes (Trujillo-Arias et al., 2017; Cabanne et al. *in review*) possam ter isolado o Cerrado em leste e Oeste. Nós reconhecemos que a população homogênea da Caatinga pode ser um viés da nossa amostragem com pouca representatividade do leste da Caatinga, podendo, portanto, ter uma subestruturação negligenciada, já que tais corredores também são presentes na Caatinga (Thomé et al., 2016; Ledo & Colli, 2017).

#### *Isolamento por distância*

Além disso, estes corredores no Cerrado (Ledo & Colli, 2017; Trujillo-Arias et al., 2017) e na Caatinga (Thomé et al., 2016; Ledo & Colli, 2017) podem explicar a ausência de evidências para o isolamento por distância nas populações de *L. angustirostris* destes biomas (Figura 8b e Figura 8c), uma vez que em virtude da fragmentação causada pelo aparecimento

de áreas florestais, ocorrem recorrentes re-colonização dessas áreas quando as condições climáticas favorecem áreas mais secas, este processo faz com que a população não atinja o tempo necessário para o equilíbrio entre fluxo gênico e deriva genética (Castric & Bernatchez, 2003; Pinsky et al., 2010), justificando a ausência de isolamento por distância nesses biomas. Por outro lado, o isolamento por distância encontrado no Chaco (Figura 8a) pode ser explicado pela ausência de evidências de corredores nesta área (Cabanne et al, *in review*) o que sugere uma área não fragmentada ao longo das flutuações climáticas, diferente do que era esperado em relação à existência de corredores florestais ao longo dos rios Bermejo e Pilcomayo (Nores, 1992). A ausência de evidência de corredores florestais, a presença de isolamento por distância na população do Chaco de *L. angustirostris* e a modelagem paleoclimática desta mesma espécie (Capítulo II) sugerem uma estabilidade climática desta região.

#### *Implicações taxonômicas*

Ainda que a avaliação dos padrões de plumagem (Bolívar-Leguizamón & Silveira, 2015) e a avaliação genética neutra (Capítulo II) já abordaram as questões taxonômicas, estes estudos não recuperado unidades completamente diagnosticáveis que pudessem ser associadas às subespécies. Neste trabalho, utilizando um número maior de marcadores, os resultados sugerem uma nova proposta taxonômica para esta espécie. Com base nas populações genéticas (Figura 3), o fluxo gênico negligenciável (Tabela 2) somados aos nichos climáticos diferentes (Figura 8), propomos a existência de três subespécies para *Lepidocolaptes angustirostris* sendo elas: (I) a subespécie do Chaco, *L. angustirostris angustirostris*; (II) a subespécie do Cerrado, *L. angustirostris bivittatus* e a subespécie da Caatinga *L. angustirostris bahiae*. Sendo assim sinonimizar-se-iam as atuais oito subespécies em apenas três subespécies sendo elas: *L. a. certhiolus* (1913), *L. a. praedatus* (1916) e *L. a. hellmayri* (1925) seriam sinônimas a *L. angustirostris angustirostris* (1818), *L. a. griseiceps* (1974) sinônima a *L. angustirostris bivittatus* (1822), e *L. a. coronatus* (1930) sinônima a *L. angustirostris bahiae* (1903).

## CONCLUSÕES

A independência dos biomas sustentada por populações independentes sem fluxo gênico e por nichos climáticos independentes é uma forte força na diversificação da avifauna nas áreas abertas da América do Sul. A ausência de barreiras geográficas evidentes entre os biomas sugere que a restrição de fluxo gênico possa ser explicada por diferentes forças seletivas em virtude da independência de nicho de cada bioma. Além disso, pode-se reforçar a influência das mudanças climáticas nos processos de diversificação, neste caso, associar as mudanças

climáticas a alteração da paisagem com o aparecimento de áreas florestais nos biomas abertos levando a diversificação dentro dos biomas, tal influência é apoiada pela alta estruturação no Cerrado, bioma com maiores registros de conexões florestais em períodos glaciais e interglaciais (Ledo & Colli, 2017), e a presença de isolamento por distância, sugerindo um tempo de equilíbrio entre deriva genética e fluxo gênico (Castric & Bernatchez, 2003) no Chaco, que não apresenta evidência de rotas de conexões florestais. Nesse sentido, as condições climáticas distintas e respostas às mudanças climáticas independentes (Muller, 1973) nos biomas associadas ao aparecimento de áreas florestais nos biomas parecem ser as maiores forças atuando na diversificação das áreas abertas da América do Su

## REFERÊNCIAS

- AbSaber, A.N. (1977) Os domínios morfoclimáticos da América do Sul. Primeira aproximação. *Geomorfologia*, **53**, 1–23.
- Akaike, H. (1974) “A new look at the statistical model identification.” *IEEE Transactions on Automatic Control*, **19**, 716–723.
- Andrade-Lima, D. (1982) Present-day forest refuges in northeastern Brazil. *Biological Diversification in the Tropics*. (ed. by G.T. Prance), pp. 245–251. Columbia University Press, New York.
- Bartoletti, L.F. de M., Peres, E.A., Sobral-Souza, T., Fontes, F. von H.M., Silva, M.J. da, & Solferini, V.N. (2017) Phylogeography of the dry vegetation endemic species *Nephila sexpunctata* (Araneae: Araneidae) suggests recent expansion of the Neotropical Dry Diagonal. *Journal of Biogeography*, 1–14.
- Batalha-Filho, H., Fjeldså, J., Fabre, P.H., & Miyaki, C.Y. (2013) Verbindungen zwischen atlantischen und amazonischen Waldvogelfaunen spiegeln distinkte historische Ereignisse wider. *Journal of Ornithology*, **154**, 41–50.
- Bolívar-Leguizamón, S. & Silveira, L.S. (2015) Morphological variation and taxonomy of *Lepidocolaptes angustirostris* (Vieillot, 1818) (Passeriformes: Dendrocolaptidae). *Papéis avulsos de Zoologia*, **55**, 281–316.
- Busby, J.R. (1991) BIOCLIM A Bioclimate Analysis and Prediction System. *Plant Protection Quarterly*, **6**, 8–9.
- Castric, V. & Bernatchez, L. (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchell). *Genetics*, **163**, 983–996.
- Cheng, H., Sinha, A., Cruz, F.W., Wang, X., Edwards, R.L., D’Horta, F.M., Ribas, C.C., Vuille, M., Stott, L.D., & Auler, A.S. (2013) Climate change patterns in Amazonia and

- biodiversity. *Nature Communications*, **4**, .
- Collevatti, R.G., Terribile, L.C., Rabelo, S.G., & Lima-Ribeiro, M.S. (2015) Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. *Frontiers in Plant Science*, **6**, 1–15.
- Colli, G.R. (2005) As origens e a diversificação da herpetofauna do Cerrado. *Cerrado: Ecologia, Biodiversidade e Conservação*. (ed. by A. Scariot, J.C. Sousa-Silva, and J.M. Felfili), pp. 249–264. Ministério do Meio Ambiente, Brasília, Distrito Federal.
- Evanno, G., Regnaut, S., & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Faircloth, B.C. (2015) PHYLUCES is a software package for the analysis of conserved genomic loci. *Bioinformatics*, **32**, 786–788.
- Frankham, R., Ballou, J.D., & Briscoe, D.A. (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.
- Furley, P.A. & Metcalfe, S.E. (2007) Dynamic changes in savanna and seasonally dry vegetation through time. *Progress in Physical Geography*, **31**, 633–642.
- Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., & Siepel, A. (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, **43**, 1031–1035.
- Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics, 4th edition*. Cornell University, Massachusetts.
- Hayes, F.E. (2001) Geographic variation, hybridization, and the leapfrog pattern of evolution in the Suiriri flycatcher (*Suiriri suiriri*) complex. *The Auk*, **118**, 457–471.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., & Andrew, R.L. (2017) The K = 2 conundrum. *Molecular Ecology*, 3594–3602.
- Katoh, K. & Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, 772–780.
- Kukla, G. (2005) Saalian supercycle, Mindel/Riss interglacial and Milankovitch's dating. *Quaternary Science Reviews*, **24**, 1573–1583.

- Ledo, R.M.D. & Colli, G.R. (2017) The historical connections between the Amazon and the Atlantic Forest revisited. *Journal of Biogeography*, .
- Ledru, M.-P., Braga, P.I.S., Soubies, F., Fournier, M., Martin, L., Suguio, K., & Turcq, B. (1996) The last 50,000 year in the Neotropics (Southern Brazil): evolution of vegetation and climate. *Palaeogeography, Palaeoclimatology, Palaeocology*, **123**, 239–257.
- Ledru, M. (2002) Late Quaternary history and evolution of the cerrados as revealed by palynological records. *Radiocarbon**The cerrados of Brazil*, 33–50.
- Lemmon, A.R., Emme, S., & Lemmon, E.C. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.
- Marantz, C., Aleixo, A., Bevier, L., & Patten, M. (2003) Available at: <http://www.hbw.com/species/narrow-billed-woodcreeper-lepidocolaptes-angustirostris>.
- Meyer, M. & Kircher, M. (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb.prot5448-prot5448.
- Muller, P. (1973) The dispersal centers of terrestrial vertebrates in the Neotropical realm: a study in the evolution of the Neotropical biota and its native landscapes. *Biogeographica*, **2**, 1–244.
- Nix, H.A. (1986) A Biogeographic Analysis of Australian Elapid Snakes. *Atlas of Elapid Snakes of Australia*. (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Nores, M. (1992) Bird Speciation in Subtropical South America in Relation to Forest Expansion and Retraction. *The Auk*, **109**, 346–357.
- Nylander, J.A.A. (2004) *MrAIC.pl* program distributed by the author. Evolutionary Biology centre, Evolutionary Biology centre.
- Pennington, R.T., Prado, D.E., & Pedry, C.A. (2000) Neotropical seasonally dry forests and Quaternaryvegetation changes. *Journal of Biogeography*, **27**, 261–273.
- Pinsky, M., Humberto, L., Montes Jr, R., & Palumbi, S.R. (2010) Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution*, **64**, 2688–2700.
- Pocco, M.E., Guzmán, N., Plischuk, S., Confalonieri, V., Lange, C.E., & Cigliano, M.M. (2018) Diversification patterns of the grasshopper genus *Zoniopoda* Stål (Romaleidae, Acridoidea, Orthoptera) in open vegetation biomes of South America. *Systematic Entomology*, **43**, 290–307.
- Porzecanski, A.L. & Cracraft, J. (2005) Cladistic analysis of distribution and endemism



- (CADE): using raw distribution of birds to unravel the biogeography of the South American aridlands. *Journal of Biogeography*, **32**, 261–275.
- Prado, C.P.A., Haddad, C.F.B., & Zamudio, K.R. (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology*, **21**, 921–941.
- Prado, D. (1993) What is the Gran Chaco vegetation in South America? I: A review. Contribution to the study of flora and vegetaion of the Chaco. V. *Candollea*, **48**, 145–172.
- Prates, I., Xue, A.T., Brown, J.L., Alvarado-Serrano, D.F., Rodrigues, M.T., Hickerson, M.J., & Carnaval, A.C. (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences*, **113**, 7978–7985.
- Pritchard, J.K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rambaut, A., Suchard, M.A., Xie, D., & Drummond, A.J. (2014) Available at: <http://beast.bio.ed.ac.uk/Tracer>.
- Recoder, R.S., De Pinho Werneck, F., Teixeira, M., Colli, G.R., Sites, J.W., & Rodrigues, M.T. (2014) Geographic variation and systematic review of the lizard genus *Vanzosaura* (Squamata, Gymnophthalmidae), with the description of a new species. *Zoological Journal of the Linnean Society*, **171**, 206–225.
- Rokyta, D.R., Lemmon, A.R., Margres, M.J., & Aronow, K. (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*, **13**, 312.
- Santos, A. da S., Trigo, T.C., Oliveira, T.G. de, Silveira, L., & Eizirik, E. (2018) Phylogeographic analyses of the pampas cat (*Leopardus colocola*; Carnivora, Felidae) reveal a complex demographic history. *Genetics and Molecular Biology*, **41**, 273–287.
- Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., & Brumfield, R.T. (2014) Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology*, **63**, 83–95.
- Stephens, M. & Scheet, P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American journal of human genetics*, **76**, 449–462.
- Stephens, M., Smith, N.J., & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978.

- Thomé, M.T.C., Sequeira, F., Brusquetti, F., Carstens, B., Haddad, C.F.B., Rodrigues, M.T., & Alexandrino, J. (2016) Recurrent connections between Amazon and Atlantic forests shaped diversity in Caatinga four-eyed frogs. *Journal of Biogeography*, **43**, 1045–1056.
- Trujillo-Arias, N., Dantas, G.P.M., Arbeláez-Cortés, E., Naoki, K., Gómez, M.I., Santos, F.R., Miyaki, C.Y., Aleixo, A., Tubaro, P.L., & Cabanne, G.S. (2017) The niche and phylogeography of a passerine reveal the history of biological diversification between the Andean and the Atlantic forests. *Molecular Phylogenetics and Evolution*, **112**, 107–121.
- Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F., & Palma-Silva, C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.
- Vanzolini, P.E. (1963) Problemas faunísticos do Cerrado. *Ferri, M. (editor) Simpósio sobre o Cerrado. Editora da Universidade de São Paulo. São Paulo*
- Vitorino, L.C., Lima-Ribeiro, M.S., Terribile, L.C., & Collevatti, R.G. (2018) Demographical expansion of *Handroanthus ochraceus* in the Cerrado during the Quaternary: Implications for the genetic diversity of Neotropical trees. *Biological Journal of the Linnean Society*, **123**, 561–577.
- Warren, D.L., Glor, R.E., & Turelli, M. (2008) Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.
- Werneck, F.P. (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.
- Werneck, F.P., Gamble, T., Colli, G.R., Rodrigues, M.T., & Sites, J.W. (2012) Deep diversification and long-term persistence in the south american “dry diagonal”: Integrating continent-wide phylogeography and distribution modeling of geckos. *Evolution*, **66**, 3014–3034.
- Wright, S. (1943) Isolation by Distance. *Genetics*, **28**, 114–138.
- Zanella, F.C. (2002) Systematics and biogeography of the bee genus *Caenonomada* Asmead, 1899 (Hymenoptera: apidae: Tapinotaspidini). *Studies on Neotropical Fauna and Environment*, **37**, 249–261.

## CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

As principais conclusões desta tese são a independência dos biomas e a influência do clima na diversificação. Tanto a análise com DNA mitocondrial como a análise com NGS indicam populações associadas aos biomas, ainda que a estimativa de migração não possa ter sido feita pelo DNA mitocondrial, as estimativas com NGS mostram atual ausência de fluxo gênico e mais ainda, os nichos independentes sugerem a independência dos biomas e sugerem a influência do clima na diversificação destas populações. Além disso, as estimativas de tempo de ambas as abordagens são concordantes no Pleistoceno, e por mais que as áreas estáveis não parecem ter influência na diversificação as mudanças climáticas, principalmente as associadas a formação de corredores florestais podem ter atuado de forma efetiva na diversificação do grupo. Neste sentido, considerando a independência dos biomas e a influência de tais corredores na diversificação do Cerrado em pelo menos leste e oeste associado aos registros da literatura, temos a intenção de testar com diferentes espécies de aves se essa estrutura populacional é recorrente em aves e identificar a causa de tal diferenciação se é que ela é um padrão.

Alguns motivos ressaltam a importância das conclusões desta tese, como: (I) ser o primeiro trabalho com uma espécie de ave que tenha distribuição coincidente com a diagonal de áreas abertas; (II) conta com uma amostragem que abrange a distribuição em quase toda sua totalidade e ser uma abordagem que inclui tanto uma abordagem mais tradicional (DNA mitocondrial), quanto uma abordagem mais recente (NGS); (III) os resultados e a discussão trazem contribuições acerca dos processos envolvidos na diversificação de *Lepidocolaptes angustirostris* (IV) é possível testar se esses processos são recorrentes nas demais espécies possibilitando a descrição dos padrões de diversificação nas áreas abertas da América do Sul.

PRODUÇÃO  
RELACIONADA



# Mitogenomes of two neotropical bird species and the multiple independent origin of mitochondrial gene orders in Passeriformes

Renato Caparroz<sup>1</sup> · Amanda V. Rocha<sup>1</sup> · Gustavo S. Cabanne<sup>2</sup> · Pablo Tubaro<sup>2</sup> · Alexandre Aleixo<sup>3</sup> · Emily M. Lemmon<sup>4</sup> · Alan R. Lemmon<sup>5</sup>

Received: 21 February 2017 / Accepted: 6 February 2018  
© Springer Science+Business Media B.V., part of Springer Nature 2018

## Abstract

At least four mitogenome arrangements occur in Passeriformes and differences among them are derived from an initial tandem duplication involving a segment containing the control region (CR), followed by loss or reduction of some parts of this segment. However, it is still unclear how often duplication events have occurred in this bird order. In this study, the mitogenomes from two species of Neotropical passerines (*Sicalis olivascens* and *Lepidocolaptes angustirostris*) with different gene arrangements were first determined. We also estimated how often duplication events occurred in Passeriformes and if the two CR copies demonstrate a pattern of concerted evolution in Sylvioidea. One tissue sample for each species was used to obtain the mitogenomes as a byproduct using next generation sequencing. The evolutionary history of mitogenome rearrangements was reconstructed mapping these characters onto a mitogenome Bayesian phylogenetic tree of Passeriformes. Finally, we performed a Bayesian analysis for both CRs from some Sylvioidea species in order to evaluate the evolutionary process involving these two copies. Both mitogenomes described comprise 2 rRNAs, 22 tRNAs, 13 protein-codon genes and the CR. However, *S. olivascens* has 16,768 bp showing the ancestral avian arrangement, while *L. angustirostris* has 16,973 bp and the remnant CR2 arrangement. Both species showed the expected gene order compared to their closest relatives. The ancestral state reconstruction suggesting at least six independent duplication events followed by partial deletions or loss of one copy in some lineages. Our results also provide evidence that both CRs in some Sylvioidea species seem to be maintained in an apparently functional state, perhaps by concerted evolution, and that this mechanism may be important for the evolution of the bird mitogenome.

**Keywords** *Lepidocolaptes angustirostris* · *Sicalis olivascens* · Mitochondrial genome · Anchored Phylogenomics

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11033-018-4160-5>) contains supplementary material, which is available to authorized users.

✉ Renato Caparroz  
renatocz@yahoo.com.br

<sup>1</sup> Departamento de Genética e Morfologia, Laboratório de Genética e Biodiversidade, Instituto de Ciências Biológicas, Universidade de Brasília, Campus Universitário Darcy Ribeiro, Asa Norte, Brasília, Distrito Federal CEP 70910-900, Brazil

<sup>2</sup> División de Ornitología, Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Ciudad de Buenos Aires, Argentina

## Introduction

When Desjardins and Moraes [1] described the first complete mitochondrial genome of a bird, the domestic chicken (*Gallus gallus*), they showed the existence of a new mitogenome arrangement among the chordates. Since

<sup>3</sup> Coordenação de Zoologia, Museu Paraense Emílio Goeldi, Belém, Pará, Brazil

<sup>4</sup> Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295, Tallahassee, FL 32306-4295, USA

<sup>5</sup> Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4102, USA

then, hundreds of avian mitogenomes have been described and seven new arrangements have been discovered (see [2]). Differences among these arrangements are related to the number of copies and arrangements comprising the control region (CR), likely a result of CR and flanking region tandem duplications followed by partially random deletions or loss of one copy [3–5]. Although the mitochondrial DNA appears to be under selection for compactness, leading to loss of function and elimination of additional gene copies [6], several studies have showed some evidence for the maintenance of function in both CRs, demonstrating a pattern of concerted evolution (e.g. [7]). In these cases, stabilizing selection or gene conversions have been evoked to explain the high degree of similarity between the two CRs (e.g. [8]).

Based on evolutionary process, Gibb et al. [4] proposed the following nomenclature for avian mitochondrial gene orders: ancestral or standard avian [1], duplicate tRNA<sup>Thr</sup>-CR (d\_tThr-CR) [9], duplicate CR (d\_CR) [7], and remnant CR2 (r\_CR2) [3]. In the ancestral avian gene order, the gene arrangement comprises cytochrome b (Cytb)/tRNA<sup>Thr</sup>/tRNA<sup>Pro</sup>/ND6/tRNA<sup>Glu</sup>/CR/tRNA<sup>Phe</sup>. The segment from the end of Cytb to CR is tandem duplicated in d\_tThr-CR. In d\_CR and r\_CR2, the first segment from tRNA<sup>Pro</sup> to tRNA<sup>Glu</sup> and the second segment from the end of Cytb to tRNA<sup>Thr</sup> are reduced or deleted. Additionally, in d\_CR, both CR (CR1 and CR2) are complete and putatively functional, while in r\_CR2, duplicated CR2 is degenerated. The last avian gene order described (hereafter tPro-CR) differs from the d\_tThr-CR as tRNA<sup>Thr</sup> is not part of the duplicated segment or it is degenerated after duplication [10].

Although the standard avian gene arrangement has been found in many species from several orders, the other derived gene arrangements have been only found in some species/genera of few orders. For instance, the d\_tRNA<sup>Thr</sup>-CR have found in Procellariiformes [4, 9] and Bucerotiformes [11], d\_CR in Psittaciformes [7] and Piciformes [4], r\_CR2 in Tinamiformes and Cuculiformes [3, 12, 13], and dd\_tRNA<sup>Thr</sup> in Charadriiformes [14]. Other two gene order were exclusively described for some Ardeidae species [2]. However, according to Schirtzinger et al. [8], the study of the evolutionary dynamics of avian mitogenome organization remains an ongoing challenge for the field of molecular evolution. Furthermore, only few orders were investigated in relation to evolutionary history of mitogenome rearrangements based on mapping these characters onto phylogenies (e.g. [7, 8]).

Particularly in Passeriformes, > 200 mitogenomes have been described, representing 47 of the 100 families of this order. The increase in the number of mitogenomes described results from the advent of next generation sequencing and due to the application of this methodology in phylogenetic studies on birds. As part of searching for a greater number of loci to increase resolution of the evolutionary relationships

in birds, many authors have described bird mitogenomes as a byproduct of these studies (e.g. [15]).

At least four different mitogenome arrangements occur in Passeriformes. The ancestral avian arrangement has been found in the basal lineage of Passeriformes (*Acanthisitta chloris*) [16] and in several supraorders, such as Meliphagoidea, Corvoidea and Passeroidea (e.g. [17]). The other three arrangements showed a more restricted taxonomic distribution and have been only found in all suboscine (d\_CR and r\_CR2), in some Sylvioidea species (d\_CR and r\_CR2), and in *Notiomystis cincta* and *Turdus philomelos* (tPro-CR) (see [10]). While the presence of similar rearrangements in suboscines and more derived lineages of oscines (e.g. Sylvioidea) indicates independent origin of these rearrangements, it is still unclear how often mitochondrial duplication events have occurred more generally in Passeriformes and in particular within Sylvioidea.

In this paper, we describe the complete mitochondrial genome of two species of Neotropical passerines with different gene arrangements: the greenish yellow finch (*Sicalis olivascens*, Thraupidae), which showed the ancestral avian arrangement, and the narrow-billed woodcreeper (*Lepidocolaptes angustirostris*, Dendrocolaptidae), which showed the r\_CR2 arrangement. In addition, we investigate the distribution of these arrangements on a mitogenome Bayesian phylogenetic tree of Passeriformes to estimate how often duplication events have occurred in this order and if the two CR copies show a pattern of concerted evolution in some Sylvioidea species.

## Materials and methods

### DNA extraction, library preparation, enrichment, and sequencing

Total DNA was extracted from vouchered tissue samples using DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer's protocol. One tissue sample for each species was used in this work: *S. olivascens* sample (MACN1307, Catamarca, Argentina) deposited in the Museo Argentino de Ciencias Naturales and *L. angustirostris* sample (MPEG 68210; São João dos Patos, Maranhão, Brazil) deposited in the Museu Paraense Emílio Goeldi. Mitogenome sequences were obtained as a by-product of the anchored hybrid enrichment approach described by Lemmon et al. [18] and Prum et al. [19]. Illumina indexed genomic libraries were prepared using a protocol modified (see details in [18]) from Meyer and Kircher [20] in a Beckman–Coulter Biomek® FXP Liquid-handling robot. A single pool containing 16 libraries (more specimens of *L. angustirostris* were included for phylogeographic analysis) was then enriched for the target using an Agilent Custom SureSelect kit (Agilent Technologies)

containing a single pool of all probes described in Lemmon et al. [18]. All procedures were performed by the Center for Anchored Phylogenomics (<http://www.anchoredphylogeny.com>). The enriched library pool was then 250 bp paired-end Illumina HiSeq 2000 sequenced on one PE150 Illumina HiSeq2000 lane by the Translational Science Laboratory in the College of Medicine at Florida State University.

## Mitogenome assembly and annotation

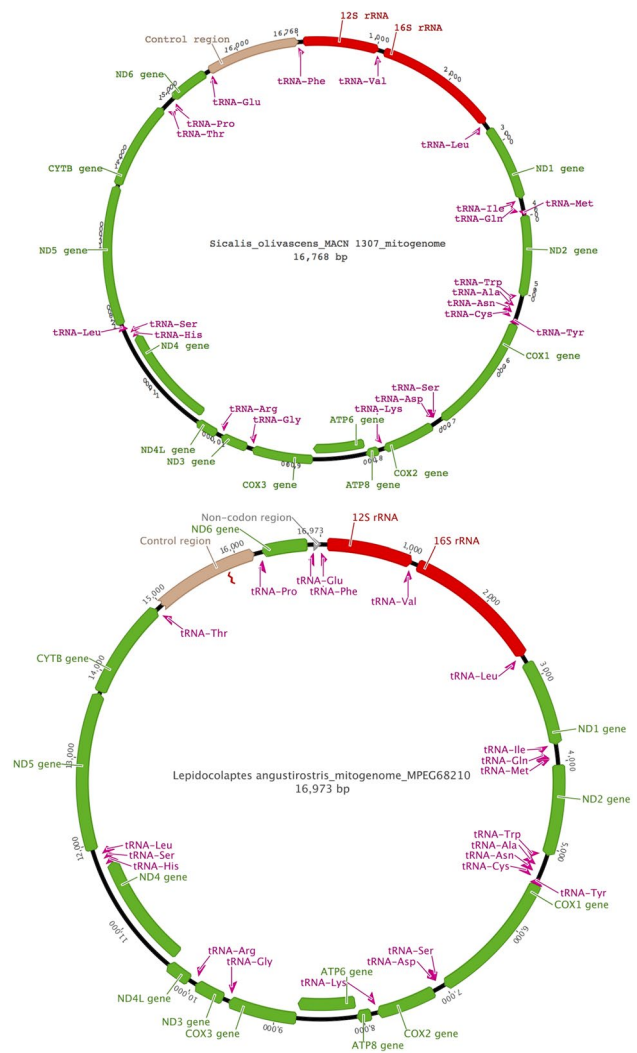
Paired reads were merged following Rokyta et al. [21]. Reads failing to meet the probability criterion were kept separate but still used in the assembly. Merged and unmerged reads were aligned to the reference mitogenomes using GENEIOUS 4.7 [22]; for *S. olivascens* and *L. angustirostris* was used the *Thraupis episcopus* (GenBank: KM078765) and *Mionectes oleagineus* (GenBank: NC024682) mitogenome as reference, respectively. Complete mitochondrial genomes were annotated based on the automatic methods of DOGMA [23] using default parameters and manually adjusted based on comparisons to the respectively reference mitogenomes using GENEIOUS.

## Alignment and bayesian phylogenetic inference

For the phylogenetic study we used an ingroup of 68 taxa, which included at least one representative of most families of Passeriformes for which mitogenomes have previously been described. We followed the taxonomic arrangement described in the Handbook Birds of the World [24]. With the exception of the two mitogenomes described here, the mitogenome sequences were obtained from GenBank. One Psittaciformes mitogenome (*Nestor notabilis*, GenBank: AY325307) was chosen as outgroup because it is the sister group of passerines based on previous molecular phylogenies (e.g. [19]). Sequences were aligned using MUSCLE v3.6 [25] with default parameters on GENEIOUS.

For Bayesian phylogenetic inference of mitogenomes, we used only the protein-coding genes. The one exception is the ND6 gene which was excluded because it was included in the region involved in the mitochondrial genome rearrangement analysis. The total alignment consisted of 10,826 bp obtained from 12 protein-codon genes. Three partitioning schemes were considered: all genes, partitioned by gene or partitioned by codon position. The optimal partitioning scheme (partitioned by codon position) was selected by PARTITIONFINDER v1.1.1 [26] under the Bayesian information criterion. The general time-reversible model was unlinked across partitions. Among-site rate variation was modeled using a discrete gamma distribution with four categories and a proportion of invariant sites. The approximation of posterior tree distributions was obtained by Markov Chain Monte Carlo (MCMC) using MRBAYES v.3.2 [27].

Multiple four chain Metropolis-coupled analyses (with default heating) were run 20,000,000 generations, sampling every 2000 generations. Run convergence and parameter posteriors were assessed using TRACER v1.5. The first 20% of the sampled trees were discarded as burn-in and the remaining trees were used to construct a 50% majority-rule consensus tree. The gene order arrangement was mapped on Bayesian phylogenetic tree using MESQUITE v.3.04 [28], considering four discrete categories: ancestral (0), duplicate CR (1), remnant CR2 (2) and tPro-CR (3). We used the parsimony criterion (Fitch) to reconstruct the ancestral state on the Bayesian phylogenetic tree. Finally, we also performed a Bayesian analysis for both CRs (CR1 and CR2) from some Sylvioidea species in order to evaluate the evolutionary



**Fig. 1** Gene maps of the *S. olivascens* (top) and *L. angustirostris* (bottom) mitochondrial genomes. COX1-3 indicates cytochrome c oxidase subunits 1–3; CYTB, cytochrome b; ATP6–8, ATPase subunits 6 and 8; ND1–6/4L, NADH dehydrogenase subunits 1–6/4L. Transfer RNA genes are designated by single-letter amino acid codes (Tables S1 and S2)

process involving the two CRs copies. The latter analysis was performed using multiple four chain Metropolis-coupled analyses (with default heating) with 20,000,000 generations, sampling every 2000 generations using MRBAYES.

## Results and discussion

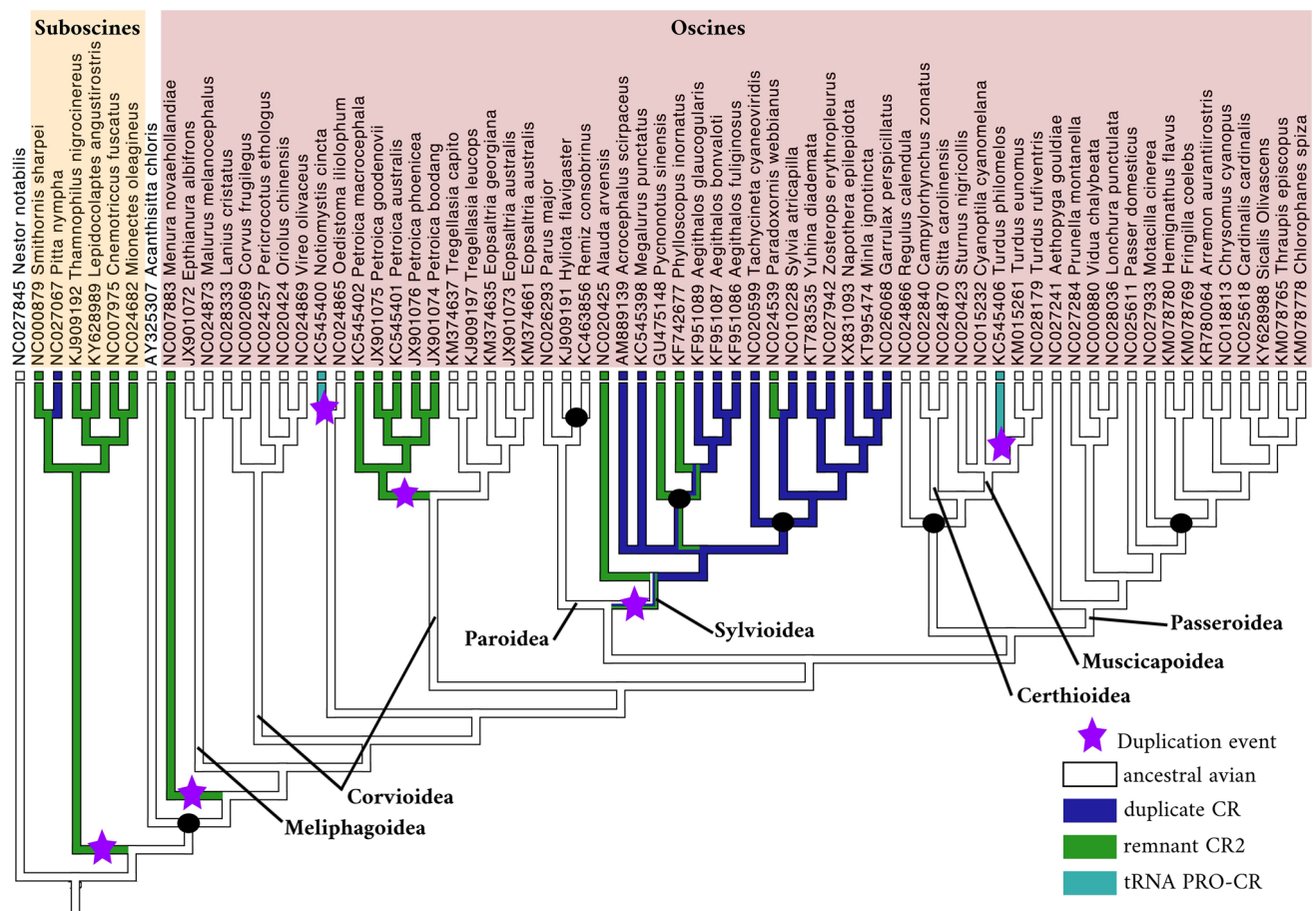
A total of 47,255 reads were mapped to *S. olivascens* (average coverage = 732 reads) while 63,814 reads were mapped to *L. angustirostris* (average coverage = 644 reads). Genbank accessions are KY628988 and KY628989, respectively.

Both mitogenomes described comprise 2 rRNAs, 22 tRNAs, 13 protein-codon genes and the CR (Fig. 1). However, *S. olivascens* showed the ancestral avian mitogenome order (T/P/ND6/E/CR), while *L. angustirostris* showed the r\_CR2 arrangement (Fig. 1), as expected for dendrocolaptids. In the latter, CR1 is found between tRNA<sup>Thr</sup> and

tRNA<sup>Pro</sup> and there is also an extra 76 bp non-coding region (CR2) flanked by tRNA<sup>Glu</sup> and tRNA<sup>Phe</sup>.

## Mitogenome phylogeny of Passeriformes

The bayesian reconstruction yielded a well-supported topology with high support (posterior probabilities > 0.98) in most nodes (Fig. 2). Evolutionary relationships across the majority of passerine taxa are consistent with those previously described (e.g. [17]), sustaining a basal split between Tyranni (suboscine) and Passeri (oscine). The Tyranni clade is composed of two main lineages corresponding to Old (*Smithornis* and *Pitta*) and New World (*Lepidocolaptes*, *Cnemotriccus* and *Mionectes*) suboscines. The Lyrebird *Menura* is the sister lineage of the other oscines, while Passeroidea (finch-like birds) is the most recently derived lineage. However, in contrast, Acanthisittidae was not recovery as the basal lineage of Passeriformes, but the clade support was low.



**Fig. 2** Ancestral state reconstruction of the avian mitochondrial gene arrangements onto the Bayesian phylogenetic tree of Passeriformes. Most nodes showed posterior probability distribution higher than 0.98 (values < 0.98 are indicate by black circles). Numbers before the spe-

cies names are the GenBank accessions. Bicolor or tricolor branches show that both gene arrangements are equally parsimonious as ancestral state



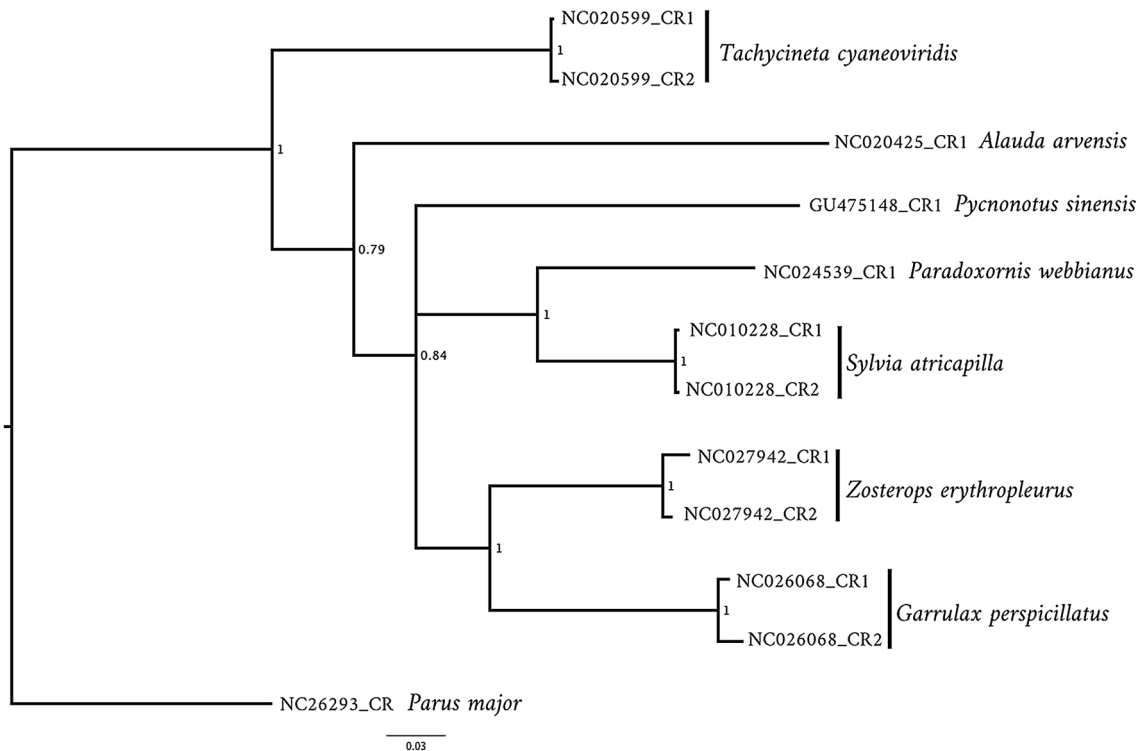
## Evolution of gene rearrangements in Passeriformes

Reconstruction of the ancestral state of gene rearrangements revealed that the standard avian order is the ancestral state in Passeriformes and the most frequent arrangement in this order (Fig. 2). Furthermore, considering the most commonly suggested model for gene rearrangement where one duplication event is followed by the reduction or loss of one copy (e.g. [4]), this analysis also showed at least six independent duplication events: (I) in the basal diversification of Suboscines; (II) in Menuridae lineage; (III) in Notiomystidae lineage; (IV) in Petroica lineage; (V) in *T. philomelos* lineage; and (VI) in the basal diversification of Sylvioidea (Fig. 2). After duplication event, reduction or loss of parts of one copy (yielding r\_CR2 arrangement) was observed to occur at least in four of these lineages, but likely initiating independently in different families of Sylvioidea. Although unlikely, there may have been a single duplication event in the basal diversification of Passeriformes followed by multiples independently reversals to standard avian order. According to Gibb et al. [4], gene duplications can be stable for long evolutionary periods due to continued gene conversion which eventually may revert to the original order.

In most Suboscines species studied and *M. novaehollandiae*, the CR2 was smaller than 301 bp in size (range from 76

to 301, Fig. S1) and this small segment aligned fragmented-way with different region of the suboscines' CRs (data not shown). This pattern is as expected for the derived arrangement r\_CR2, in which most parts of the segments involved in the duplication are deleted or gradually degenerated [3]; only a small portion of the ancestral CR remains. The only exception was observed in the CR2 of *Pitta nympha*, which is the longest segment (1136 bp) and did not align with any suboscines CR analyzed. In addition, the 3' segment is composed by around 25 bp tandemly repeated around 25 times.

In contrast, the CR2 observed in the derived arrangements of the Sylvioidea clade showed signs of drastic reduction only in four (Alaudidae, Pycnonotidae, Phylloscopidae and Timaliidae) of the eight families studied (Fig. 2 and Fig. S1). In the other species of this clade, CR2 were either a little longer, or similar in size, than the respective CR1. In all these latter cases, the duplicated CR (CR2) showed high similarity with their respective CRs (data not shown) and contained many of the conserved primary sequence motifs (e.g. F, D and C boxes, and CSB-1) that are typical of functional CRs in birds [29] and other vertebrates [30]. Furthermore, the CR2 of these Sylvioidea species are more closely related to paralogous segments (respective CRs) than the orthologous ones (Fig. 3). Thus, our data supports the hypothesis first raised by Singh et al. [5] that the duplication of the segment containing the CR occurred early on in



**Fig. 3** Bayesian analysis of mitochondrial CRs (CR1) and duplicated CRs (CR2) of some Sylvioidea species. Posterior probabilities are given in each node. Numbers before the names are the Genbank accessions

Sylvioidea diversification and that both CRs of some Sylvioidea species may be maintained functional over evolutionary time, either through stabilizing selection or through occasional gene conversion as previously described in other bird orders (e.g. [7]).

In conclusion, our results reinforce the idea of multiple origins of some mitochondrial gene arrangements in Passeriformes, with ancestral state reconstruction suggesting at least six independent duplication events. Our results also reinforce evidence that both CRs in some Sylvioidea species seem to be maintained in an apparently functional state, perhaps by concerted evolution, but the molecular mechanisms involved should be further investigated.

Finally, our conclusions were based on almost exclusively in published mitogenome, and in few cases in incomplete genomes, such as *A. chloris*, and this may directly affect our findings if different genetic orders are found in some of these lineages.

**Acknowledgements** We are grateful to all students and technicians from the Center for Anchored Phylogenomics at the Florida State University. We are also grateful to two anonymous reviewers for their valuable contribution to the improvement of this manuscript. This study is part of the GENPAC – Geographical Genetics and Regional Planning for Natural Resources in Brazilian Cerrado (MCT/CNPq/FNDCT/FAPs/MEC/CAPES/PRO-CENTRO-OESTE Nº 031/2010 – Proc. 564036/2010-2), and it was also supported by CNPq Nº 14/2014 (Proc. 445025/2014-0), PPBio – Rede Cerrado (CNPq Nº 35/2012 – Proc. 457444/2012-6), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina – PIP 2012–2014 0862), the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, Argentina – PICT 2014–2154). RC received PDE scholarship (CNPq – Proc. 202796/2014-0) and AVR received scholarship from CAPES and CNPq.

## Compliance with ethical standards

**Conflict of interest** There is no conflict of interest.

## References

- Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *J Mol Biol* 212:599–634
- Zhou X, Lin Q, Fang W, Chen X (2014) The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. *BMC Genom* 15:573
- Mindell DP, Sorenson MD, Dimcheff DE (1998) Multiple independent origins of mitochondrial gene order in birds. *Proc Natl Acad Sci USA* 95:10693–10697
- Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D (2007) Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. *Mol Biol Evol* 24:269–280
- Singh TR, Shneor O, Huchon D (2008) Bird mitochondrial gene order: insight from three warbler mitochondrial genomes. *Mol Biol Evol* 25:475–477
- Rand DM, Harrison RG (1986) Mitochondrial DNA transmission genetics in crickets. *Genetics* 114:955–970
- Eberhard JR, Wright TF, Bermingham E (2001) Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. *Mol Biol Evol* 18(7):1330–1342
- Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, Hernandez A, Mueller H, Graves GR, Fleischer RC, Wright TF (2012) Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Mol Phylogenet Evol* 64(2):342–356
- Abbott CL, Double MC, Trueman JW, Robinson A, Cockburn A (2005) An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* albatrosses. *Mol Ecol* 14(11):3605–3613
- Gibb GC, England R, Hartig G, McLenachan PA, Smith BLT, McComish BJ, Cooper A, Penny D (2015) New Zealand Passerines help clarify the diversification of major songbird lineages during the Oligocene. *Genome Biol Evol* 7(11):2983–2995
- Sammler S, Bleidorn C, Tiedemann R (2011) Full mitochondrial genomes of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genom* 12:35
- Bensch S, Härlid A (2000) Mitochondrial genomic rearrangements in songbirds. *Mol Biol Evol* 17:107–113
- Haddrath O, Baker AJ (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc R Soc Lond B* 268:939–945
- Verkuil YI, Piersma T, Baker AJ (2010) A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). *Mol Phylogenet Evol* 57(1):411–416
- Amaral FR, Neves LG, Resende MF Jr, Mobili F, Miyaki CY, Pellegrino KC, Biondo C (2015) Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS ONE* 10:e0138446
- Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, Penny D (2004) Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. *Mol Biol Evol* 21:974–983
- Barker FK (2014) Mitogenomic data resolve basal relationships among passeriform and passeridan birds. *Mol Phylogenet Evol* 79:313–324
- Lemmon AR, Emme SA, Lemmon EM (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol* 61:727–744
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR (2015) A comprehensive phylogeny of birds (Aves) using targeted next generation DNA sequencing. *Nature* 526:569–573
- Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. <https://doi.org/10.1101/pdb.prot5448>
- Rokyta DR, Lemmon AR, Margres MJ, Aronow K (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genom* 13:312
- Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2009) Geneious v4.7. <http://www.geneious.com>. Accessed 26 June 2015
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255
- del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds) (2017) Handbook of the birds of the world alive. Lynx Edicions, Barcelona. <http://www.hbw.com/>. Accessed 20 Nov 2016
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797

26. Lanfear R, Calcott B, Ho SY, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29(6):1695–1701
27. Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542
28. Maddison WP, Maddison DR (2015) Mesquite: a modular system for evolutionary analysis. Version 3.04. <http://mesquiteproject.org>. Accessed 18 Dec 2015
29. Marshall HD, Baker AJ (1997) Structural conservation and variation in the mitochondrial control region of fringilline finches (*Fringilla* spp.) and the greenfinch (*Carduelis chloris*). *Mol Biol Evol* 14:173–184
30. Southern S, Southern PJ, Dizon AE (1988) Molecular characterization of a cloned dolphin mitochondrial genome. *J Mol Evol* 28:32–42