Andressa Nuss

Filogeografia e padrões de diversificação de três espécies de aves da Mata Atlântica (*Myiothlypis leucoblephara, Myrmoderus loricatus* e *Myrmoderus squamosus*)

Phylogeography and patterns of diversification of three Atlantic Forest birds (*Myiothlypis leucoblephara, Myrmoderus loricatus and Myrmoderus squamosus*)

São Paulo, Setembro de 2019

Andressa Nuss

Filogeografia e padrões de diversificação de três espécies de aves da Mata Atlântica (*Myiothlypis leucoblephara*, *Myrmoderus loricatus* e *Myrmoderus squamosus*)

Phylogeography and patterns of diversification of three Atlantic Forest birds (*Myiothlypis leucoblephara*, *Myrmoderus loricatus and Myrmoderus squamosus*)

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutor em Ciências Biológicas, na Área de Biologia - Genética.

Orientadora: Cristina Yumi Miyaki

São Paulo, Setembro de 2019

Ficha Catalográfica

Nuss, Andressa

Filogeografia e padrões de diversificação de três espécies de aves da Mata Atlântica (*Myiothlypis leucoblephara, Myrmoderus loricatus* e *Myrmoderus squamosus)*

95 páginas

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo.

Departamento de Genética e Biologia Evolutiva.

1. Passeriformes 2.NGS 3.Região Neotropical 4.Demografia histórica

I.Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

_____	_____
Prof.(a) Dr.(a)	Prof.(a) Dr.(a)
_____	_____
Prof.(a) Dr.(a)	Prof.(a) Dr.(a) Cristina Yumi Miyaki

Dedico essa tese a todos os que lutam pela biodiversidade, tarefa tão árdua e complexa nessa pátria mal educada, comandada por covardes gananciosos.

Resistam como cerrado ao fogo, como mangue a maré, como vaga-lume a escuridão. Resistam!

“Nesses tempos de céus de cinzas e chumbos, nós precisamos de árvores desesperadamente verdes.”

Mário Quintana

Agradecimentos

A minha orientadora Cris, que aceitou o desafio de me supervisionar, por todo o conhecimento compartilhado, a estrutura fabulosa do laboratório e o suporte necessário para realizar um projeto com metodologia científica de ponta.

Nada como ser acolhida com corações quentes ao chegar na maior cidade da América Latina. Aos colegas de laboratório, com quem tive o prazer de conviver: Aline Macedo, Maria Svensson, Ana Bia Bueno, Bruno Lima, Beatriz Raicoski, Fernanda Midori; e em especial a quatro meninas incríveis que ganharam minha total admiração: Rafaella Monteiro, Carla Sardelli, Denise Mendes e Ana Cristina Fazza. Foram 4 anos de amizade, suporte e ótimos momentos de descontração.

Ao Claydson Assis e Gregory Thom que além da amizade e parceria, compartilharam comigo na prática e na teoria seus conhecimentos do extraordinário mundo ornitológico.

Ao Fábio Raposo e Henrique Batalha-Filho que estiveram sempre disponíveis para retirar dúvidas e auxiliar no projeto.

Aos que gentilmente cederam amostras; Adrian Azpiroz, Darío Lijtmaer (MNCN), Richard Prum e Marcos Raposo.

Aos funcionários do Instituto de Biociências, Lucilene, Patrícia, Helenice, Gilmar, Elder e especialmente ao Allan, por todo suporte necessário para a conclusão desse doutorado.

Aos membros da banca examinadora por gentilmente aceitarem participar da defesa de tese.

Agradecimento especial ao Guilherme, por ter me ajudado tanto e de tantas formas, sendo meu fiel companheiro faça chuva ou faça sol, mar adentro ou campo afora, mesmo longe sempre muito perto.

A todos os meus amigos, novos e antigos, fisicamente próximos ou não, expresso o meu profundo agradecimento por me acompanharem nessa jornada. Me sinto muito afortunada ao olhar para traz e ver todas essas múltiplas pegadas ao lado das minhas.

A minha família pelo amor e compreensão, tornando tudo isso possível.

Aos financiadores: CAPES pela implementação da bolsa de doutorado e ao Biota/FAPESP 2013/50297-0 por ter subsidiado esse projeto.

A São Paulo, selva desvairada de pedra e arte. Metrópole catalizadora de transformação, suas engrenagens a frente de outros tempos, proporcionaram meu crescimento pessoal exponencial. Nunca mais serei a mesma, ainda bem!

A Universidade de São Paulo pela estrutura fornecedora de experiências únicas. Movimentei meu corpo e minha alma no interior dessa renomada redoma. Serão lembradas com muito carinho as remadas no por do sol, pedaladas na garoa fina, canções que libertaram minha voz no coral e música erudita de excepcional qualidade gratuita disponível para quem quiser apreciar.

Sumário

Pág.

Resumo ----------i
Abstract ----------ii
Introdução geral ----------1
Capítulo 1. ----------18
Capítulo 2. ----------59
Discussão geral e conclusões ----------92

Resumo

A Floresta Atlântica possui grande heterogeneidade de ambientes e estima-se que hoje está reduzida a 11% a 16 % de sua área original. Além disso, abriga alto grau de endemismo, sendo considerada como um dos 5 *hotspots* mundiais de biodiversidade. Para investigar os processos históricos que podem ter sido responsáveis pela distribuição geográfica da diversidade de linhagens genéticas de espécies desse domínio realisamos na presente Tese um estudo filogeográfico de três espécies de passeriformes endêmicos da Floresta Atlântica, *Myiothlypis leucoblephara*, *Myrmoderus loricatus* e *squamosus.* Para acessar a estrutura genética e realizar análises demográficas das espécies foram produzidos dados de milhares de locos, por meio da técnica de *Genotyping by Sequencing*, de um total de 151 indivíduos de 83 localidades no Brasil, Uruguai e Argentina. A presente Tese inclui dois capítulos. No capítulo 1 a análise de ~8.700 SNPs de 86 indivíduos de *Myiothlypis leucoblephara* indicou a presença de dois grupos genéticos distribuidos em gradientes norte/sul e leste/oeste com claro padrão de isolamento por distância. Ainda, também detectamos expansão populacional no último máximo glacial (~20 mil anos atrás). No capítulo 2 realizamos um estudo filogeográfico das espécies irmãs *Myrmoderus loricatus* (35 indivíduos) e *M. squamosus* (30 indivíduos) utilizando em torno de ~15.500 SNPs. Diversas análises indicaram padrão de isolamento por distância para ambas espécies e a estimativa de data de separação dessas espécies poderia estar associada à alta atividade tectônica do Vale do Paraíba do Sul. Os dados apresentados na presente Tese permitiram testar algumas hipóteses de diversificação para a Floresta Atlântica e inferir como os processos demográficos atuam moldando a distribuição das linhagens. Além disso, nossos resultados reforçam a importância do uso de grandes quantidades de marcadores para caracterizar populações e inferir como os eventos demográficos aconteceram ao longo da evolução das espécies.

Abstract

The Atlantic Forest has a high heterogeneity of types of habitats and today it is reduced to 11% to 16% of its original area. In addition, it harbors an elevated degree of endemism and is considered as one of the top five global biodiversity hotspots. In order to investigate the historical processes that may be responsible for the geographic distribution of genetic lineages, we performed a phylogeographic study of three Atlantic Forest endemic passerine species, *Myiothlypis leucoblephara, Myrmoderus loricatus* and *M. squamosus.* Data from thousands of loci were produced using Genotyping by Sequencing to access the genetic structure and perform demographic analyzes of a total of 151 individuals from 83 locations in Brazil, Uruguay and Argentina. This Ph.D. Dissertation includes two chapters. In chapter 1 we analyzed ~8,700 SNPs from 86 individuals of *Myiothlypis leucoblephara*. The results indicated the presence of two genetic clusters distributed in north/south and east/west gradients with clear isolation by distance pattern. In addition, we also detected population expansion in Last Glacial Maximum (~20 thousand years ago). In chapter 2 we performed a phylogeographic study of the sister species *Myrmoderus loricatus* (35 individuals) and *M. squamosus* (30 individuals) based on ~15,500 SNPs. Different analyses indicated the presence of isolation by distance for both species and the estimated date of split of these species could be associated with high tectonic activity of Paraíba do Sul Valley. The data presented in this Ph.D. Dissertation allowed us to test some hypotheses of diversification for the Atlantic Forest and to infer how demographic processes act in shaping the distribution of lineages. Moreover, our results reinforce the importance of using large amounts of markers to characterize populations and infer how demographic events happened along the evolution of species.

Floresta Atlântica

A Floresta Atlântica (FA) é uma das maiores florestas úmidas das Américas, originalmente cobrindo cerca de 150 milhões de hectares distribuídos ao longo de uma faixa no leste da América do Sul. A FA foi uma formação florestal praticamente contínua ao longo de grande parte da região litorânea do Brasil, tendo como limite norte o estado do Ceará e o limite sul o estado do Rio Grande do Sul. Limitando sua distribuição a leste encontra-se a costa Atlântica e a oeste, parte do Estado do Mato Grosso do Sul e o leste do Paraguai, além de Misiones na Argentina (GALINDO-LEAL; CÂMARA, 2003).

A ação antrópica é responsável há séculos pela devastação de mais de 92% da mata nativa a partir do início da colonização europeia (DEAN, 1996). Nesse processo de ocupação, a FA experimentou desmatamento e exploração desordenada dos recursos naturais. Como resultado houve grave fragmentação do habitat e hoje existem apenas manchas disjuntas da floresta em locais que desfavorecem a instalação humana, principalmente em topografia mais acidentada que dificulta a mobilidade, impede atividade agrícola e acesso fácil ao escoamento de produção (PINTO *et al*. 2006). A maior parte da FA remanescente se encontra em pequenos fragmentos, menores que 100 hectares (RANTA *et al.* 1998) isolados uns dos outros e compostos por floresta de vegetação secundária em estágios médios de sucessão (METZGER, 2000; METZGER *et al.* 2009). Boa parte dos remanescentes de FA mais expressivos e extensos estão situados no estado de São Paulo, no Parque Estadual da Serra do Mar (LEITÃO FILHO 1994), o qual possui 332 mil hectares que abrangem parte de 24 municípios. Devido à sua vasta extensão e à sua localização, ele constitui um verdadeiro corredor ecológico, interligando os mais significativos remanescentes de FA do país (INSTITUTO FLORESTAL 2006).

Essa degradação do habitat pode, por exemplo, levar à extinção de populações pela perda de habitat ou pela perda de variabilidade genética devido ao isolamento e à perda de habitat. A resposta das espécies à perda de habitat pode variar amplamente e é um assunto complexo que exige estudos, e ainda não é possível mensurar os efeitos genético, ecológico e estrutural ao longo do tempo que a degradação do ambiente vai causar (METZGER *et al.* 2009). A presença de um mosaico de manchas de habitat (ao invés de uma paisagem contínua) pode refletir fortemente na dinâmica populacional de

migração, fluxo gênico e deriva genética que podem alterar a distribuição da composição alélica da espécie. Essa composição alélica é importante para que as populações se adaptem às transformações ambientais, assim, em muitos casos a diminuição da variabilidade genética (ex. por perda de alelos) leva ao declínio da população (AVISE 2000, PRIMACK; RODRIGUES, 2001).

Dentro desse contexto de perda de habitat, Myers (1988) cunhou o termo "hotspots" que se refere a áreas prioritárias para a conservação da biodiversidade por possuírem concentração excepcional de espécies com alto grau de endemismo além de terem experimentado degradação e perda de habitat. Para qualificar essas áreas, os dois parâmetros adotados pelo autor foram: conter no mínimo 0,5% de espécies de plantas globalmente endêmicas e que tenha perdido 70% ou mais de sua vegetação original. Assim, nesse estudo o autor classifica a costa Atlântica brasileira juntamente com Madagascar e o oeste do Equador como os "hotspots" mais importantes do planeta. A região costeira brasileira, onde a FA ocorre, é fortemente caracterizada pela industrialização em larga escala, com alto grau de urbanização, sendo responsável por gerar em torno de 70% do PIB nacional (CI-BRASIL *et al*. 2000). A alta densidade humana define diretamente a grande importância do litoral para o país uma vez que abriga mais da metade da população (FUNDAÇÃO SOS MATA ATLANTICA, 2008; CI-BRASIL *et al*. 2000). A FA originalmente possuía uma extensão de 1.227,600 km², restando no inicio do século XXI em torno de 11.4% a 16% da vegetação original (RIBEIRO *et al*. 2009; GALINDO LEAL; CÂMARA, 2003). Apesar de ter restado menos de 17% da cobertura vegetal, os dados são ainda mais preocupantes, pois apenas 35.9% dos remanescentes florestais estão sob proteção (MYERS *et al*. 2000) e mesmo nessas áreas a conservação das espécies não está garantida. Conforme Pinto *et al.* (2006), as unidades de conservação de proteção integral são 684 e abrangem menos de 2% da extensão do domínio. Infelizmente regiões biogeográficas distintas que ocorrem dentro da FA não estão igualmente representadas nessas áreas protegidas. Assim, esse sistema de unidades de conservação não abrange toda a biodiversidade desse domínio.

A biota da floresta é composta tanto por espécie mais antigas (Pré-Plioceno, entre 5 e milhão de anos atrás) quanto novas (Plioceno, entre 1 milhão 11,5 mil anos atrás; SILVA; CASTELETI, 2003). Analisando a composição da biodiversidade da FA, estima-se que esses complexos ecossistemas contenham 2,7% das plantas e 2,1% dos vertebrados endêmicos conhecidos mundialmente. Cerca de 620 espécies de aves habitam a FA sendo 181 endêmicas (STOTZ *et al*. 1996). Esses números somados aos

dos mamíferos, anfíbios e répteis resultam em cerca de 1.361 espécies, sendo 567 endêmicas (MYERS *et al*. 2000). De acordo Brooks *et al*. (2002), mais de 50% das plantas e 57% dos vertebrados globalmente ameaçados de extinção, segundo os critérios da União Mundial para a Natureza (IUCN), encontram-se nos "hostpots". Para os vertebrados terrestres da FA, cerca de 8,5% das espécies e aproximadamente uma em cada quatro das espécies endêmicas estão ameaçadas de extinção (PAGLIA 2005). Dado esse elevado grau de endemismo e também a forte fragmentação florestal, a FA abriga mais de 60% espécies presentes na lista oficial da fauna brasileira ameaçada de extinção (TABARELLI *et al*. 2003; PAGLIA 2005).

Esse alto índice de biodiversidade pode estar associado à heterogeneidade de habitats presentes na FA. Essa característica parece estar associada à grande amplitude de latitude e longitude (1º a 30º e 35º a 60º, respectivamente), estendendo-se por regiões tropicais e subtropicais, bem como variação de altitude (nível do mar até 2.900 m) e grande amplitude de temperatura (4ºC a 32ºC). Essa combinação de fatores proporciona uma grande variedade de macro e micro-habitats ao longo do domínio (GALINDO LEAL; CÂMARA, 2003; SILVA; CASTELETI, 2003). Hoje essa biodiversidade está contida em diversos tipos de formações florestais. As consideradas como *latu sensu* são: Floresta Ombrófila Densa e Mista (ou Floresta de Araucária); Floresta Estacional Semidecidual e Decidual; e ecossistemas associados, que são os campos sulinos e as áreas com influência flúvio-marinha, compreendidas pelos manguezais e restingas. A FA *stricto sensu* (ou Floresta Pluvial Atlântica) das regiões sudeste e sul, é composta por três formações florestais distintas: Florestas de Planície Litorânea, Florestas de Encostas e Florestas de Altitude (JOLY *et al*. 1999; RIZINNI 1997; OLIVEIRA FILHO; FONTES, 2000).

Diversas hipóteses têm sido levantadas visando explicar a origem da diversidade em florestas subtropicais e tropicais. Dentre elas está a hipótese dos museus que associa a maior concentração de endemismo (e maior taxa de diversificação) em áreas de elevadas altitudes (onde condições ambientais seriam mais resilientes às mudanças climáticas) do que em regiões de baixa altitude e com oscilações climáticas maiores (FJELDSAÅ *et al*. 1997; FJELDSAÅ; LOVETT, 1997). Outra hipótese proposta é a de que os rios podem ter atuado como barreiras para o fluxo gênico podendo resultar em especiação por isolamento (WALLACE 1852). De maneira similar, eventos tectônicos podem ter alterado a paisagem de forma a criar barreiras como, por exemplo, levando a especiação também por isolamento (SILVA; STRAUBE, 1996). A hipótese dos

gradientes de habitat ou de diversificação parapátrica por seleção postula que a seleção sobre caracteres relacionados à aptidão biológica pode, ao longo do tempo, resultar em isolamento reprodutivo entre grupos de indivíduos de ambientes distintos resultando no surgimento de diferentes espécies adaptadas a diferentes condições ambientais (SMITH *et al*. 1997). Outra hipótese é a dos refúgios florestais que se baseia em dados paleoclimáticos que indicam que no Pleistoceno o planeta passou por ciclos de avanço e recuo de geleiras continentais devido a mudanças globais de temperatura e umidade (SANT'ANNA NETO; NERY, 2005). Essa hipótese sugere que algumas florestas úmidas podem ter se retraído até tornarem-se remanescentes isolados (também chamados de paleorefúgios), circundados por vegetação aberta durante os períodos de máximo glacial (uma vez que nesses períodos o clima era seco e frio) e expandiam-se durante os períodos mais quentes e úmidos (períodos interglaciais). Como consequência, a distribuição de espécies dependentes de florestas úmidas possivelmente acompanhou os ciclos de expansão e retração das florestas. Se o tempo de isolamento dos organismos nesses remanescentes foi suficiente para promover diferenciação, mesmo com contato posterior devido à expansão florestal, essas linhagens poderiam permanecer sem fluxo gênico (HAFFER 1969; VANZOLINI; WILLIAMS, 1970).

O entendimento dos processos evolutivos que geraram a biodiversidade é importante para a sua conservação. Esse conhecimento permite identificar linhagens que podem ocorrer em áreas sob maior risco de impacto (MORITZ *et al.* 2000). Dentro desse panorama temos um importante desafio, uma vez que várias espécies de aves da FA estão na lista de espécies com maior risco de extinção dentre as aves neotropicais (STOTZ *et al*. 1996). Ainda, há a previsão de que diversas espécies desse domínio devem estar extintas em um curto período de tempo (BROOKS; BALMFORD, 1996). Estima-se que devido às mudanças climáticas 600 a 900 espécies de aves sejam extintas até o ano de 2100, a maioria (89%) em ambientes tropicais (ŞEKERCIOĞLU *et al*. 2012). Junto com a cobertura da vegetação original, parte do conhecimento que foi perdido é irrecuperável. Assim, diversos padrões e processos da história evolutiva envolvendo as comunidades estarão para sempre extintos sem que pudéssemos sequer ter a chance de compreendê-los.

Filogeografia

A filogeografia é um campo de estudo relacionado aos princípios e processos que regem as distribuições geográficas das linhagens genealógicas, especialmente

aquelas dentro e entre espécies estreitamente relacionadas. Como o termo indica, filogeografia inclui a distribuição espacial das linhagens genéticas (AVISE 2000), portanto ela visa reconstruir a filogenia baseada em dados genéticos de forma geograficamente contextualizada, permitindo testar hipóteses sobre a relação causal entre fenômenos geográficos, a distribuição das espécies, e os mecanismos que levam à diversificação (AVISE 2000). Dentro desse contexto, a filogeografia é uma disciplina que disponibiliza informações históricas para o entendimento de padrões de fluxo gênico, colonização, expansão e efeito da distribuição atual da variação genética (por exemplo adaptação local) ao longo de toda a distribuição de uma espécie (AVISE 2000). Uma vez que a filogeografia procura entender as forças microevolutivas que geraram esse padrão de diversidade intra-específica atual, uma das vertentes de investigação mais utilizadas é a identificação, se houver, de barreiras ao fluxo gênico dentro das espécies (AVISE *et al*. 1987). O surgimento dessa disciplina promoveu importantes avanços em diversas áreas, como por exemplo, biologia da conservação. A análise dos níveis de diversidade genética e a caracterização da distribuição geográfica desta variabilidade dentro de uma espécie são importantes para identificar e priorizar áreas nas quais programas de manejo e conservação devem ser elaborados (MORITZ; FAITH 1998). Dessa forma, as análises de padrões filogeográficos atuam de forma estreitamente direta na verificação de estruturação genética e por consequência na interpretação dos fatores que levaram ao surgimento da diversidade genética observada (MORITZ *et al.* 2000).

A abordagem filogeográfica foi originalmente desenvolvida utilizando marcadores mitocondriais (AVISE *et al*. 1987). Ou seja, com o mapeamento da distribuição de grupos de haplótipos de mtDNA de uma espécie ou complexo de espécies para inferir a história das populações (PUORTO *et al*. 2001). Com esses dados também se estima a variabilidade genética e se caracteriza a estrutura populacional (WILSON 1985). O DNA mitocondrial apresenta algumas características que facilitam sua utilização e o tornam não apenas uma ferramenta apropriada para estudos genéticos, como também uma importante ponte de ligação entre a genética de populações e a filogenia (AVISE *et al*. 1987). Dentre dessas características estão a herança materna, variabilidade intra-específica relativamente alta sendo sua taxa de evolução maior do que a de segmentos do DNA nuclear de cópia única e presença de múltiplas cópias por célula (AVISE 2000).

Com os avanços tecnológicos nos processos de sequenciamento e o crescente desenvolvimento de análises bioinformáticas mais complexas, o processamento dos dados moleculares ganhou nova dimensão com a possibilidade de realizar estudos sub-genômicos. Segundo Hickerson *et al*. (2010), a tendência das abordagens filogeográficas é chegar ao estudo genômico de multi-táxons com modelos coalescentes, testando sinal de seleção natural, com modelagem de nicho ecológico e estudos de especiação ecológica. Esta ambiciosa síntese deverá permitir sugerir relações causais entre geografia, alterações climáticas, interações ecológicas e composição genética das populações.

As recentes tecnologias que envolvem o sequenciamento de nova geração (NGS) são ferramentas poderosas para investigar as forças evolutivas que moldam os genomas. Muitas estatísticas sumárias usadas em estudos de demografia, seleção natural, e estrutura populacional são derivadas de estimativas de variação nucleotídica entre múltiplos indivíduos (NIELSEN 2005). O número de sítios segregantes e as frequências alélicas nesses locais estão entre as características mais importantes dos dados de uma perspectiva evolutiva, e são a base de testes de neutralidade comumente utilizados (FU; LI, 1993). Dentro do atual panorama, as demandas são crescentes para que os dados gerados tenham importante papel nas estratégias de conservação, as quais, segundo Moritz (2002), deveriam levar em conta os padrões de distribuição da diversidade genética e também os processos evolutivos envolvidos na origem de tais padrões. Portanto avaliar os dados de forma multidisciplinar, buscando tecnologias mais informativas e abrangentes, leva a um acréscimo fundamental para o nosso conhecimento dos processos históricos biogeográficos, consequentemente a aproximação das respostas para eventos que geraram diversificação.

Sequenciamento de nova geração - NGS

A atual era genômica conta com várias tecnologias implementadas no mercado (ex. 454/Roche, Illumina, ABI SOLiD, Ion Torrent, PacBio), cada uma das quais possui seu próprio conjunto de características (MARDIS 2017). As primeiras plataformas de NGS surgiram em 2005 e um desses instrumentos foi o 454 FLX (Roche) que usa o método de pirosequenciamento. Esse método depende da detecção de pirofosfato liberado durante a incorporação do nucleotídeo e essa amplificação é realizada em uma solução oleosa emulsificada, denominada PCR em emulsão (BERKA *et al.* 2010). Outra plataforma foi a HiSeq System (Illumina), a qual realiza o sequenciamento por síntese

(SBS; MARDIS 2017). As bibliotecas de SBS são amplificadas *in situ* em um suporte sólido banhado por fluidos com reagentes para a síntese e detecção da reação de sequenciamento. A plataforma de sequenciamento da Illumina se destacou entre as concorrentes, sendo atualmente a mais utilizada (HEAD *et al.* 2014). A plataforma Ion Torrent (Life Technologies) possui uma abordagem de sequenciamento de DNA diferenciada, visto que a identificação das bases se baseia na alteração de pH quando os nucleotídeos são incorporados na cadeia nova e não pelo uso de ddNTPs ou reações luminosas, como era o caso das tecnologias utilizadas até então (MARDIS 2017). O método empregado para o sequenciamento na plataforma SOLiD (Applied Biosystems) faz uso de fragmentos de DNA ligados a esferas (*beads*), quando a base é incorporada na cadeia sendo sintetizada há emissão de um sinal luminoso do fluoróforo acoplado à base correspondente e esse sinal é capturado pelo sistema óptico do sequenciador (MARDIS 2017).

Essas tecnologias tem em comum a capacidade de gerar rapidamente enormes quantidades de dados de sequências de forma relativamente econômica. Assim, um grande número de marcadores genéticos é analisado, o que por sua vez, auxilia na busca da base genética de variações fenotípicas ecológicas importantes. Muitas questões relevantes podem ser respondidas com dados de sequências em grande quantidade (GLENN 2011). Por exemplo, a comparação do nível de variabilidade genética em regiões associadas a genes e em regiões intergênicas pode permitir inferir sobre a ação da seleção natural nas populações. Em se tratando de estudos de ecologia molecular, houve avanços importantes. Enquanto habitualmente estudos baseados em espécies-modelo eram predominantes, com esses avanços os pesquisadores puderam estudar populações naturais de organismos e abordar questões ecológicas e evolutivas em uma escala de precisão impensável anteriormente, especialmente pelo crescente número de espécies com genomas sequenciados, viabilizando cada vez mais utilizar essa abordagem em espécies não-modelo. Nesse contexto, um grande número de espécies com relevância ecológica e conservacionista pode se beneficiar da era genômica (KOHN *et al*. 2006; EKBLOM; GALINDO, 2011).

Uma metodologia de redução da porção do genoma a ser sequenciada muito utilizada atualmente é o *genotyping-by-sequencing* (GBS; ELSHIRE *et al*. 2011). Esse método é muito similar ao *Restriction-site associated DNA sequencing* (RAD-seq; DAVEY, BLAXTER, 2010). Ambas envolvem uma etapa inicial de redução de complexidade do genoma alvo (descrita a seguir), mas para a preparação da biblioteca

de GBS os fragmentos gerados não são selecionados pelo tamanho e seu custo é comparativamente menor (ELSHIRE *et al*. 2011). Assim, o DNA total é digerido com uma ou mais enzimas de restrição gerando vários fragmentos que são ligados a adaptadores em ambas as extremidades. Esses adaptadores servem, então, como iniciadores em uma reação de PCR para aumentar a quantidade desses segmentos (BAIRD *et al*. 2008). Em seguida essa biblioteca é sequenciada por NGS. Com isso milhares de marcadores em todo o genoma são sequenciados (DAVEY; BLAXTER, 2010). A introdução recente de instrumentos capazes de produzir milhões de leituras de sequência de DNA em uma única corrida está mudando rapidamente a paisagem da genética, fornecendo a capacidade de responder perguntas complexas e interdisciplinares com inimaginável rapidez (MARDIS 2017).

As espécies-modelo

Myiothlypis leucoblephara

A espécie *Myiothlypis leucoblephara* (Vieillot 1817), também conhecida como pula-pula-assobiador, pertence à família Parulidae Wetmore, Friedmann, Lincoln, Miller, Peters, van Rossem, Van Tyne & Zimmer 1947, é endêmico da FA. Ocorre no Paraguai, Argentina, Uruguai e no sul e sudeste do Brasil. Habita preferencialmente o interior de matas úmidas e sombreadas em regiões de serra, locomovendo-se por meio de pulos no solo ou pouco acima dele, na parte baixa da vegetação (SICK 1997). Seu tarso avantajado lhe permite utilizar a técnica de forrageio "glean", que se baseia predominante na escolha de alimento no solo ou próximo dele, sem uso de movimento acrobático (FITZPATRICK, 1985; REMSEN JR; ROBINSON, 1990). Sua dieta é basicamente insetívora, com predominância de coleópteros e himenópteros, e também aracnídeos (CHATELLENAZ 2008). Possui cauda longa e larga, suas asas são verde-escuro e possui o píleo acinzentado, apresenta círculo em torno do olho e sobrancelha estreita brancos, mede 14,5 cm e possui coberteiras inferiores da cauda amareladas (SICK 1997).

A família Parulidae apresenta em torno de 22 espécies e dentre estas, pelo menos 7 espécies que ocorrem na FA (LIMA 2013). Estudos filogenéticos moleculares recentes têm revelado que muitas dessas famílias tradicionais dentro dos passeriformes, não são agrupamentos monofiléticos, levando a uma série de revisões para as atuais classificações (KLICKA *et al*. 2007). Lovette *et al.* (2010), em um estudo com 110 táxons, sugerem diversas modificações dentro da família, subdividindo-a em 14

gêneros. Segundo os resultados das análises filogenéticas moleculares, a polifilia deste gênero ficou evidente, sendo sugerida a utilização do gênero *Myiothlypis* Cabanis, 1850 para 16 táxons que pertenciam anteriormente ao gênero *Basileuterus*. Entre as espécies que sofreram essa revisão está *Myiothlypis leucoblephara,* alteração incluída na lista de espécies de aves do Brasil de 01/01/2014 pelo Comitê Brasileiro de Registros Ornitológicos(CBRO, 2015)

Batalha Filho *et al.* (2012), em um estudo com *Myiothlypis leucoblephara* baseado em marcadores nucleares e mitocondriais, abordaram algumas questões, dentre elas, se a estruturação genética atual da espécie concordava com os padrões observados para outros organismos da FA e como o último máximo glacial afetou o tamanho populacional efetivo dessa espécie. A abordagem utilizada baseada em modelos demográficos permitiu inferir a história evolutiva de *M. leucoblephara* e a dinâmica de diversificação dentro da FA. No entanto, o trabalho revelou o primeiro exemplo de um organismo desse domínio sem uma forte estruturação genética populacional e cujo tamanho da população não se alterou durante o último máximo glacial. A controvérsia dos resultados pode estar relacionada à falta de marcadores, e como principal consequência disso, ocorreria a limitação e o enfraquecimento das inferências realizadas, evidenciando um possível viés na análise.

Myrmoderus loricatus e Myrmoderus squamosus

As espécies *Myrmoderus loricatus* (Lichtenstein, 1823) e *Myrmoderus squamosus* (Pelzeln, 1868) (comumente chamados de formigueiro-assobiador e papa-formiga-de-grota, respectivamente) pertencem à Família Thamnophilidae Swainson, 1824. Essa família faz parte da ordem Passeriformes e possui um elevado número de espécies no Brasil, em torno de 238 (CBRO, 2015). Seus representantes são majoritariamente insetívoros e estão praticamente restritos a florestas de altitude mais baixas da região neotropical, habitando dos Andes até o sul do Brasil (ZIMMER; ISLER, 2003). O gênero *Myrmeciza,* considerado um dos mais diversos dentro da família, foi alvo de uma revisão taxonômica por apresentar uma série de incertezas e controversa classificação das espécies (ISLER *et al*. 2013). Devido a classificações genéricas e divergências entre autores, a composição do gênero vinha sendo modificada ao longo do tempo (Zimmer 1932; Peters 1951) e a polifilia do gênero foi confirmada por estudos moleculares (eg. IRESTEDT *et al.* 2004; BRUMFIELD *et al*. 2007). Assim, Isler *et al.* (2013), baseados em características morfológicas, comportamentais,

ecológicas e filogenéticas, sugeriram a realocação de *M. loricatus* e *M. squamosus* do gênero *Myrmeciza* para o atualmente reconhecido gênero *Myrmoderus*, o que foi aceito na lista de aves do Brasil de 01/01/2014 do Comitê Brasileiro de Registros Ornitológicos (CBRO, 2015).

As duas espécies alvo do presente projeto são consideradas irmãs e, juntamente com *M. ferruginea* e *M. ruficauda*, formam um clado monofilético bem suportado. No entanto, diferentemente de Sibley e Monroe (1990), Sick (1997) descreve *M. squamosus* como subespécie de *M. loricatus*, considerando aspectos como plumagem muito semelhante, bem como seus hábitos e sons. Possuem distribuição parapátrica, com áreas sobrepostas na região central da FA, da qual são endêmicas, habitando tanto áreas de várzea até florestas úmidas, do nível do mar até 1000 m ou superior (RIDGELY; TUDOR, 2009).

Raposo do Amaral *et al.* (2013), em um estudo de filogeografia multilocos, detectaram pouco ou nenhum fluxo gênico entre as espécies em questão, justificado pelo desenvolvimento de barreiras reprodutivas ou exclusão competitiva. Além disso, os resultados apresentados pelos autores ressaltaram a importância de utilizar diversos marcadores independentes visando obter resultados mais robustos. O efeito de amostragem inadequada pode, dentre outras consequências, comprometer as estimativas de tempo de divergência e a reconstrução da história demográfica das espécies, uma vez que a história de um ou poucos genes por vezes não representa a divergência dentro da espécie (EDWARDS; BEERLI, 2000).

Capítulos da tese

A presente Tese apresenta dois capítulos que descrevem estudos filogeográficos usando dados de milhares de SNPs de *Myiothlypis leucoblephara* (Capítulo1) e dos táxons irmãos *Myrmoderus squamosus* e *Myrmoderus loricatus* (Capítulo 2). Ambos estudos visaram contribuir com aporte teórico no entendimento da Biogeografia da Floresta Atlântica e na história evolutiva das espécies em questão. Pretendemos submeter os manuscritos às revistas: *Molecular Phylogenetics and Evolution* e *Molecular Ecology*, respectivamente.

AVISE, J.C. **Phylogeography: The History and Formation of Species**. Harvard University Press, Cambridge, 2000.

AVISE, J.C.; ARNOLD, J.; BALL, R. M.; BERMINGHAM, E.; LAMB, T.; NEIGEL, J. E.; REEB, C. A. *et al*. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. **Annual Review of Ecology and Systematic**, v.18, p. 489–522, 1987.

BAIRD, N. A; ETTER, P. D.; ATWOOD, T. S.; CURREY, M. C.; SHIVER, A. L.; LEWIS, Z. A.; SELKER, E. U. *et al*. Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PloS One**, v. 3, p. 3376, 2008.

BATALHA-FILHO, H.; CABANNE, G.S.; MIYAKI, C.Y. Phylogeography of an Atlantic forest passerine reveals demographic stability through the last glacial maximum. **Molecular Phylogenetics and Evolution**, v. 65, p. 892-902, 2012.

BERKA, J.; CHEN, Y.J.; LEAMON, J.H., LEFKOWITZ, S.; LOHMAN, K.L.; MAKHIJANI, V.B.; WEINER, M.P. Bead emulsion nucleic acid amplification. U.S. Patent No. 7,842,457. **U.S. Patent and Trademark Office**, 2010.

BROOKS, T. M.; BALMFORD, A. Atlantic forest extinctions. **Nature**, v. 380, n. 6570, p. 115-115, 1996.

BROOKS, T. M.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A.; RYLANDS, A. B.; KONSTANT, W. R.; FLICK, P. *et al*. Habitat loss and extinction in the hotspots of biodiversity. **Conservation Biology**, v. 16, n. 4, p. 909-923, 2002.

BRUMFIELD, R.T.; TELLO, J.G.; CHEVIRON, Z.A.; CARLING, M.D.; CROCHET, N.; ROSENBERG, K.V. Phylogenetic conservatism and antiquity of a tropical specialization: Army-ant-following in the typical antbirds (Thamnophilidae). **Molecular Phylogenetics and Evolution**, v. 45, p. 1–13, 2007.

CBRO – Comitê Brasileiro de Registros Ornitológicos. Listas das aves do Brasil, 2015. Disponível em http://www.cbro.org.br (acessado em 01/05/2019).

CHATELLENAZ, Mario L. Ecología alimentaria de dos especies simpátricas del género *Basileuterus* en el Noreste de Argentina. **El Hornero**, v. 23, n. 2, p. 87-93, 2008.

CI-BRASIL (Conservation International do Brasil), Fundação SOS Mata Atlântica, Fundação Biodiversitas, IPÊ, SMA-SP & SEMAD-MG. **Avaliação e Ações Prioritárias para Conservação da Biodiversidade da Mata Atlântica e Campos Sulinos**. MMA/SBF, Brasília, 2000.

DAVEY, J. W.; BLAXTER, M. L. RADSeq: next-generation population genetics. **Briefings in Functional Genomics**, v. 9, p. 416-423, 2010.

EDWARDS, S. V.; BEERLI, P. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. **Evolution**, v. 54, p. 1839–1854, 2000.

EKBLOM, R.; GALINDO, J. Applications of next generation sequencing in molecular ecology of non-model organisms. **Heredity**, v. 107, p. 1-15, 2011.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS One**, v. 6, p. e19379, 2011.

FITZPATRICK, John W. Form, foraging behavior, and adaptive radiation in the Tyrannidae. **Ornithological Monographs**, p. 447-470, 1985.

FJELDSAÅ, J.; EHRLICH, D.; LAMBIN, E.; PRINS, E. Are biodiversity 'hotspots' correlated with current ecoclimatic stability? A pilot study using the NOAA-AVHRR remote sensing data. **Biodiversity & Conservation**, v. 6, n. 3, p. 401-422, 1997.

FJELDSAÅ, J.; LOVETT, J.C. Geographical patterns of old and young species in African forest biota: the significance of specific montane areas as evolutionary centres. **Biodiversity & Conservation**, v. 6, n. 3, p. 325-346, 1997.

FU, Y.X.; LI, W.H. Statistical tests of neutrality of mutations. **Genetics**, v. 133, p. 693–709, 1993.

FUNDAÇÃO SOS MATA ATLÂNTICA/INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Atlas dos Remanescentes Florestais da Mata Atlântica,**

Período de 2000 a 2005, 2008. Disponível em www.sosma.org.br (acessado em 10⁄05⁄2018).

GALINDO-LEAL, C.G.; CÂMARA, I.G. **The Atlantic Forest of South America: biodiversity status, threats, and outlook**. Island Press, 2003.

GLENN, T. C. Field guide to next-generation DNA sequences. **Molecular Ecology Resources**, v. 11, p. 759-769, 2011.

HAFFER, J. Speciation in Amazonian forest birds. **Science**, v. 165, n. 3889, p. 131-137, 1969.

HEAD, S.R.; KOMORI, H. K.; LAMERE, S. A.; WHISENANT, T.; VAN NIEUWERBURGH, F.; SALOMON, D. R.; ORDOUKHANIAN, P. Library construction for next-generation sequencing: overviews and challenges. **Biotechniques**, v. 56, p. 61-77, 2014.

HICKERSON, M. J.; CARSTENS, B. C.; CAVENDER-BARES, J.; CRANDALL, K. A.; GRAHAM, C. H.; JOHNSON, J. B.; RISSLER, L. *et al*. Phylogeography's past, present, and future: 10 years after. **Molecular Phylogenetics and Evolution**, v. 54, p. 291-301, 2010.

INSTITUTO FLORESTAL. 2006. **Parque Estadual da Serrado Mar: Plano de Manejo**. Instituto Florestal - Secretaria do Meio Ambiente do Estado de São Paulo. http://www.iflorestal.sp.gov.br (acesso em 25/10/2008).

IRESTEDT, M.; FJELDSÅ, J.; NYLANDER, J. A. A.; ERICSON, P. G. P. Phylogenetic relationships of typical antbirds (Thamnophilidae) and test of incongruence based on Bayes factors. **BMC Evolutionary Biology**, v. 4, p. 23, 2004.

ISLER, M. L.; BRAVO, G. A.; BRUMFIELD, R. T. Taxonomic revision of *Myrmeciza* (Aves: Passeriformes: Thamnophilidae) into 12 genera based on phylogenetic, morphological, behavioral, and ecological data. **Zootaxa**, v. 3717, n. 4, p. 469-497, 2013.

JOLY, C.A.; AIDAR, M.P.M.; KLINK, C.A. Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. **Ciência e Cultura**, v. 51, n. 5/6, p. 331-348, 1999.

KLICKA, J.; BURNS, K.; SPELLMAN, G. M. Defining a monophyletic Cardinalini: a molecular perspective. **Molecular Phylogenetics and Evolution**, v. 45, p. 1014–1032, 2007.

KOHN, M. H.; MURPHY, W. J.; OSTRANDER, E. A.; WAYNE, R. K. Genomics and conservation genetics. **Trends in Ecology & Evolution**, v. 21, p. 629-637, 2006.

LEITÃO FILHO, H.F. Diversity of arboreal species in Atlantic rain forest. **Anais da Academia Brasileira de Ciências,** v. 66, p. 91-96, 1994.

LIMA, L. M. **Aves da Mata Atlântica: riqueza, composição, status, endemismos e conservação.** Tese de Doutoramento. Universidade de São Paulo. 2013

LOVETTE, I. J.; PÉREZ-EMÁN, J. L.; SULLIVAN, J. P.; BANKS, R. C.; FIORENTINO, I.; CÓRDOBA-CÓRDOBA,S.; ECHEVERRY-GALVIS, M. *et.al.* A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). **Molecular Phylogenetics and Evolution**, v. 57, n. 2, p. 753-770, 2010.

MARDIS, E. R. DNA sequencing technologies: 2006-2016. **Nature Protocols**, v. 12, p. 213-218, 2017.

METZGER, J.P. Tree functional group richness and landscape structure in a Brazilian tropical fragmented landscape. **Ecological Applications**, v. 10, n. 4, p. 1147-1161, 2000.

METZGER, J.P.; MARTENSEN, A.C.; DIXO, M.; BERNACCI, L.C.; RIBEIRO, M.C.; TEIXEIRA, A.M.G.; PARDINI, R. Time-lag in biological responses to landscape changes in a highly dynamic Atlantic forest region. **Biological Conservation**, v. 142, n. 6, p. 1166-1177, 2009.

MORITZ, C.; FAITH, D.P. Comparative phylogeography and the identification of genetically divergent areas for conservation. **Molecular Ecology**, v. 7, p. 419-429, 1998.

MORITZ, C.; PATTON, J.; SCHNEIDER, C.; SMITH, T. Diversification of rainforest faunas: an integrated molecular approach. **Annual Review of Ecology and Systematics**, v.31, p. 533-563, 2000.

MORITZ, C. Strategies to protect biological diversity and the evolutionary processes that sustain it. **Systematic Biology**, v.51, p. 238-254, 2002.

MYERS, N. Threatened biotas:" hot spots" in tropical forests. **Environmentalist**, v. 8, n. 3, p. 187-208, 1988.

MYERS, N.; MITTERMEIER, R.A.; MITTERMEIER, C.G.; FONSECA, G.A.B.D.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, p. 853 - 858, 2000.

NIELSEN, R. Molecular signatures of natural selection. **Annual Review of Genetics**, v. 39, p. 197-218, 2005.

OLIVEIRA FILHO, A.T.; FONTES, M.A.L. Patterns of floristic differentiation among Atlantic Forests in southeastern Brazil and the influence of Climate. **Biotropica**, v. 32, n. 4b, p. 793-810, 2000.

PAGLIA, A.P. Panorama geral da fauna ameaçada de extinção no Brasil. In A. B. M. MACHADO, C. SOARES MARTINS & G.M. DRUMOND (eds.). **Lista da Fauna Brasileira Ameaçada de Extinção – Incluindo a Lista das Quase Ameaçadas e Deficientes em Dados**. p. 17-22. Fundação Biodiversitas, Belo Horizonte, 2005.

PINTO, L.P.; BEDÊ, L.; PAESE, A.; FONSECA, M.; PAGLIA, A.; LAMAS, I. *et al*. Mata Atlântica Brasileira: os desafios para conservação da biodiversidade de um hotspot mundial. **Biologia da Conservação: Essências.** RiMa, São Carlos, p. 69-96, 2006.

PRIMACK, R.B.; RODRIGUES, E. **Biologia da Conservação**. E. Rodrigues, 2001.

PUORTO, G.; DA GRAÇA SALOMÃO, M.; THEAKSTON, R. D.G.; THORPE, R. S.; WARRELL, D. A.; WÜSTER, W. Combining mitochondrial DNA sequences and morphological data to infer species boundaries: phylogeography of lanceheaded pitvipers in the Brazilian Atlantic forest, and the status of *Bothrops pradoi* (Squamata: Serpentes: Viperidae). **Journal of Evolutionary Biology**, v. 14, p. 527-538, 2001.

RANTA, P.; BLOM, T.O.M.; NIEMELA, J.; JOENSUU, E.; SIITONEN, M. The fragmented Atlantic rain forest of Brazil: size, shape and distribution of forest fragments. **Biodiversity & Conservation**, v. 7, n. 3, p. 385-403, 1998.

RAPOSO DO AMARAL, F.; ALBERS, P. K.; EDWARDS, S. V.; MIYAKI, C. Y. Multilocus tests of Pleistocene refugia and ancient divergence in a pair of Atlantic Forest antbirds (*Myrmeciza*). **Molecular Ecology**, v. *22*, n. 15, p. 3996-4013, 2013.

REMSEN JR, J. V.; ROBINSON, Scott K. A classification scheme for foraging behavior of birds in terrestrial habitats. **Studies in Avian Biology**, v. 13, p. 144-160, 1990.

RIBEIRO, C.A.; METZGER, J.P.; MARTENSEN, A.C.; PONZONI, F.J.; HIROTA, M.M. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological Conservation**. v. 142, p. 1141–1153, 2009.

RIDGELY, R. S.; TUDOR, G. **Field Guide to the Songbirds of South America**. University of Texas Press, 2009. 736 p.

RIZZINI, C.T.; **Tratado de Fitogeografia do Brasil**. 2 ed. Âmbito Cultural Edições Ltda, 1997.

SANT'ANNA-NETO, J.L.; NERY, J.T. Variabilidade e mudanças climáticas no Brasil e seus impactos. In: Souza GRG, Suguio K, Oliveira MAS, Oliveira PL (Eds). **Quaternário do Brasil**. Holos Editora, p. 28–50, 2005.

ŞEKERCIOĞLU, C.H.; PRIMACK, R.B.; WORMWORTH, J. The effects of climate change on tropical birds. **Biological Conservation**, v. 148, p. 1-18, 2012.

SIBLEY, C. G.; MONROE, B. L. **Distribution and Taxonomy of Birds of the World**. Yale University Press, 976 p, 1990.

SICK, H. **Ornitologia Brasileira**. Nova Fronteira, 862 p, 1997.

SILVA, J.M.C.; CASTELETI, C.H.M. Status of the biodiversity of the Atlantic Forest of Brazil. In: C. Galindo-Leal & I.G. Câmara (eds.). **The Atlantic Forest of**

South America: Biodiversity Status, Threats, and Outlook. Center for Applied Biodiversity Science and Island Press, p. 43-59, 2003.

SILVA, J.M.C.; STRAUBE, F.C. Systematics and biogeography of scaled woodcreepers (Aves: Dendrocolaptidae). **Studies on Neotropical Fauna and Environment**, v. 31, p. 3-10, 1996.

SMITH, T.B.; WAYNE, R.K.; GIRMAN, D.J.; BRUFORD, M.W. A role for ecotones in generating rainforest biodiversity. **Science**, v. 276, n. 5320, p. 1855-1857, 1997.

STOTZ, D.F.; FITZPATRICK, J.W.; PARKER III, T.A.; MOSKOVITS, D.K.; SNOW, D. **Neotropical Birds: Ecology and Conservation**, University of Chicago Press, 478 p, 1996.

TABARELLI, M.; VICENTE, A.; BARBOSA, D.C.A. Variation of seed dispersal spectrum of woody plants across a rainfall gradient in north-eastern Brazil. **Journal of Arid Environments**, v. 53, n. 2, p. 197-210, 2003.

VANZOLINI, P.E.; WILLIAMS, E. E. South American anoles: the geographic differentiation and evolution of the *Anolis chrysolepis* species group (Sauria, Iguanidae). **Arquivos de Zoologia**, v. 19, n. 1-2, p. 1-298, 1970.

WALLACE, A.R. On the monkeys of the Amazon. **Proceedings of the Zoological Society of London**, v. 20, p. 107-110, 1852.

ZIMMER, K.; ISLER, M.L. White-bibbed antbird (*Myrmeciza loricata*). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.). **Handbook of the Birds of the World Alive**. Lynx Edicions, Barcelona, 2003. Disponível em http://www.hbw.com/node/56825 (acessado em 03/11/2018).

Subgenomic data reveals isolation by distance in an endemic Atlantic Forest passerine

Introduction

The Atlantic Forest (AF) is one of the world's most biodiverse ecosystems, originally covering about 150 million hectares, along the coastal region of eastern Brazil to eastern Paraguay and northeastern Argentina (Myers et al., 2000; Galindo-Leal and Câmara, 2003). The AF presents less than 16% of its original area (Ribeiro et al., 2009). It is considered as one of the top five priority conservation biomes worldwide. It is home to more than 8000 endemic species (Myers et al., 2000) and more than 60% species the official list of threatened Brazilian fauna (Tabarelli et al., 2003; Paglia, 2005). The AF has a set of heterogeneous geographic and climatic variables. Elevation ranges from sea level up to 2,900m, with abrupt changes in the landscape and high variation of humidity average, depth of soil and air temperature, resulting in different phytophysiognomies (Galindo-Leal and Câmara, 2003). Several hypotheses have been proposed to explain the origin of diversity in subtropical and tropical forests (Moritz et al., 2000). Isolation in paleorefuges and by geographic barriers (e.g. rivers, mountains) are two of the most discussed hypotheses (Pellegrino et al., 2005; Cabanne et al., 2007, Costa et al., 2000; Carnaval et al., 2009; d'Horta et al., 2011). The refuge hypothesis suggests that especially in the late Pleistocene the rainforest possibly contracted during glacial maxima into forest refugia isolated by open vegetation areas. Within these refugia populations of forest-dependent taxa seem to have been isolated and could have originated differentiated lineages. During interglacial cycles, when the global temperature and humidity increased, humid forests expanded and these lineages got in contact again, but if they were differentiated enough, there would be no gene flow between them (Brown and Ab'Saber, 1979; Graham et al., 2006; Sant'Anna-Neto and Nery, 2005; Carnaval and Moritz, 2008).

Ecological and historic processes of lineages diversification have been widely studied under the light of phylogeography, and understanding micro and macro evolutionary processes allows us to plan conservation strategies (Moritz, 2002). Currently, this subject has become a huge challenge to be faced, once researchers estimated that due to climate change 600 to 900 species of birds shall be extinct by the year 2100, mostly (89%) in tropical environments (Şekercioğlu et al., 2012).

We inferred the demographic history of the AF endemic bird *Myiothlypis leucoblephara* (Vieillot 1817) also known as white-rimmed warbler. This species belongs to the family Parulidae (Passeriformes). It occurs in Paraguay, Argentina, Uruguay and in the south and southeast of Brazil (Fig. 1). It inhabits preferably the

interior of humid and shaded forests in mountainous regions, foraging close to the ground, in the low part of the vegetation (Sick, 1997). The present study is one of the few to carry out a phylogeographic study using thousands of markers. We inferred the evolutionary history of the white-rimmed warbler to contribute to the understanding of the processes underlying the origin of the biodiversity of the AF. We described the demographic history and the level of population structure of this bird and suggest possible past events associated to these results. From the phylogeographic perspective, the use of thousands of markers allows for more refined identification of lineages within species and inference of the demographic history (Davey and Blaxter, 2010; Edwards et al., 2015; Mardis, 2017).

Material and Methods

Samples and data processing

We obtained samples from 104 individuals of white-rimmed warbler, but after the read quality filtering 86 individuals from 55 locations in Brazil, Uruguay, and Argentina were retained for further analyses (Fig. 1, Table 1). Tissue samples (muscle, blood or feather tip) were preserved in absolute ethanol and stored at -20ºC. DNA was extracted with the DNeasy kit (Qiagen Inc.). The total amount of extracted DNA was quantified using Qubit (Thermo Fisher Scientific) and then concentrated to ~100 ng/µl. DNA library construction (Genotyping-by-sequencing digested with *Pst*I; Elshire et al., 2011) and sequencing (single-end in HiSeq 2500, Illumina) were outsourced at Ecomol consultoria, ESALQ/USP.

FastQC (Andrews, 2010) was used to check the general quality of the reads. Then we used ipyrad (Eaton and Overcast, 2016), which is a set of scripts intended to process reads of libraries constructed using restriction enzyme. This processing consists of seven steps with various parameters. We tested different values of these parameters in seven independent ipyrad runs to evaluate our data. The final values are in bold as follows. Step 1: This step used the barcode file information - the list of samples and their respective indexing codes - to separate the reads per sample. Step 2: This step used the Phred index (a measure of the quality of each nucleotide) recorded in the sequencing files to the poor quality bases. Sites with Phred values below **33** were transformed into Ns, reads with more than **four** Ns were discarded, and we applied a **strict filter for adapters**. Step 3: This step records the number of times each read was observed by clustering those with at least **90%** of similarity and removes replicated ones. Step 4:

This step used the Lynch maximum likelihood equation (Lynch, 2008) to jointly estimate the rate of sequencing error and the mean heterozygosity from the base frequencies at each site of the reads of each individual. Based on these values, the binomial probability of a site being homozygous or heterozygous was calculated. Step 5: Based on the rates estimated in the previous step, this step generated consensus sequences for each sample. Reads were discarded when they presented coverage lower than **eight**, more than **five** undetermined sites (Ns), more than **four** heterozygous sites (Hs). Loci with more than **two** alleles were also discarded. Step 6: In this step, the consensus sequences were pooled between samples using the similarity values generated in step 3. In the case of a heterozygous individual, one of its alleles was randomly sampled and used in this step, although both alleles remain present in the set of the final data. Step 7: In this step sequences were aligned potential paralogs under the following parameters: maximum of **50%** shared polymorphic sites in a locus, maximum of **five** indels and **five** SNPs per locus. The final dataset was obtained as output files.

We also checked two quality indicative flags for all samples in VCFtools (Danecek et al., 2011): "--indv-freq-burden" (number of variants [e.g. singletons, doubletons] for each individual in comparison to the mean frequency of variants of the total sampling, this allows to detect individuals whose sequences may not have the same quality as the others and they were excluded) and "--singletons" (identifies the loci and individuals that present singletons). After this quality checking and filtering processes using fastQC, ipyrad, and VCFtools we detected that 18 samples (not included in Table 1 and Fig. 1) presented general low quality and less than 4,000 reads (see “Nreads” in Table S1). After the exclusion of these 18 samples ipyrad was run again with the same parameters described above. Then using VCFtools we randomly selected one SNP per locus (unlinked SNPs) and finally SNPs matrices with various percentages of completeness of data were obtained (Table S2).

Genetic structure

Genetic ancestry estimation is a broad term that is related to several distinct issues in population genetics (Pritchard et al., 2000). Here we used an R package for Landscape and Ecological Associations studies (Frichot and François, 2015) to infer population structure, the number of distinct populations and to assign individuals to these populations. In this package, the function sNMF estimates the ancestral coefficients using a non-negative sparse factorization matrix and can analyze large bi-allelic datasets

efficiently without loss of accuracy, even in scenarios of departure of linkage and Hardy Weinberg equilibria (Frichot et al., 2014). The number of ancestral populations (K) varied from 1 to 10, for each K 100 independent runs were performed and four values of the alpha regularization parameter (1, 10, 100 and 1000) were tested. The evaluation of the best K followed the criterion of cross-entropy, which is based on the implementation of a masked data matrix. The values of each run were plotted on a graph and the best K was the one with the lowest value of cross-entropy. SNP matrices with 95%, 90% and 85% of completeness of data were analyzed in sNMF and the results obtained were similar.

We also used Principal Component Analysis (PCA) to analyze the genetic diversity among the sampled individuals using the SNP matrix with 90% completeness. This analysis is based on eigenvalue decomposition of a correlation matrix and the data was plotted to visualize the genetic distance between individuals using the package ADEGENET 2.0 (Jombart and Collins, 2015).

We used the estimated effective migration surface (EEMS) method, which requires a repository to be implemented (e.g., C++) to visualize spatial population structure from geo-referenced genetic samples using the concept of effective migration to model the relationship between geography and genetics, resulting in a visual representation of population structure that allows to highlight potential areas with higher and lower-than-average migration and infers the location of possible barriers and corridors of historic gene flow (Petkova et al., 2015). EEMS estimates and maps the genetic differentiation by mapping effective migration surface among demes based on a spatially explicit approach using a Markov Chain Monte Carlo (MCMC) to test all possible routes of migration between two demes. This approach allows to estimate demographic parameters through sampling from their posterior distributions given the observed genetic dissimilarity between individuals (computed in an Euclidian matrix) based on SNP datasets excluding singletons in ADEGENET 2.0. We produced habitat polygons using GIS tools (QGIS, 2012) based on the species geographic distribution, and 500 demes were distributed over the inhabited area. To describe the within-deme and between-deme components of genetic dissimilarity, EEMS calculates D(a,b) that is the dissimilarity between one individual from deme a and another individual from deme b. Then, the within-deme component for a and b is simply D(a,a) and D(b,b), respectively. The between-deme component is D(a,b) - [D(a,a) + D(b,b)] / 2 and it represents dissimilarity that is due to the spatial structure of the population and is not a

consequence of the local diversity in the two demes (Petkova et al., 2015). Each MCMC run was performed for $20x10^6$ generations with the first $10x10^5$ generations excluded as burn-in and with 9999 interactions. Maps were generated using additional features of the EEMS R package.

Also in ADEGENET 2.0 we performed an isolation by distance (IBD) test with 999 replications. This analysis is a Mantel (r) test calculated as the usual Pearson correlation coefficient between the two matrices (genetic and geographic distances): $Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_{ij}\, y_{ij}$, where Z is standardized by the variances in the two matrices (Wright, 1943).

A phylogeographic study based on two markers (one mitochondrial and one nuclear) with 62 samples of Myiothlypis leucoblepharus (Batalha Filho et al., 2012) revealed the first example of an AF organism without a strong population genetic structure and whose population size seem not to have changed during the last glacial maximum. To compare our results based o thousands of SNPs with those from this previous study, we obtained a subset of 27 samples shared between both studies (Fig. 1, Table 1) utilizing VCFtools. This SNP matrix (1 SNP per locus) presented 90% of completeness and was used in sNMF, PCA and Mantel tests analyses performed as described for our whole dataset.

Demographic inference

To infer the history of the species under a demographic approach we used (in a python 3 environment) the package Momi2 (MOran Models for Inference 2, Kamm et al., 2019). This method is based on the variant lookdown construction based on the Moran model that is an stochastic process used to produce a sampling distribution similar to the coalescent (Moran, 1962). Associated with Momi2 we used a Jupyter notebook to run the analyses, which is a software that supports interactive science data and scientific computing across programming languages. Momi2 computes the expected sample frequency spectrum (SFS), which is a histogram of allele counts used to infer the history of population size changes, migrations, and other demographic events affecting populations, once different population histories leave characteristic signatures on the summary statistic of SFS. We simulated four different demographic scenarios/models (Fig. S1) considering previous findings for the species (Batalha-Filho et al., 2012): model 1- panmictic population with constant size, model 2- exponential population growth, model 3- exponential population contraction, and model 4- exponential population contraction followed by exponential population growth. Model 2

was subdivided into 2I, 2II, 2III, and 2IV, to test different time parameters. We set uniform priors for current effective population size (Ne; 1), size of instantaneous population size (lower 10,000/upper 1,500,000), growth rate (model 2: upper 0.01/lower 0.0001; model 3: upper 0.0001/lower 0.01, and model 4: expansion upper 0.01/lower 0.0001 and contraction upper -0.0001/lower -0.01) and time (model 2: 10 thousand years ago [kya]-500 kya; model 2I: 10-50 kya; model 2II: 50-100 kya; model 2III: 100-300 kya; model 2IV: 300-500 kya; model 3: 10-130 kya; model 4: contraction 10-250 kya and expansion 10-130 kya). We assumed an average mutation rate of $2.5x10^{-9}$ (Nadachowska-Brzyska et al., 2015) and we assumed one year as the generation time (Cabanne et al., 2008) to estimate divergence times. We performed 1000 runs with 2000 interactions for each model.

In order to select the best-fit model to the observed data we implemented an information theory procedure based on an estimator of the relative quality of statistical models, the Akaike information criterion (AIC), which extracts the log-likelihood of each model, calculates the AIC and then calculates delta AIC values and AIC weights. The best model shows the lowest AIC score. Delta AIC and the AIC weight are indications of how confident we can be that the best fitting model is the correct model. To obtain 95% confidence intervals for the parameters of the best-fit model we performed 100 parametric bootstraps (Kamm et al., 2019).

Results

The final data processing produced raw VCF files with 153,758 SNPs for the 86 individuals sampled. The average number of consensus of clustered reads (see "Reads cons", Table S1) was 66,162.7; the average size of concatenated contigs (see "N sites", Table S1) was 683.048 bp, and the average site depth was 24.5 (see "Avg depth", Table S1). After randomly selecting one SNP per loci we obtained 8,707 SNPs and 4,902 SNPs for matrices with 90% and 95% data, respectively (Table S2). The population structure analyses in sNMF presented the best value of cross-entropy (value = 0.2404) with alpha parameter of 1000 (Fig. S2, Table S3) with K=2. Individuals sampled in the north (MG and RJ states) were associated with the red cluster (Figs. 2, S3). Individuals sampled in the central region (SP, PR, and SC) showed mixed ancestry. Individuals predominantly carrying the other genetic set (green, Figs. 2, S3) were sampled in the south/west of the species distribution (RS, Argentina, and Uruguay). The second-best model estimated in sNMF supported the presence of three ancestral populations (K=3)

with alpha parameter =100 and cross-entropy value = 0.241 (Table S3). With K=3 almost all individuals presented a shared genetic ancestry component and showed a gradual increase of this component from north to south (in yellow, Fig. S4a). Also, we analyzed one sample of *M. leucoblephara lemurum* Olson (1975) and based on our results (Figs. 2, S4a) it was not possible to genetically differentiate this sample from individuals of *M. leucoblephara leucoblephara* since it has the same genetic ancestry (green, Fig. 2) of RS and Argentina samples.

The PCA result was congruent with the sNMF result and indicated the same gradual change of the genetic diversity. The first component showed a north to south change, while the second component showed longitudinal variation (east to west) also indicating isolation by distance (Fig. 3). The results of EEMS based on 7,675 SNPs (no singletons) showed that the central portion of the species' distribution presented the highest genetic variability, this is where the two genetic groups indicated by sNMF co-occur (Fig. 4). The spatial distribution of this genetic variability in the effective migration map showed no clear pattern associated with regions or localities, suggesting that gene flow is occurring irregularly along the distribution area (Fig. 4). EEMS also indicated the presence of a pattern of isolation by distance (IBD), as the dissimilarity between individuals increases according to their geographic distance (Fig. S5). The Mantel test for IBD showed a significant effect (p-value = 0.001) of geographic distance on the genetic distance, indicated by a moderate correlation between the two matrices (r = 0.529; Fig. 5). This indicates that spatial variation affects allele frequencies that change slowly from one end of the landscape to the other. We also detected an IBD signal based on 10,693 SNPs (matrix completeness of 90%) of 27 samples shared between the present study and Batalha Filho et al. (2012). The population structure obtained in sNMF presented the best value of cross-entropy (value = 0,4585) with K=1 and alpha parameter=1. The second-best model supported the presence of K=2 with alpha parameter =1 and cross-entropy value = 0.4810. Individuals sampled in the north (MG and RJ states) were associated with the red cluster (Fig. S4b). Those from the central region (SP, PR) showed mixed ancestry. Individuals carrying the other genetic set (green, Figs. S4b) were sampled in the south/west of the species distribution (PR, SC, RS, Argentina). The PCA result (Fig. S6) was congruent with the sNMF result and indicated the same gradual change of the genetic diversity, detected mainly by the first component. The Mantel test for IBD showed a significant effect (p-value = 0.001) of

geographic distance on the genetic distance, indicated by a moderate correlation between the two matrices (r = 0,559; Fig. S6).

The simulated model 2 was the best one as indicated by the AIC value (46590.14). This scenario included demographic expansion with a growth rate of 0.0001 per year starting 27,238 years ago (Table S4). Therefore, additional scenarios based on this model were tested using four different time parameters (submodels 2I-2IV) to evaluate the estimated date of expansion (Fig. S1). After including these four submodels in the analysis (total of eight scenarios), we obtained similar AIC values for model 2 and submodel 2I (Table S4); thus reinforcing that the expansion possibly started 27,238 years ago. The values and 95% confidence intervals (obtained based on 100 bootstrap replicates) of parameters of model 2 were: effective population size of diploid individuals of the current population (Ne) of 1,500,000 (1,366,558.70-1,466,346.95); effective population size of diploid individuals of the ancestral population (Ne Anc.) of 98,431.4 (95,448.86-101,542.12); time of population expansion (T1) at 27,238.6 years ago (20,273.7-46,977.89), and growth rate of 0,0001 (0.0001-0.0001).

Discussion

Genetic structure and demographic inference

Phylogeography has helped provide insights on historical processes responsible for the high species diversity of the AF (e.g., Thomé et al., 2010; d'Horta et al., 2011; Cabanne et al., 2016) and here we used a sub genomic dataset to model the diversification and demographic history of the white-rimmed warbler. Our results support a clear pattern of isolation by distance (IBD; Figs. 2, 3, 5, S3, S4a), both in the north/south and east/west directions (Fig. 3), with genetic distance reflecting geographic distance. This result is similar to those obtained in other two phylogeographic studies of AF organisms (birds-Cabanne et al., 2007; land planarians- Álvarez-Presas et al., 2014). In these studies the authors observed that pairs of populations close to each other were more genetically similar to each other than populations farther away from each other, since IBD is a consequence of limited dispersal across the space, as described by Wright (1943). Preliminary song analyses of white-rimmed warbler also point out differences between individuals sampled in the extremities of the distribution, with individuals singing fewer notes for longer intervals and with lower frequencies in the north (G. Macedo, pers. comm.).

The IBD observed here (Figs. 3, 5, S5) could be the outcome of a relatively strong territorial behavior (with territories usually shared by a couple). Unfortunately, there is no estimate of the territory size (Duarte, 2017), but this species responds very well to playback, which is a typical territorialist behavior (Sick, 1997; Uezu et al., 2005). Thus, the IBD could be due to short dispersal distances, but there is no field data that could support this hypothesis. However, our results (Figs. 2, S3, S4a) indicated that there seem to have enough gene flow over short distances to prevent local genetic differentiation, and this could be driven by juvenile dispersal (Brown, 1969).

As no population structure was observed, no potential geographic barriers to gene flow (e.g., rivers, mountains) or vicariant event (e.g., past forest refugia in the Pleistocene; Carnaval and Moritz, 2008; Carnaval et al., 2009) suggested by previous studies of AF birds (Cabanne et al., 2008; d'Horta et al., 2011; Maldonado-Coelho, 2012; Amaral et al., 2013) could be supported for the white-rimmed warbler. However, Carnaval et al. (2014) presented a palaeomodel for the AF that did not indicate dramatic forest retraction linked to glaciations in the southern region and thus, subtropical AF could have harbored forested areas, with isolated patches of forest persisting in the south. This hypothesis is congruent with results obtained for birds (Batalha-Filho et al., 2012; Cabanne et al., 2013) and amphibians (Amaro et al., 2012; Carnaval et al., 2014). As the white-rimmed warbler occurs in subtropical AF, its absence of genetic structure is also congruent with the hypothesis of a stable forest in this area.

According to our scenarios results (Fig. S1, Table S4), there was a signal of demographic expansion within the upper Pleistocene, starting 27,238 years ago (Table S4), which is coincident with the Last Glacial Maximum (LGM, 18 to 48 kya;, Behling, 2002; Behling et al., 2005; Sant'Anna-Neto and Nery, 2005; Anderson et al., 2007). This result suggests that the white-rimmed warbler started to increase its population size near to the end of the LGM (14 to 20 kya). This is in accordance with one of the predictions derived from a forest refuge model that hypothesizes that organisms from the southern part of the AF are expected to show signal of demographic expansion as a consequence of postglacial expansion of the forest in this region (Carnaval and Moritz, 2008). This demographic expansion was also observed in phylogeographic studies of other AF vertebrate species that showed – as opposed to our results for the white-rimmed warbler – genetic structure associated to latitude (Cabanne et al., 2008; Carnaval et al. 2009; Martins et al., 2009; Ribeiro et al., 2010; d'Horta et al., 2011; Martins, 2011; Maldonado-Coelho, 2012). Besides, our results (Fig. S1, Table S4) are

in accordance with Françoso et al. (2016), that described a phylogeographic study of a bumblebee associated to colder conditions and whose distribution expanded during the LGM. The white-rimmed warbler is cold-adapted, it occurs at higher altitudes in lower latitudes (e.g., 22°, 800 to 2200 m) and in higher latitudes it occurs at sea level (30° at sea level; Sick, 1997; Barry, 1992). Possibly the cold tolerance ability could have conferred some adaptive advantage during the LGM (Amaral et al. 2013; Françoso et al. 2016).

As mentioned before, in comparison to other AF organisms (Carnaval et al., 2009; Martins et al., 2009; Batalha-Filho et al., 2010; Ribeiro et al., 2010; Thomé et al., 2010; d'Horta et al., 2011; Cabanne et al., 2011; Maldonado-Coelho, 2012) the white-rimmed warbler did not exhibit any population genetic structure. This could be associated to its relatively flexible habitat requirements, that seem to allow it to be an abundant species in both large and small fragments, as long as they are connected (Uezu et al., 2005), and with a medium degree of sensitivity to disturbances and tolerance to habitat fragmentation (Stotz et al., 1996). Thus, it is possible that barriers that preclude gene flow in other species do not influence as strongly the white-rimmed warbler.

We found that the eastern portion of the state of São Paulo harbors a higher than average effective diversity (blue area, Fig. 4b). This coincides with an area that harbors higher genetic diversity of bumblebees (Françoso et al., 2016). Also, some authors suggested that the coastal area of São Paulo could have been a refugium from where populations expanded (Carnaval et al., 2009; Martins, 2011). Thus, the coast of São Paulo seems to be an important area in terms of conservation. Fortunately the largest protected areas within the AF are located in this region (Ribeiro et al., 2009) and their conservation is key for the maintenance of this biodiversity.

The resolution power of thousands of markers

We observed some similarities and some differences in the results of phylogeographic analyses of the white-rimmed warbler based on few (Batalha-Filho et al.; 2012) and on thousands of markers (present study). Batalha-Filho et al. (2012) used 914 bp of a mitochondrial gene and 512 bp of a nuclear intron and detected absence of strong population genetic structure and presence of genetic clusters for both loci, which did not correspond to any geographic structure. Based on thousands of SNPs no strong genetic structure was detected, but a gradient of change of genetic composition involving two ancestral populations was observed in the north-south axis of the

distribution of this species (K=2, Figs. S2, S3, Table S3). When we used a subset data with common 27 samples between the studies, we still detect IBD (Figs. S4b, S6) attributing our population structure results not to the increase of locations (40 locations and 24 samples increased) but to the substantial amount of markers considered. Our results underscore the limitations of Sanger-based phylogeography in accessing the structure of population (Raposo do Amaral et al., 2018), once previous results did not detect IBD or identify genetic structure associated with extremities, where the evolutionary history of white-rimmed warbler seems to be overlooked due to lack of statistical power of datasets of two loci.

Besides, Batalha-Filho et al. (2012) observed demographic stability during the LGM and demographic expansion during the upper Pleistocene (300,000 - 500,000 years ago, ya). In our study we detected an expansion starting around 27,238 ya (95% CI: 20,273.7 - 46,977.89; Table S4) that is coincident with the LGM and much younger than the previously estimated date. Also, our scenario 2IV [that was based on the expansion interval estimated by Batalha-Filho et al. (2012)] was the second less likely (AIC 47793.5, Tab. S4). The distinct temporal scales could also be associated to the different amount of markers used in the two studies. Data from hundreds of unlinked loci allow us to assess the demographic history across the genome, and, because it is possible to sequence each individual at large numbers of loci, serve as independent instantiations of the coalescent process (Wang & Bradburd, 2014), while a single or a few loci may recover particular stories leading to errors such as lack of detection of incomplete lineage sorting (Alves et al., 2003; Edwards et al., 2015). Therefore inferring demographic history based on one or few loci may recover gene histories instead of organismal histories (Alves et al., 2003). Besides, mitochondrial DNA may not be under neutral evolution (Godinho el al., 2008; Nabholz et al., 2009). The mitochondrial genome of the white-rimmed warbler (F. Raposo do Amaral, pers. comm.) does not present any site for *Pst*I and thus, it is unlikely that any of the SNPs analyzed here belong to this genome. We used one SNPs per locus, thus recombination is not an issue (Hare, 2001). Also, there is mounting evidence that the uncertainty of estimates of demographic parameters decrease with increasing numbers of loci (Balakrishnan et al., 2010). This was already discussed by Taberlet et al. (1998) in a comparative study among different taxa that observed the absence of congruence among phylogeography patterns, affirming that molecular genetic studies may be of limited use in dating a vicariant event that occurred during the Quaternary. Based on 3,741 SNPs Emerson et

al. (2010) resolved fine-scale genetic divergences among intraspecific populations of a mosquito that have been separated for less than 20 kya. They argue that in cases of recent population differentiation, such as postglacial range expansion, few markers may fail to provide sufficient resolution to infer patterns of population relatedness with high level of certainty. Also, Ströher et al. (2019) studied seven species of ants and emphasized that the use of subgenomic dataset in their comparative phylogeography resulted in more precision at various timescales, phylogenetic and demographic histories, than few markers.

Phylogeography is experiencing a revolution brought by next-generation sequencing methods, incorporating new questions and new aspects to be investigated (Edwards et al., 2015) as novel or increased numbers of molecular markers at a genome-wide scale (Etter et al., 2012) can be accessed. The present study showed that the use of a larger dataset allowed the detection of an IBD pattern that was not found based on few markers. This is in agreement with other studies that show that the re-analyses of the demographic history of organisms using more data shall result in more robust results (Emerson et al., 2010; McCormack et al., 2012; Ng et al., 2017; Raposo do Amaral et al., 2018). Also, the use of SNPs promise to provide more comparable datasets between phylogeographic studies than those based on organelle genomes (Edwards et al., 2015).

The analytical gain increases the resolution and accuracy of historical demography investigations and this strategy is just starting to be applied to Atlantic Forest organisms.

Conclusion

Here we inferred the evolutionary history of an AF bird species based on thousands of markers and we recovered a strict pattern of IBD in multiple analyses, suggesting that discrete barriers do not limit movements in this species. Also our approach revealed population expansion in LGM. The analyses of AF taxa have been unveiling a complex mixture of historical processes (Costa, 2003). Thus, it is clear that the prediction of possible future scenarios for this biome and its components will be extremely complex and should take into account specific ecological traits in an integrative approach to reconstruct the demographic history of a species (Cabanne et al., 2016; Batalha-Filho and Miyaki, 2016). The subgenomic data used here was efficient in recovered the demographic history of the white-rimmed warbler that seems to have been influenced by climatic changes during the Pleistocene. Also, our results reinforce the importance

of using large quantities of markers to infer how demographic events happened and to characterize populations.

References

Alves, P.C., Ferrand, N., Suchentrunk, F., Harris, D.J., 2003. Ancient introgression of Lepus timidus mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. Mol. Phylogenet. Evol. 27, 70–80.

Álvarez-Presas, M., Sánchez-Gracia, A., Carbayo, F., Rozas, J., Riutort, M., 2014. Insights into the origin and distribution of biodiversity in the Brazilian Atlantic forest hot spot: a statistical phylogeographic study using a low-dispersal organism. Heredity 112, 656-665.

Amaro, R.C., Rodrigues, M.T., Yonenaga-Yassuda, Y., Carnaval, A.C., 2012. Demographic processes in the montane Atlantic rainforest: molecular and cytogenetic evidence from the endemic frog *Proceratophrys boiei*. Mol. Phylogenet. Evol. 62, 880-888.

Amaral, F. R., 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*). Mol. Ecol., 22, 3996–4013.

Anderson, D.E., Goudie, A.S., Parker, A.G., 2007. Global Environments Through the Quaternary. Oxford University Press, Oxford.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqcReference Source.

Balakrishnan, C.N., Lee, J.Y., Edwards, S.V., 2010. Phylogeography and phylogenetics in the nuclear age. In: Grant, P., Grant, R. (Eds.), Searching for the Causes of Evolution: from Field Observations to Mechanisms. Princeton University Press, Princeton, pp. 65–88.

Barry, R. G., 1992. Mountain Weather and Climate. Psychology Press, London.

Batalha-Filho, H., Cabanne, G.S., Miyaki, C.Y., 2012. Phylogeography of an Atlantic forest passerine reveals demographic stability through the last glacial maximum. Mol. Phylogenet. Evol. 65, 892-902.

Batalha-Filho H., Miyaki, C.Y., 2016. Late Pleistocene divergence and postglacial expansion in the Brazilian Atlantic Forest: multilocus phylogeography of *Rhopias gularis* (Aves: Passeriformes). J. Zool. Syst. Evol. Res. 54, 137-147.

Batalha-Filho, H., Waldschmidt, A.M., Campos, L.A.O., Tavares, M.G., Fernandes-Salomão, T.M., 2010. Phylogeography and historical demography of the Neotropical

stingless bee *Melipona quadrifasciata* (Hymenoptera, Apidae): incongruence between morphology and mitochondrial DNA. Apidologie 41, 534–547.

Behling, H., 1998. Late Quaternary vegetational and climatic changes in Brazil. Rev. Palaeobot. Palynol. 99, 143–156.

Behling, H., 2002. South and southeast Brazilian grasslands during Late Quaternary times: a synthesis. Palaeogeogr. Palaeoclimatol. Palaeoecol. 177, 19–27.

Behling, H., Pillar, V.D.P., Bauermann, S.G., 2005. Late Quaternary grassland (Campos), gallery forest, fire and climate dynamics, studied by pollen, charcoal and multivariate analysis of the São Francisco de Assis core in western Rio Grande do Sul (southern Brazil). Rev. Palaeobot. Palynol. 133, 235–248.

Brown, J.L., 1969. Territorial behavior and population regulation in birds: a review and re-evaluation. Wilson Bull. 293-329.

Brown K.S., Ab'Saber, A.N., 1979. Ice-ages forest refuges and evolution in Neotropics: correlation of paleoclimatoligical, geomorphological and pedological data with biological endemism. Paleoclimas 5, 1-30.

Cabanne, G.S., Calderón, L., Trujillo Arias, N., Flores, P., Pessoa, R., d'Horta, F. M., Miyaki, C.Y., 2016. Effects of Pleistocene climate changes on species ranges and evolutionary processes in the Neotropical Atlantic Forest. Biol. J. Linn. Soc. 119, 856-872.

Cabanne, G.S., d'Horta, F.M., Sari, E.H.R., Santos, F.R., Miyaki, C.Y., 2008. Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): biogeography and systematic implications. Mol. Phylogenet. Evol. 49, 760–773.

Cabanne, G.S., d'Horta, F.M., Meyer, D., Silva, J.M.C., Miyaki, C.Y., 2011. Evolution of *Dendrocolaptes platyrostris* (Aves: Furnariidae) between the South American open vegetation corridor and the Atlantic forest. Biol. J. Linn. Soc. 103, 801–820.

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. Biol. J. Linn. Soc. 91, 73–84.

Cabanne, G.S., Sari, E.H.R., Meyer, D., Santos, F.R., Miyaki, C.Y., 2013. Matrilineal evidence for demographic expansion, low diversity and lack of phylogeographic structure in the Atlantic forest endemic greenish schiffornis *Schiffornis virescens* (Aves: Tityridae). J. Ornithol. 154, 371-384.

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

Carnaval, A.C., Moritz, C., 2008. Historical climate modeling predicts patterns of current biodiversity in the Brazilian Atlantic forest. J. Biogeogr. 35, 1187–1201.

Carnaval, A.C., Waltari, E., Rodrigues, M.T., Rosauer, D., VanDerWal, J., Damasceno, R., Prates, I., Strangas, M., Spanos, Z., Rivera, D., Pie, M., Firkowski, C.R., Bornschein, M.R., Ribeiro, L.F., Moritz, C., 2014. Prediction of phylogeographic endemism in an environmentally complex biome. Proc. R. Soc. London 281, 1-10.

Costa, L.P., 2003. The historical bridge between the Amazon and the Atlantic forest of Brazil: a study of molecular phylogeography with small mammals. J. Biogeogr. 30, 71–86.

Costa, L.P., Leite, Y.L., Fonseca, G.A.B., Fonseca, M.T., 2000. Biogeography of South American Forest mammals: endemism and diversity in the Atlantic forest. Biotropica 32, 872-881.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., Mcvean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9, 416-423.

d'Horta, F., Cabanne, G.S., Meyer, D., Miyaki, C.Y., 2011. The genetic effects of Late Quaternary climatic changes over a tropical latitudinal gradient: diversification of an Atlantic Forest passerine. Mol. Ecol. 20, 1932–1935.

Duarte, T.B.F., 2017. Efeito da estrutura da paisagem na conectividade para ocorrência de aves florestais em fragmentos de mata atlântica em um cenário urbano. Master's Thesis. Universidade Federal de Ouro Preto.

Eaton, D.A.R., Overcast, I., 2016. Ipyrad: interactive assembly and analysis of RADseq data sets. Available from: http://ipyrad.readthedocs.io/.

Edwards, S.V., Shultz, A.J., Campbell-Staton, S.C., 2015. Next-generation sequencing and the expanding domain of phylogeography. Folia Zool. 64, 187-207.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One 6, 19379.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E., Holzapfel, C.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proc. Natl. Acad. Sci. USA 107, 16196-16200.

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A., 2012. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo, V., Rockman, M.V. (Eds.), Molecular Methods for Evolutionary Genetics. Human Press, New York, pp. 157-178.

Françoso, E., Zuntini, A.R., Carnaval, A.C., Arias, M.C. 2016. Comparative phylogeography in the Atlantic forest and Brazilian savannas: Pleistocene fluctuations and dispersal shape spatial patterns in two bumblebees. BMC Evol. Biol. 16, 267.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., François, O., 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics 196, 973–983.

Frichot, E., François, O., 2015. LEA: an R package for landscape and ecological association studies. Methods Ecol. Evol. 6, 925-929.

Galindo-Leal, C.G., Câmara, I.G., 2003. The Atlantic Forest of South America: biodiversity status, threats, and outlook. Island Press, Washington.

Godinho, R., Crespo, E.G., Ferrand, N., 2008: The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. Mol. Ecol. 17, 4670-4683.

Graham, C.H., Moritz, C., Williams, S.E., 2006. Habitat history improves prediction of biodiversity in rainforest fauna. Proc. Natl. Acad. Sci. USA 103, 632–636.

Hare, M.P., 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. 16, 700-706.

Jombart, T., Collins, C., 2015. An introduction to adegenet 2.0. 0. Imp Coll London-MRC Cent Outbreak Anal Model, 43.

Kamm, J.A., Terhorst, J., Durbin, R., Song, Y.S., 2019. Efficiently inferring the demographic history of many populations with allele count data. J. Am. Stat. Assoc. 1-42.

Lynch, M., 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from 529 high-coverage genome-sequencing projects. Mol. Biol. Evol. 25, 2409–2419.

Maldonado-Coelho, M., 2012. Climatic oscillations shape the phylogeographical structure of Atlantic Forest fire-eye antbirds (Aves: Thamnophilidae). Biol. J. Linn. Soc. 105, 900–924.

Mardis, E.R., 2017. DNA sequencing technologies: 2006-2016. Nat. Protoc. 12, 213-218.

Martins, F.M., 2011. Historical biogeography of the Brazilian Atlantic forest and the Carnaval–Moritz model of Pleistocene refugia: what do phylogeographical studies tell us? Biol. J. Linn. Soc. 104, 499-509.

Martins, F.M., Templeton, A.R., Pavan, A.C.O., Kohlbach, B.C., Morgante, J.S., 2009. Phylogeography of the common vampire bat (*Desmodus rotundus*): marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. BMC Evol. Biol. 9, 294.

McCormack, J.E., Maley, J.M., Hird, S.M., Derryberry, E.P., Graves, G.R., Brumfield, R.T., 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol. Phylogenet. Evol. 62, 397-406.

Moran, P.A.P., 1962. The Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

Moritz, C., 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst. Biol. 51, 238-254.

Moritz, C., Patton, J.L., Schneider, C.J., Smith, T.B., 2000. Diversification of rainforestfaunas: an integrated molecular approach. Annu. Rev. Ecol. Syst. 31, 533–563.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B.D., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853-858.

Nabholz, B., Glemin, S., Galtier, N., 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evol. Biol. 9, 54.

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., Ellegren, H., 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr. Biol. 25,1375-1380.

Ng, N.S., Wilton, P.R., Prawiradilaga, D.M., Tay, Y.C., Indrawan, M., Garg, K.M., Rheindt, F.E., 2017. The effects of Pleistocene climate change on biotic differentiation in a montane songbird clade from Wallacea. Mol. Phylogenet. Evol. 114, 353-366.

Olson, S.L., 1975. Geographic variation and other notes on Basileuterus leucoblepharus (Parulidae). Bull. B. O. C. 95, 101-104.

Paglia, A.P., 2005. Panorama geral da fauna ameaçada de extinção no Brasil. In: Machado, A.B.M., Soares Martins, C., Druond, G.M. (Eds.), Lista da Fauna Brasileira

Ameaçada de Extinção – Incluindo a Lista das Quase Ameaçadas e Deficientes em Dados. Fundação Biodiversitas, Belo Horizonte, pp. 17-22.

Pellegrino, K.C.M., Rodrigues, M.T., Waite, A.N., Morando, M., Yassuda, Y.Y., Sites, J.W., 2005. Phylogeography and species limits in the *Gymnodactylus darwinii* complex (Gekkonidae, Squamata): genetic structure coincides with river system in the Brazilian Atlantic Forest. Biol. J. Linn. Soc. 85, 13–26.

Petkova, D., Novembre, J., Stephens, M., 2015. Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. 48,94–100.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

QGIS Development Team, 2012. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Raposo do Amaral, F., Maldonado-Coelho, M., Aleixo, A., Luna, L.W., Rêgo, P.S.D., Araripe, J., Souza T.O., Silva, W.A.G., Thom, G., 2018. Recent chapters of Neotropical history overlooked in phylogeography: shallow divergence explains phenotype and genotype uncoupling in *Antilophia manakins*. Mol. Ecol. 27, 4108-4120.

Ribeiro, C.A., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biol. Conserv. 142, 1141–1153.

Ribeiro, R.A., Lemos-Filho, J.P., Ramos, A.C.S., Lovato, M.B., 2010. Phylogeography of the endangered rosewood *Dalbergia nigra* (Fabaceae): insights into the evolutionary history and conservation of the Brazilian Atlantic forest. Heredity 106, 46–57.

Sant'Anna-Neto, J.L., Nery, J.T., 2005. Variabilidade e mudanças climáticas no Brasil e seus impactos. In: Souza, G.R.G., Suguio, K., Oliveira, M.A.S., Oliveira, P.L. (Eds.), Quaternário do Brasil. Holos Editora, Ribeirão Preto, pp. 28-50.

Şekercioğlu, C.H., Primack, R.B., Wormworth, J., 2012. The effects of climate change on tropical birds. Biol. Conserv. 148, 1-18.

Sick, H., 1997. Ornitologia Brasileira. Nova Fronteira, Rio de Janeiro.

Stotz, D.F., Fitzpatrick, J.W., Parker, T.A., Moskovits, D.K., 1996. Neotropical Birds: Ecology and Conservation. University of Chicago Press, Chicago.

Ströher, P.R., Meyer, A.L., Zarza, E., Tsai, W.L., McCormack, J.E., Pie, M.R., 2019. Phylogeography of ants from the Brazilian Atlantic Forest. Org. Divers. Evol. 10, 1-11.

Tabarelli, M., Vicente, A., Barbosa, D.C.A., 2003. Variation of seed dispersal spectrum of woody plants across a rainfall gradient in north-eastern Brazil. J. Arid. Environ. 53, 197-210.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453-464.

Thomé, M.T.C., Zamudio, K.R., Giovanelli, J.G.R., Haddad, F.B., Baldissera, F.A.Jr., Alexandrino, J., 2010. Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. Mol. Phylogenet. Evol. 55, 1018–1031.

Uezu, A., Metzger, J.P., Vielliard, J.M.E., 2005. Effects of structural and functional connectivity and patch size on the abundance of seven Atlantic Forest bird species. Biol. Conserv. 123, 507-519.

Wang, I. J., Bradburd, G. S. 2014. Isolation by environment. Mol. Ecol. 23, 5649-5662.

Wright, S., 1943. Isolation by distance. Genetics 28, 114-138.

Table 1. Sampling localities of *Myiothypis leucoblephara leucoblephara* and of *M. l. lemurum* (YPM ORN100955). # - Locality number, Long - Longitude, Lat - Latitude. Samples - collections where samples are deposited: LGEMA- Laboratório de Genética e Evolução Molecular de Aves, Universidade de São Paulo; MNT- Museu Nacional, Universidade Federal do Rio de Janeiro; MACN- Museo Argentino de Ciencias Naturales Bernardino Rivadavia; MCT- Museu de Ciência e Tecnologia, Pontifícia Universidade Católica do Rio Grande do Sul; YPMORN- Yale Peabody Museum of Natural History, Yale University. In bold are the 27 samples shared between the present study and Batalha Filho *et al*. (2012).

#	Locality	State/Country	Long	Lat	Samples
1	Beira da Serra Serra Azul; Mateus Leme	MG, Brazil	-44.43	-20.13	**LGEMA12020; LGEMA12017**
2	RPPN Serra do Caraça	MG, Brazil	-43.83	-20.17	LGEMA15761; LGEMA15774
3	Serra da Moeda, Várzea do Lopes; Itabirito	MG, Brazil	-43.93	-20.28	MNT1485
4	Poços de Caldas	MG, Brazil	-46.54	-21.77	LGEMA13431
5	Resende	RJ, Brazil	-44.76	-22.41	**LGEMA12024; LGEMA12014;LGEMA12023; LGEMA12015**
6	PN Serra dos Órgãos; Teresópolis	RJ, Brazil	-43.01	-22.46	**LGEMA13006; LGEMA13000; LGEMA13019**
7	Cunha	SP, Brazil	-44.94	-23.12	**LGEMA12007**
8	Cunha	SP, Brazil	-44.85	-23.17	**LGEMA12006**
9	Cunha	SP, Brazil	-45.03	-23.24	**LGEMA12003**
10	Morro Grande; Cotia	SP, Brazil	-45.95	-23.69	LGEMA879; LGEMA883; LGEMA2954; LGEMA3029
11	Parque Estadual Morro Grande	SP, Brazil	-46.97	-23.71	LGEMA2931;LGEMA2930; LGEMA2925
12	Parque Estadual da Serra do Mar, Núcleo Curucutu; Juquitiba	SP, Brazil	-46.93	-24.01	LGEMA15229
13	Buri	SP, Brazil	-48.57	-23.72	**LGEMA2962; LGEMA2966**
14	Wencenslau Braz	PR, Brazil	-49.79	-23.93	**LGEMA1388; LGEMA1384**
15	Pinhalão	PR, Brazil	-50.05	-23.97	**LGEMA1296**
16	Ortigueira	PR, Brazil	-50.92	-24.2	**LGEMA11408; LGEMA11412**
17	Oberá; Provincia de Misiones	Argentina	-54.94	-27.44	MACN6924; MACN6947
18	Secc. Apepú Parque Nacional Iguazú; Misiones	Argentina	-54.3	-25.56	MACN6838; MACN3562
19	Parque Uruguaí, Uruzú; Misiones	Argentina	-54.17	-25.68	**LGEMA1409**
20	Paraje Maria Soledad Dpto Gral Belgrano; Misiones	Argentina	-53.98	-25.86	**LGEMA10345; LGEMA10358; LGEMA10364**
21	Parque Nacional Chaco; Provincia de Chaco	Argentina	-59.61	-26.81	MACN4841; MACN4792
22	Reserva Natural Educativa Colonia Benítez; Provincia de Chaco	Argentina	-58.95	-27.32	MACN4178
23	EBCO; Provincia de Corrientes	Argentina	-58.68	-27.55	MACN1697; MACN1585; MACN1663
24	Parque Nacional Mburucuya Provincia; Corrientes	Argentina	-58.02	-28.02	MACN7735; MACN5085
25	Rancho Queimado	SC, Brazil	-49.09	-27.65	**LGEMA2164; LGEMA2153; LGEMA2150**
26	Morro Grande, Pousada Rancho Fundo; Nova Roma	SC, Brazil	-49.77	-28.71	LGEMA15065
27	Pró-mata; São Francisco de Paula	RS, Brazil	-50.13	-29.45	MCT577
28	São Francisco de Paula	RS, Brazil	-50.57	-29.45	MCT0783
29	Maquiné	RS, Brazil	-50.32	-29.52	LGEMA17995; LGEMA17996
30	BR116 Km 232; Novo Hamburgo	RS, Brazil	-51.14	-29.63	LGEMA17992; LGEMA17993
31	RS 239; Riozinho	RS, Brazil	-50.46	-29.65	LGEMA17994
32	Morro da Borússia; Ozório	RS, Brazil	-50.3	-29.85	LGEMA18000
33	Santa Cruz do Sul	RS, Brazil	-52.38	-29.71	LGEMA17986; LGEMA17990
34	BR293; Santa Maria	RS, Brazil	-53.77	-29.81	LGEMA17984
35	Eldorado do Sul	RS, Brazil	-51.3	-30.05	MCT1623
36	Barra do Ribeiro	RS, Brazil	-51.14	-30.43	MCT3126
37	BR472; Uruguaiana	RS, Brazil	-57.12	-29.88	LGEMA17950
38	Parque Estadual do Espinilho; Barra do Quaraí	RS, Brazil	-57.49	-30.19	MCT4833
39	BR293; Quarai	RS, Brazil	-56.3	-30.45	LGEMA17956
40	BR 293; Caçapava do Sul	RS, Brazil	-53.46	-30.51	LGEMA17981
41	Arroio areião; Santana da Boa vista	RS, Brazil	-53.11	-30.91	LGEMA17980
42	Dom Pedrito	RS, Brazil	-54.73	-30.94	LGEMA17965
43	Dom Pedrito	RS, Brazil	-54.51	-31.06	LGEMA17963; LGEMA17962
44	Pelotas, Santana da Boa Vista	RS, Brazil	-52.84	-31.19	MCT3266
45	Arroio do Padre, Colonia Cerrito; Pelotas	RS, Brazil	-52.38	-31.37	**LGEMA2138**
46	Bagé	RS, Brazil	-54.1	-31.38	LGEMA17964
47	Dom Pedrito	RS, Brazil	-54.14	-31.52	LGEMA17971
48	Pinheiro Machado	RS, Brazil	-53.24	-31.62	LGEMA17975
49	Cascata; Pelotas	RS, Brazil	-52.47	-31.66	LGEMA17977; LGEMA17978
50	Arroio Bagé; Bagé	RS, Brazil	-54.16	-31.73	LGEMA17974
51	Arroio passos das pedras; Pelotas	RS, Brazil	-52.73	-31.74	LGEMA17976
52	Aceguá	RS, Brazil	-54.05	-31.88	LGEMA17966, LGEMA17970
53	Cerro Largo	Uruguai	-53.65	-32.31	YPMORN100955
54	Parque Nacional El Palmar; Provincia de Entre Ríos	Argentina	-58.24	-31.89	MACN4396
55	Santo Domingo, Arroyo Guaviyú; Departamento de Paysandú	Uruguai	-57.68	-31.80	LGEMA19553

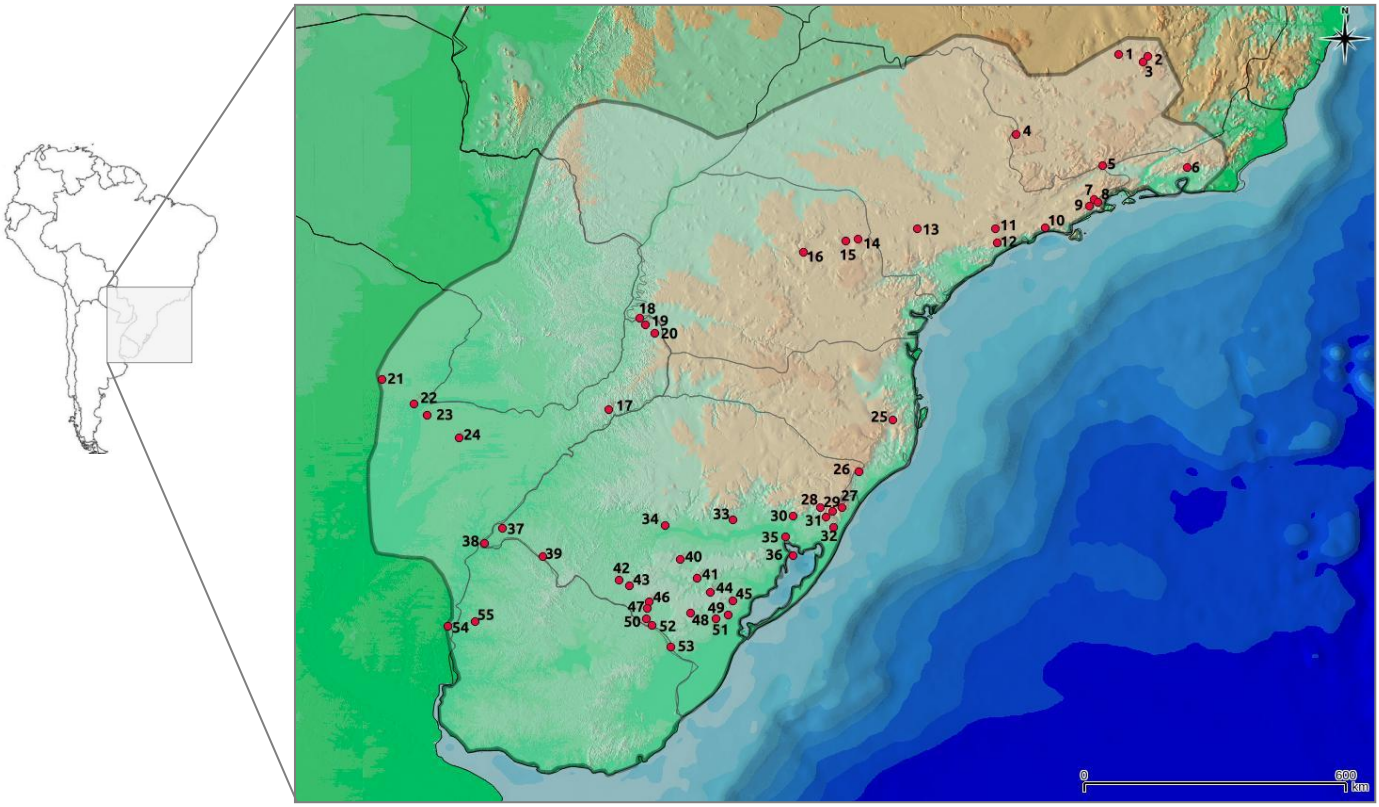


Figure 1. Map of South America highlighting the geographical distribution area of *Myiothlypis leucoblephara* (continuous line) and sampling locations (see Table 1).

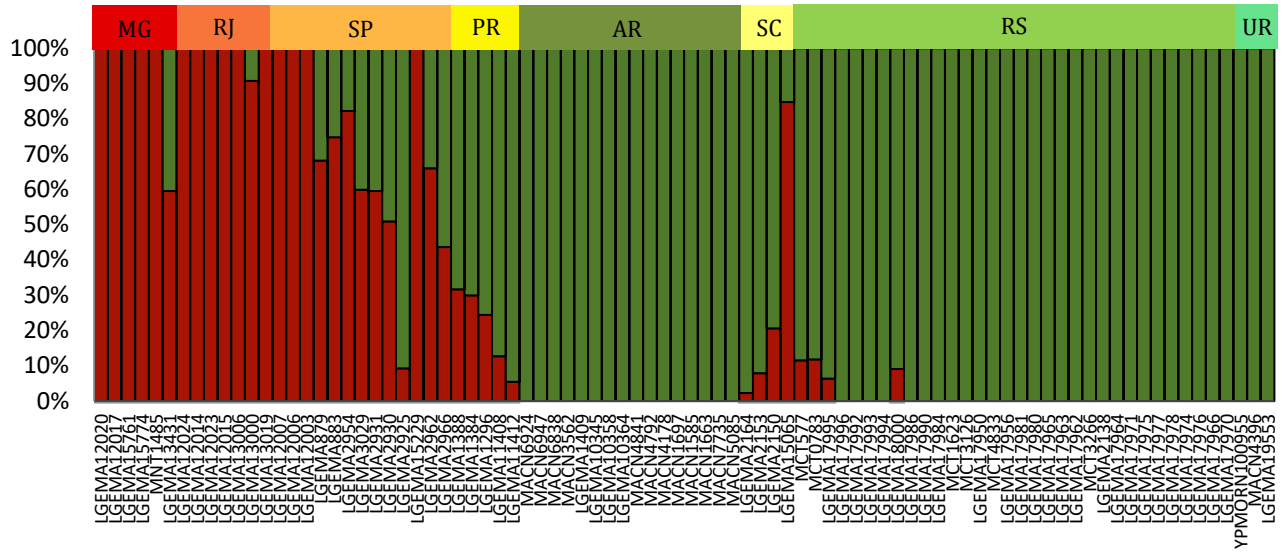


Figure 2. Results of sNMF analyses based on 8,707 SNPS of 86 individuals of *M. leucoblephara* adopting K = 2. The colors of the vertical bars represent the 2 genetic groups. The names below the graph correspond to samples (Table 1) arranged according to increasing latitude from left to right.

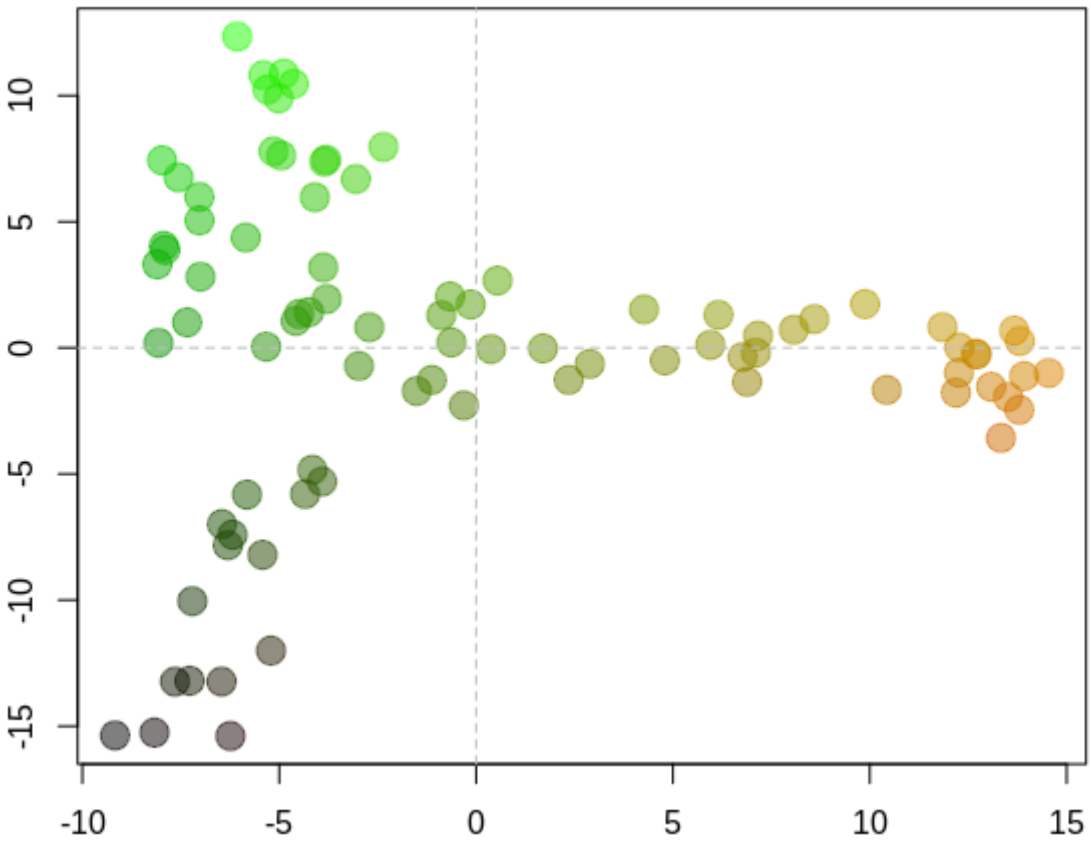


Figure 3. PCA results of 8,707 SNPs from 86 individuals of *M. leucoblephara* showing the first (X axis, north to south) and the second (Y axis, east to west) main components. Each circle represents one individual. Circles with similar colors correspond to genetically more similar individuals.

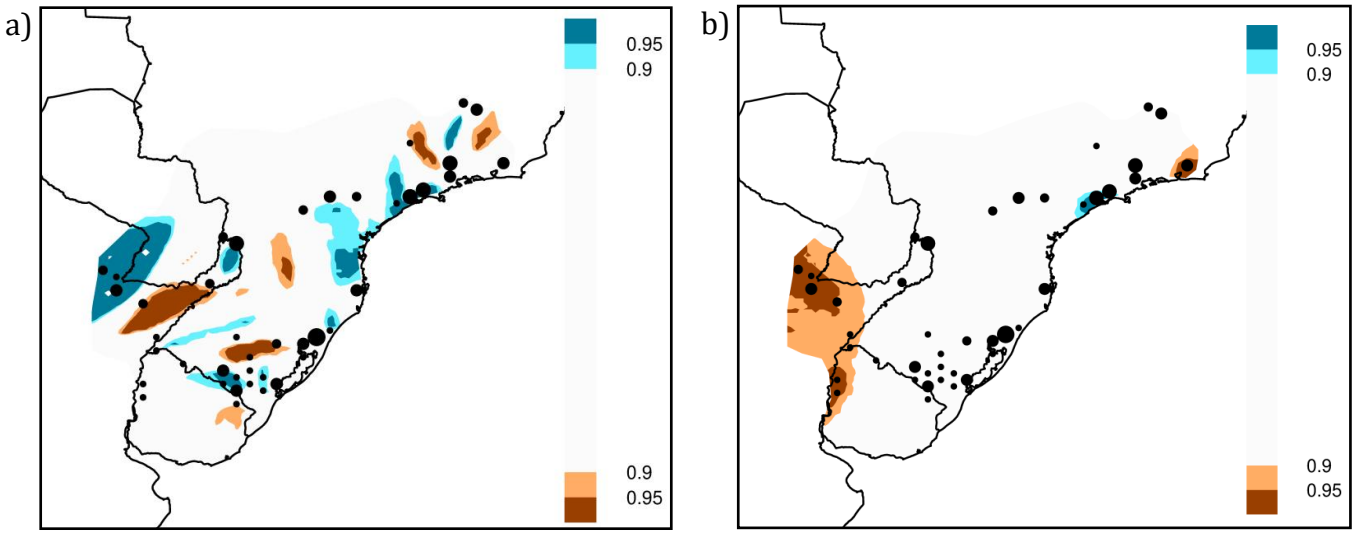


Figure 4. Spatial distribution of confidence interval of posterior distributions obtained for effective migration (a) and effective diversity (b) based on 7,675 SNPs from 86 individuals of *M. leucoblephara*. The colors represent the higher (blue) and lower (brown) than average rates.

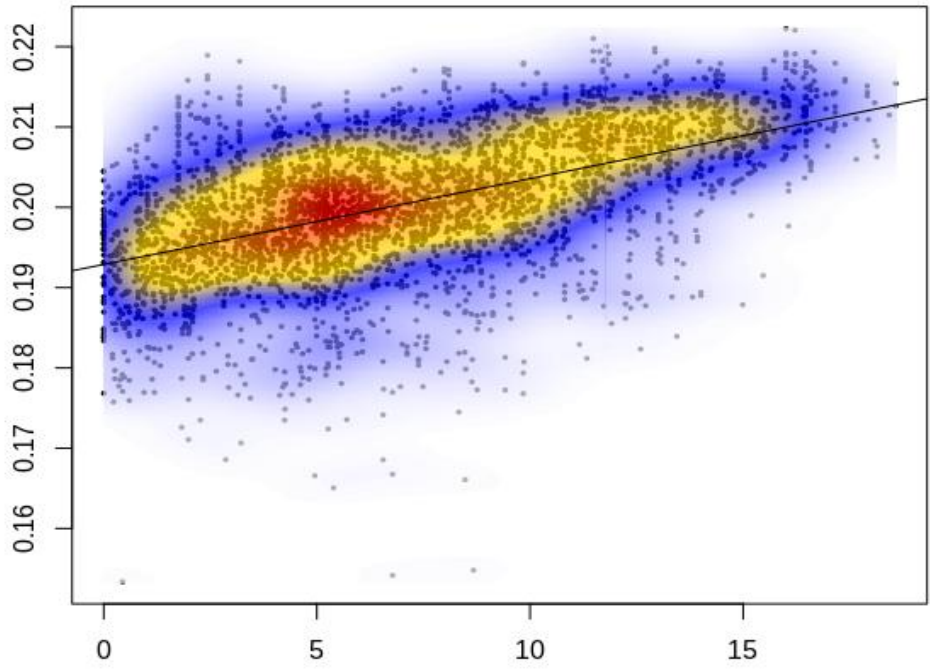


Figure 5. Plot of isolation by distance analysis showing the correlation between genetic distance (Y axis) and linear geographic distance (X axis, Euclidean distance) based on 8,707 SNPs from 86 individuals of *M. leucoblephara.*

Table S1. Results of the filtering of 86 individuals of *Myiothypis leucoblephara* dataset in ipyrad. Steps 2 to 5 are shown. **N reads-** Number of reads; **Reads Ns-** Reads discarded by number of Ns; **Reads min len-** Reads retained by minimum length of 75 pb; **Clusters total-** Total number of contigs; **Avg depth tot-** Average total depth; **Sd depth tot-** Total depth standard deviation; **Clusters hi depth-** Contigs with minimum depth of eight per site; **Avg depth-** Depth average using minimum of eight depth per site; **Sd depth-** Depth standard deviation using minimum of eight depth per site; **Hetero est-** Estimated heterozygosity per site; **Error est-** Estimated heterozygosity standard error; **Filt Max H-** Number of clusters removed with more than four heterozygous sites; **Filt Max N-** Clusters removed with more than five Ns per site; **Reads cons-** Total number of consensus of clustered reads; **N sites-** Size of all concatenated contigs; **N hetero-** Size of concatenated contigs of heterozygous sites.

	Step 2			Step 3						Step 4		Step 5				
Samples	N reads	Reads Ns	Reads min len	Clusters total	Avg depth tot	Sd depth tot	Clusters hi depth	Avg depth	Sd depth	Hetero est	Error est	Filt Max H	Filt Max N	Reads cons	N sites	N hetero
LGEMA10345	1686838	311	481135	208359	5.54	37.59	36458	21	88.16	0.010114	0.002255	1467	174	34816	3501141	17090
LGEMA10358	5471879	1818	1352365	388493	10.08	113.62	98246	32.37	224.44	0.008591	0.001968	3057	235	94944	9715394	50852
LGEMA10364	4398940	1297	1314728	306185	9.62	89.17	78622	30.27	174.31	0.008451	0.001684	2364	186	76067	7901346	39093
LGEMA11408	4790968	2351	1188538	325588	10.41	130.51	88987	31.42	248.40	0.008628	0.002174	2818	216	85946	9016201	46835
LGEMA11412	2294403	415	581122	235536	6.93	67.32	53195	22.76	140.48	0.009019	0.002142	1678	180	51334	5422037	26745
LGEMA12003	3678795	1344	1072586	283033	8.86	77.56	69569	28.61	154.74	0.008994	0.002059	2258	228	67078	6886105	34797
LGEMA12006	4899444	1039	1251005	342622	10.12	81.86	92303	30.8	155.82	0.009145	0.001896	2985	277	89038	9004830	48314
LGEMA12007	3189431	892	856104	269327	8.20	99.32	63309	27.09	203.68	0.009104	0.002033	2134	174	60997	6318185	31562
LGEMA12014	3606364	1362	1068095	280414	8.56	70.72	67583	28	142.28	0.008962	0.002084	2088	223	65268	6754811	34957
LGEMA12015	3864936	1165	1074387	310385	8.58	89.43	73141	28.47	182.78	0.009089	0.002075	2358	218	70560	7247949	38015
LGEMA12017	1985300	667	526086	227603	6.06	54.27	45815	21.1	119.74	0.009188	0.002265	1446	178	44190	4544451	22252
LGEMA12020	7339301	2511	2354342	357470	13.12	186.85	99616	40.72	352.45	0.009050	0.001796	3311	224	96068	9896932	53845
LGEMA12023	3499925	925	883235	297712	8.29	77.50	68877	27.81	159.54	0.008785	0.001909	2026	187	66658	6789041	35118
LGEMA12024	2544079	844	778508	235306	7.12	72.38	53100	23.9	151.13	0.009039	0.002116	1645	195	51257	5334178	26328
LGEMA13000	3869866	1184	1249479	291439	8.65	59.89	70702	28.22	119.46	0.008080	0.001866	1995	180	68524	7183583	33447
LGEMA13006	5306272	988	1801177	310667	10.72	91.70	82180	33.65	176.25	0.007863	0.001719	2469	187	79518	8287620	37320
LGEMA13019	4613246	2334	1448561	310342	9.61	73.25	76583	31.58	145.23	0.008403	0.001990	2375	190	74014	7697280	37606
LGEMA13431	1173344	229	534921	117592	5.07	51.13	18807	19.84	126.77	0.010111	0.002604	679	89	18038	1671268	9246
LGEMA15065	3720008	1487	1199513	273029	8.68	79.93	65586	28.84	161.40	0.008864	0.002082	2051	186	63346	6713069	34051
LGEMA15229	5267368	2021	1440687	330012	10.91	119.27	90210	33.32	226.59	0.008864	0.001816	2828	207	87169	9133499	48741
LGEMA15761	4254312	880	1249350	299882	9.36	101.19	73996	30.48	202.22	0.008921	0.001873	2400	215	71376	7447198	38466
LGEMA15774	3486106	686	1019054	265956	8.79	86.30	66305	28.07	171.37	0.008910	0.001919	2118	218	63965	6693950	33772
MACN1585	1588971	522	351903	228271	5.13	34.66	39114	19.03	82.25	0.008743	0.002346	1270	152	37692	3861863	17297
MACN1663	3329630	1132	912725	275860	8.24	73.39	65258	27.14	149.31	0.008359	0.002085	2064	176	63014	6517213	30083
MACN1697	3351143	1661	1090183	289286	7.45	50.90	60989	26.8	108.64	0.008890	0.002255	2110	238	58640	5964428	27119
LGEMA17956	4661456	1922	1623589	300573	9.69	70.25	74900	31.7	138.38	0.008096	0.001898	2238	196	72461	7601157	35254
LGEMA17962	3506935	1268	1216805	262543	8.30	73.14	62210	27.72	148.57	0.008518	0.001987	1858	172	60176	6327680	30037
LGEMA17963	2413762	550	776018	233772	6.73	45.44	51168	22.71	95.38	0.008714	0.001966	1579	173	49415	5119482	23967
LGEMA17965	4408428	831	1084608	325991	9.65	86.86	88485	28.79	165.17	0.008391	0.001768	2576	182	85723	8985834	45095
LGEMA17966	1342889	597	304123	206839	4.78	34.28	34548	17.2	82.68	0.008829	0.002656	1106	160	33282	3454071	15627
LGEMA17970	4109652	1900	1339553	302066	8.78	75.86	71291	29.49	154.32	0.008001	0.001994	2081	197	69007	7192679	32888
LGEMA17971	3250107	1350	1001234	264893	8.07	61.11	62301	26.65	124.16	0.008506	0.002056	1937	169	60192	6219726	29246
LGEMA17974	2405274	804	713827	241827	6.73	53.15	51500	23.17	113.62	0.008827	0.002121	1536	198	49766	5096681	24253
LGEMA17975	2877513	939	791063	264662	7.48	60.32	61979	24.29	123.13	0.008496	0.001935	1894	184	59899	6301923	29915
LGEMA17976	3782528	1456	1254150	278223	8.70	74.08	66787	28.84	149.39	0.008241	0.001977	1923	185	64675	6804762	32016
LGEMA17977	4350304	842	1031148	370590	8.58	83.28	87786	28.39	169.56	0.007995	0.001815	2575	180	85026	8798676	40998

	Step 2			Step 3						Step 4		Step 5				
Samples	N reads	Reads Ns	Reads min len	Clusters total	Avg depth tot	Sd depth tot	Clusters hi depth	Avg depth	Sd depth	Hetero est	Error est	Filt Max H	Filt Max N	Reads cons	N sites	N hetero
LGEMA17978	3490075	1118	1201502	273941	8.00	59.14	62191	27.31	122.13	0.008262	0.001964	1840	164	60186	6148080	28145
LGEMA17980	3983126	992	1269600	293629	8.91	76.77	71120	29.24	154.21	0.008760	0.001874	2234	205	68677	7203052	35508
LGEMA17981	3242294	1454	1086439	273810	7.55	47.30	62462	25.38	96.88	0.008524	0.001994	1781	191	60489	6318573	30558
LGEMA17984	6052726	1847	1841577	337910	11.89	93.82	97241	35.15	172.69	0.008386	0.001659	2817	229	94189	9706444	49289
LGEMA17986	7606340	2568	2477888	390737	12.40	119.33	108232	38.35	224.65	0.008696	0.001756	3200	236	104790	10727347	57114
LGEMA17990	1485758	262	387184	205465	5.10	34.35	37328	18.02	79.24	0.009271	0.002411	1153	157	36018	3677961	17924
LGEMA17992	6274683	1090	2170308	342027	11.43	106.52	91652	35.87	203.76	0.008736	0.001711	2807	195	88644	9192302	48397
LGEMA17993	2560738	505	867311	233734	6.91	48.89	50761	23.72	103.13	0.008480	0.001916	1397	150	49212	5054685	23999
LGEMA17994	6082290	1938	2090253	316708	11.86	96.01	88838	35.89	179.04	0.008686	0.001771	2617	173	86042	8683936	46065
LGEMA17995	2594040	1325	797280	255792	6.66	53.65	52671	23.88	116.60	0.008745	0.002230	1573	170	50926	5250539	25452
LGEMA17996	7319328	2337	2440557	359292	12.94	127.53	102472	39.15	236.75	0.008694	0.001831	3093	228	99143	10293447	55331
LGEMA18000	3295296	877	891753	282849	8.09	76.29	66640	26.62	155.70	0.008826	0.002011	2116	188	64331	6833159	34068
MACN3562	1257995	420	328440	186888	4.72	34.56	29873	17.82	85.17	0.009466	0.002538	1062	125	28686	2907503	14008
MACN4178	3466823	1068	1092155	302931	7.40	62.33	65103	25.65	132.82	0.008743	0.001919	2179	216	62705	6445847	30167
MACN4396	2290476	622	518822	253680	6.61	55.50	51913	23.15	121.24	0.008096	0.002187	1540	187	50186	5142677	22360
MACN4792	1943768	620	623693	219516	5.76	34.75	41562	20.66	78.06	0.009224	0.002194	1463	205	39894	4017210	18223
MACN4841	4460447	2407	1030079	327953	9.79	106.95	87857	29.73	205.29	0.008177	0.002127	2531	218	85102	8574244	41579
MACN5085	4083068	1182	1006347	308810	9.42	100.00	78326	29.84	197.12	0.008163	0.001901	2325	187	75808	7754495	36549
MACN6838	3662779	1329	819080	300244	8.77	90.14	74323	27.92	179.80	0.008523	0.002043	2263	182	71874	7284872	36143
MACN6924	4745516	1465	1345320	327545	9.83	103.86	82368	31.79	205.53	0.008438	0.001904	2457	168	79737	8246857	41471
MACN6947	3803559	1446	871468	308049	8.90	83.89	75454	28.8	167.93	0.008682	0.002006	2234	194	73022	7381965	37471
MACN7735	2977610	1229	698579	269584	7.95	66.46	63581	25.82	135.27	0.008316	0.002148	1963	165	61450	6291768	29285
YPMORN100955	3796092	757	934990	293725	9.17	82.30	77593	27.79	158.61	0.008148	0.001832	2343	198	75048	7923921	37711
MCT0783	775428	160	344786	99057	4.14	27.29	12061	18.3	76.61	0.010731	0.002795	506	74	11481	1062878	5660
MCT1623	4313853	1309	1608074	262153	9.75	132.06	63411	33.17	267.14	0.007988	0.001856	1862	171	61372	6461609	28970
MCT3126	3038981	1136	962143	251460	7.78	99.84	56843	26.74	208.85	0.008602	0.002107	1812	179	54846	5718500	26863
MCT3266	1883843	468	723474	184363	5.89	52.52	36143	21.26	117.32	0.009209	0.002293	1217	166	34759	3458054	16626
MCT4833	6839744	1529	2167942	369718	11.97	118.10	102027	36.86	222.88	0.008225	0.001690	2958	204	98858	10270635	51497
MCT577	1158852	409	277214	194426	4.31	35.50	25248	17.88	97.34	0.010082	0.003185	1037	145	24066	2466326	12069
LGEMA19553	2012241	946	650464	204668	6.32	54.92	42393	22.17	119.31	0.008318	0.002415	1309	154	40929	4173458	18037
LGEMA1409	3195070	1412	1044027	268796	7.65	53.61	61559	25.54	110.12	0.008811	0.002102	1993	218	59346	6176566	29440
LGEMA2138	6561555	2056	2018572	416059	10.34	116.44	100119	35.09	235.63	0.008412	0.001811	3089	262	96758	9886730	49866
LGEMA2150	7760894	2409	2220934	421983	12.42	143.53	118393	37.86	269.29	0.008312	0.001701	3570	209	114603	11805081	61060
LGEMA2153	6489438	2484	2130055	335622	12.16	135.84	93723	37.18	255.35	0.008756	0.001834	2751	212	90753	9192139	48910
LGEMA2164	4237611	1387	1164168	320912	9.12	94.56	78770	29.6	189.37	0.009179	0.002032	2731	248	75785	7741734	39619
LGEMA2925	2342564	768	523460	226159	7.92	27.39	54792	25.06	51.96	0.006855	0.001547	1239	104	53449	5484584	22146
LGEMA2930	1432581	526	448022	197028	4.75	39.81	29633	19.39	101.35	0.010152	0.002738	1143	153	28337	2837435	14605
LGEMA2931	3143463	631	892852	293443	7.25	73.15	62587	25.08	157.08	0.009975	0.002291	2431	281	59871	6098622	32111
LGEMA2954	7644189	2443	2398088	380336	12.97	174.24	105388	40.3	329.43	0.008785	0.001781	3347	233	101792	10347763	55695
LGEMA2962	6496903	3359	2241296	342707	11.79	132.07	91086	37.63	254.38	0.008883	0.001986	2954	205	87920	9111290	48744
LGEMA2966	2537441	1116	797435	245057	6.80	58.15	52981	23.4	123.61	0.008899	0.002172	1675	178	51127	5320315	25490
LGEMA3029	3312467	1012	976063	282494	7.93	70.79	66129	26.13	144.80	0.008861	0.001912	2091	193	63842	6727348	34082
LGEMA879	6902539	2053	2078777	473332	9.60	109.42	104565	34.77	231.04	0.009281	0.001747	3626	250	100679	10260367	56902
LGEMA883	5297066	1749	1411553	331698	11.13	108.87	92563	33.42	204.38	0.008379	0.001793	2730	181	89645	9377613	48800
LGEMA1296	4457119	743	1451458	300376	9.58	81.82	74809	31.14	162.02	0.008855	0.001672	2337	201	72265	7575101	39187

	Step 2			Step 3						Step 4		Step 5				
Samples	N reads	Reads Ns	Reads min len	Clusters total	Avg depth tot	Sd depth tot	Clusters hi depth	Avg depth	Sd depth	Hetero est	Error est	Filt Max H	Filt Max N	Reads cons	N sites	N hetero
LGEMA1384	7010362	3221	2406724	369126	11.90	97.07	98925	37.92	184.99	0.008587	0.001902	3003	199	95717	9970704	52566
LGEMA1388	1756950	669	528674	213969	5.47	32.77	42395	18.76	72.03	0.008924	0.002161	1267	160	40968	4238314	20206
LGEMA17950	1707879	322	379235	222310	5.68	42.93	43881	19.25	95.38	0.008364	0.002225	1343	170	42367	4377373	18889
LGEMA17964	4358805	826	1245016	301277	9.87	73.87	80610	30.01	140.82	0.009112	0.001852	2745	261	77599	7857531	39796
MNT1485	3431760	789	948306	277818	8.43	88.06	67984	27.06	176.70	0.008964	0.002003	2211	209	65560	6956098	35335

Table S2. Number of SNPs (1 SNP per locus) obtained from 86 samples of *Myiothlypis leucoblephara* according to the percentage of completeness of the matrix.

Matrix completeness	SNPs
25%	30030
50%	18566
85%	11209
90%	8707
95%	4902

Table S3. Lower cross-entropy K values (1 to 10) with four alpha values (100 replicates for each). In bold is the best value.

	Alpha 1	**Alpha 10**	**Alpha 100**	**Alpha 1000**
K=1	0.2451	0.2448	0.2437	0.2435
K=2	0.2418	0.2419	0.2406	**0.2404**
K=3	0.2428	0.2416	0.2423	0.2417
K=4	0.2470	0.2456	0.2469	0.2462
K=5	0.2511	0.2479	0.2492	0.2538
K=6	0.2534	0.2524	0.2514	0.2626
K=7	0.2566	0.2534	0.2548	0.2698
K=8	0.2603	0.2571	0.2615	0.2690
K=9	0.2631	0.2613	0.2632	0.2798
K=10	0.2681	0.2661	0.2691	0.2833

Table S4: Parameters obtained for each model for *M. leucoblephara* in Momi2.

Models	Ne Cur.	Ne Anc. in T1	Ne Anc. in T2	T1	T2	Growth Rate	Growth Rate in T1	AIC Value	Relative Weights
Model 1	1,500,000	-	-	-	-	-	-	48145.7	0.00E+000
Model 2	1,500,000	98,431.4	-	27,238.6	-	0.0001	-	46590.1	1.00E+000
Model 3	1,500,000	10,001.0	-	10,001.0	-	-0.01	-	47081.0	2.54E-107
Model 4	10,001.6	10,001.3	10,000.2	10,002.9	10,003.7	0.002	-0.007	51339.1	0.00E+000
Submodel I	1,500,000	98,431.4	-	27,238.6	-	0.0001	-	46590.1	1.00E+000
Submodel II	696,265.3	96,166.3	-	50,001.1	-	0.0001	-	47006.0	4.98E-091
Submodel III	190,700.8	10,000.3	-	100,000.7	-	0.0001	-	47793.5	4.89E-262
Submodel IV	190,700.8	10,001.9	-	300,000.2	-	0.0001	-	47793.5	4.89E-262

Ne cur. - Effective population size of diploid individuals of the current population, Ne Anc. - Effective population size of diploid individuals of the ancestral population, T1/T2 – date expressed in years ago, AIC - Akaike information criterion, Relative Weights - relative contribution for each model,(-) not tested in the corresponding model..

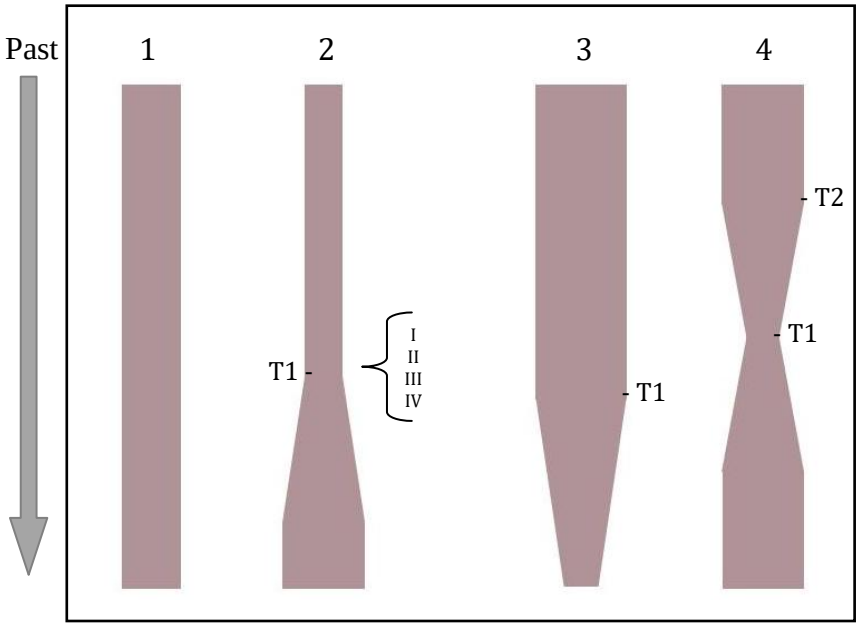


Figure S1. Historical demographic models tested for *M. leucoblephara* using SFS by simulations in Momi2. For all models (indicated by 1 to 4) we simulated populations that faced two types of historical events: population expansions (models 2 and 4) and bottlenecks (models 3 and 4). Model 2 was subsequently tested with four prior time intervals (I, II, III, IV) expressed in thousand years – kya: model 2I: 10 -50 kya; model 2II: 50-100 kya; model 2III: 100-300 kya; and model 2IV: 300-500 kya.

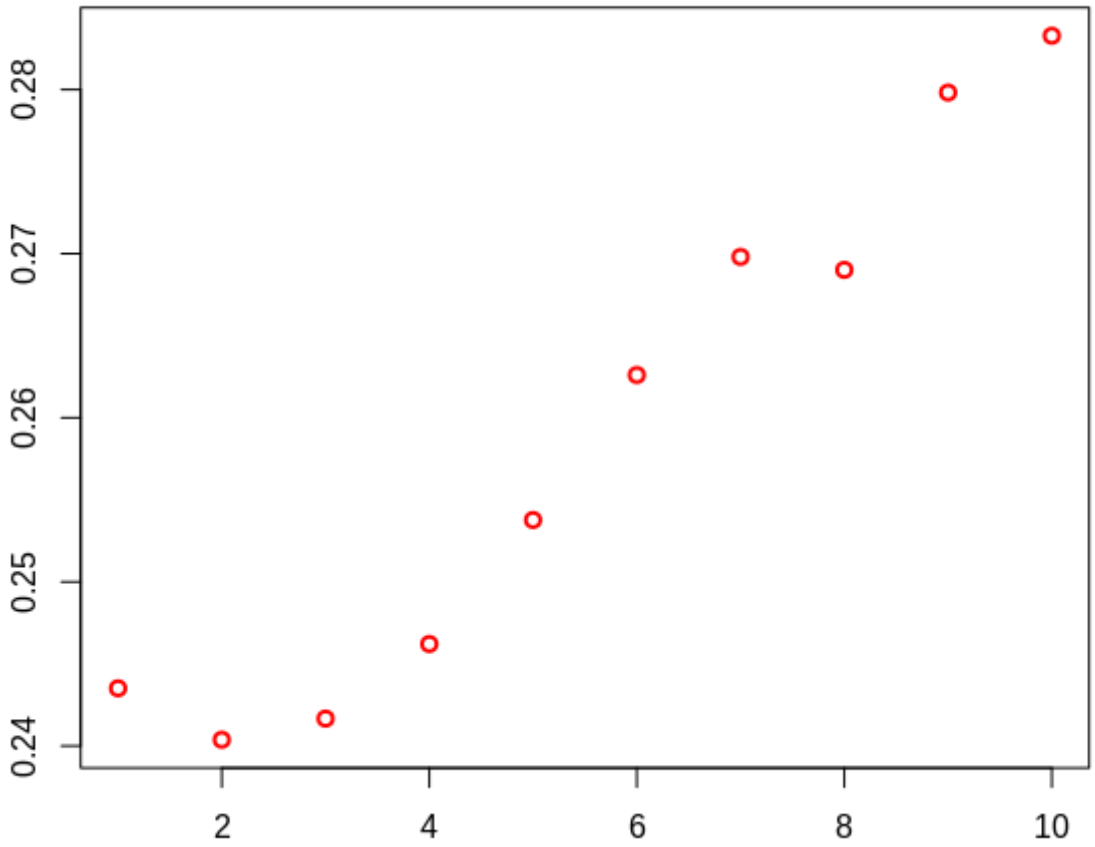


Figure S2. Cross-entropy values (alpha 1000) for various K (1 to 10) based on 100 replicates per K.

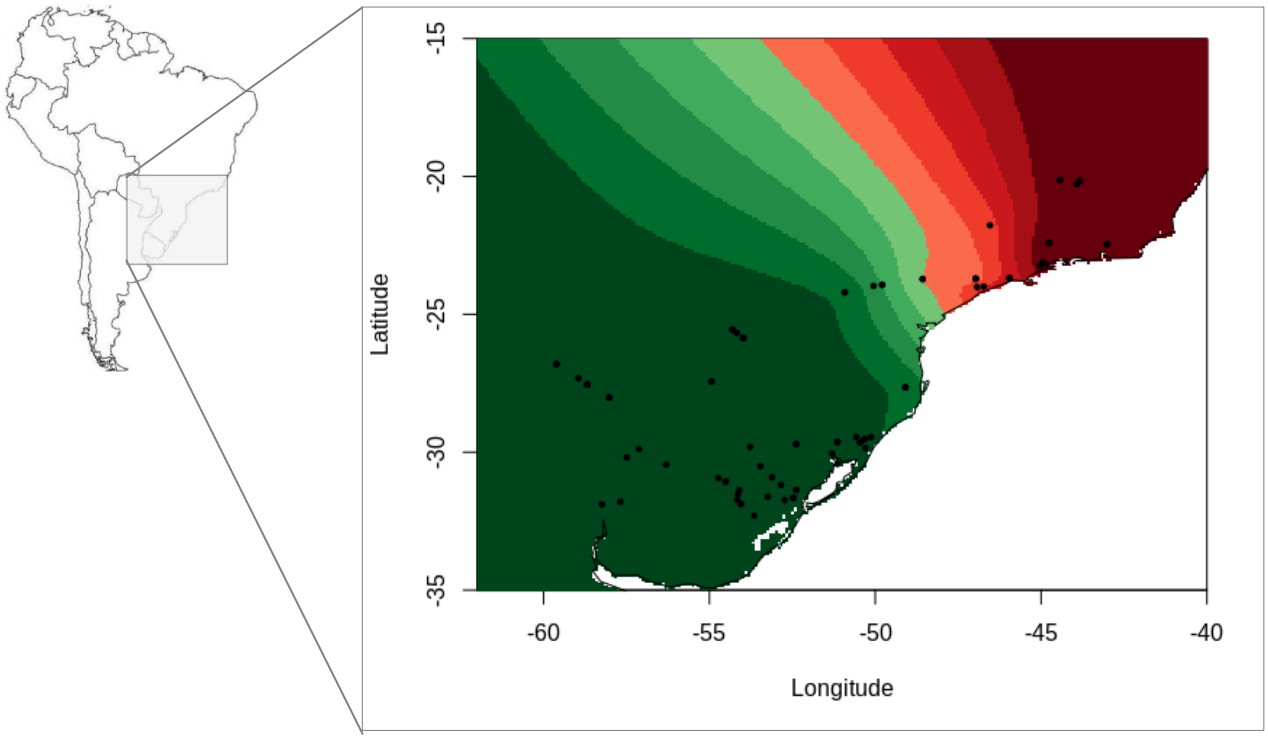
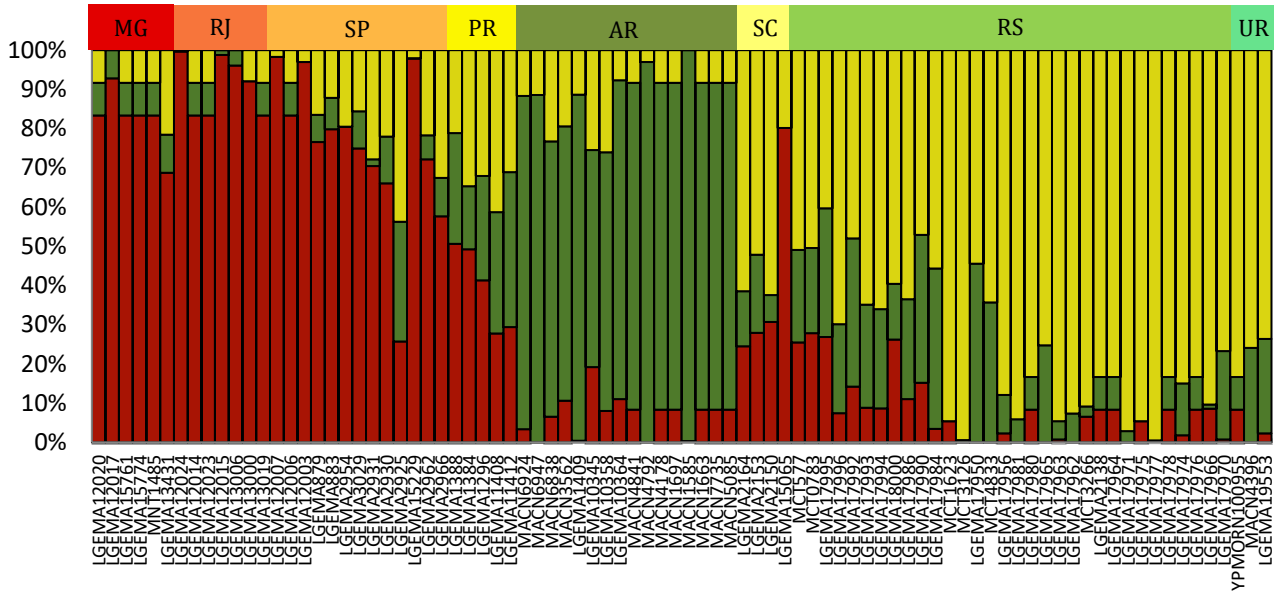


Figure S3. Geographic distribution of *M. leucoblephara* sNMF results for K = 2. Each circle represents the sampling location of an individual. The colors represent the 2 genetic pools (see Fig. 2). The lighter the color the greater is the mixture level.

(a)



(b)

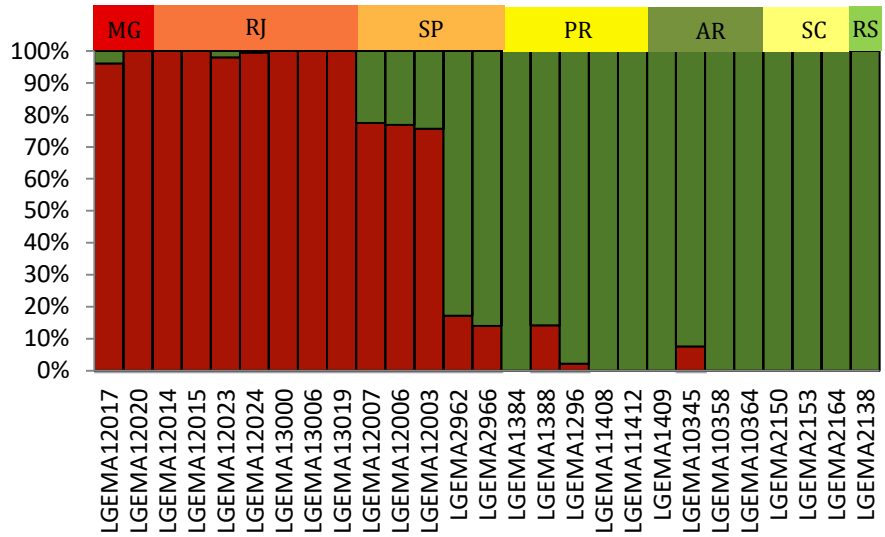


Figure S4. Results of sNMF analyses. a) Based on 8,707 SNPs of 86 individuals of *M. leucoblephara* adopting K = 3, b) Based on 10,693 SNPs of 27 individuals of *M. leucoblephara* adopting K = 2. The colors of the vertical bars represent three genetic groups. The names below the graph correspond to samples (Table 1) arranged according to increasing latitude from left to right.

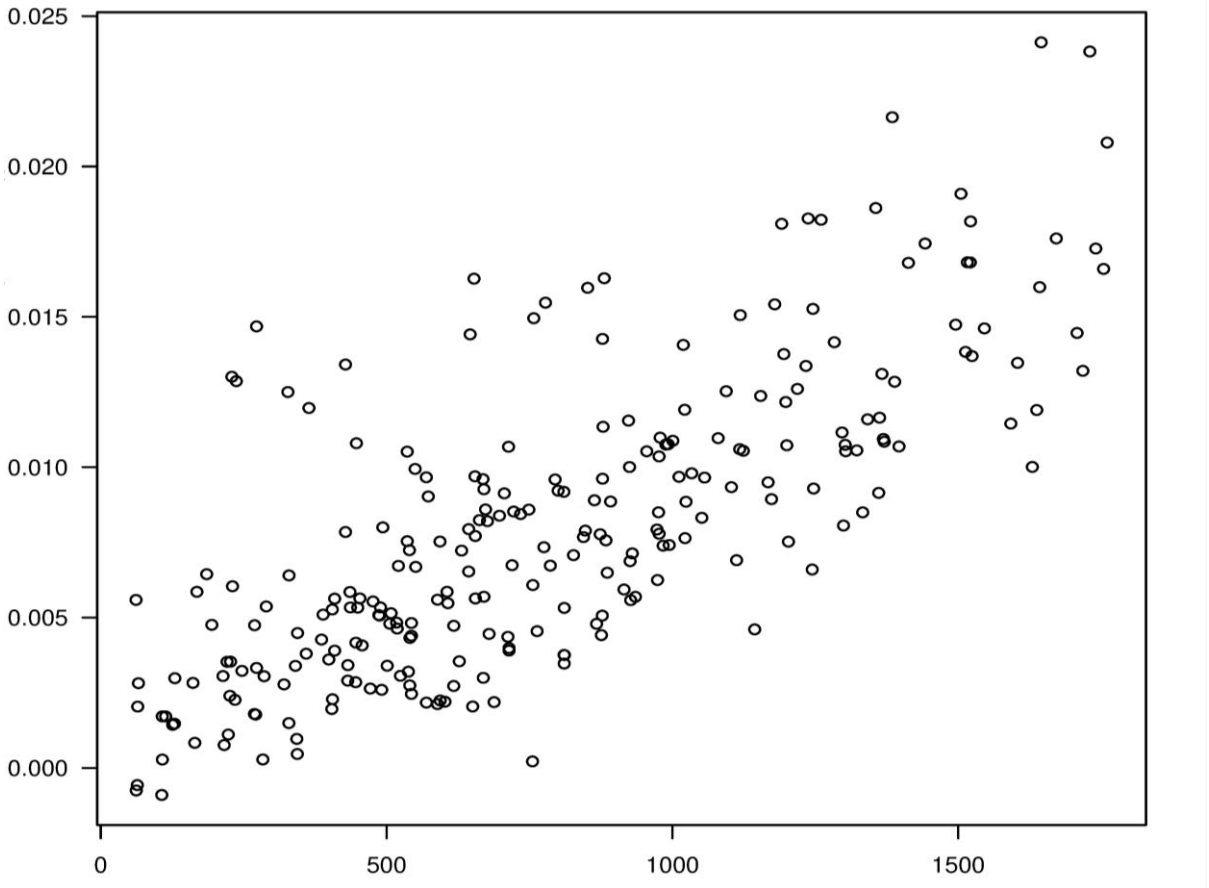
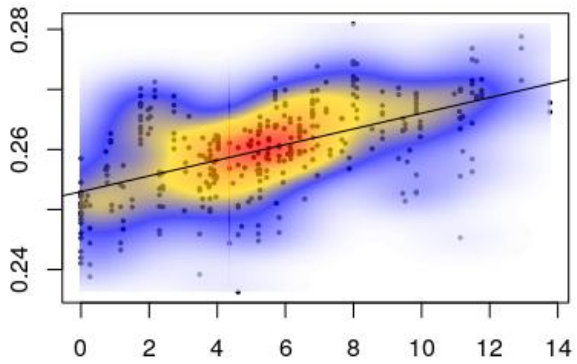


Figure S5. Observed dissimilarity between individuals computed between demes (Y axis) in relation to the geographic distance (X axis) based on 7,675 SNPs of 86 individuals of *M. leucoblephara.*

(a)



(b)

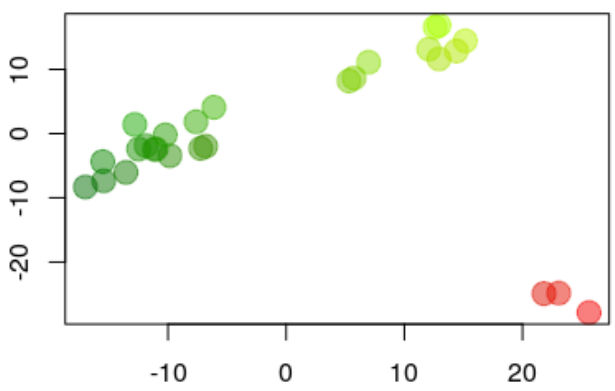


Figure S6. Results based on 10,693 SNPs of 27 individuals of *M. leucoblephara*. a) Plot of isolation by distance analysis showing the correlation between genetic distance (Y axis) and linear geographic distance (X axis, Euclidean distance), b) PCA plot showing the first (X axis) and the second (Y axis) main components.

Historical demographic inference of two Atlantic Forest antbirds, *Myrmoderus loricatus and Myrmoderus squamosus* (Aves: Thamnophilidae)

Introduction

The Neotropical region has been an area of high interest for biogeographers (Cracraft 1985). This region contains one of the world's most biodiverse ecosystems, the Atlantic Forest (AF). It contains 2.7% of the world's endemic vertebrates and 2.1% of plants (Myers *et al.* 2000). Unfortunately more than 60% of AF species are listed as nationally threatened (Tabarelli *et al.* 2003; Paglia 2005). Regarding avian biodiversity the AF harbors around 620 species with 199 endemics (Stotz *et al.* 1996, Myers *et al.* 2000). The AF originally covered about 1.477.500 km^2 along the coastal area of eastern Brazil to eastern Paraguay and northeastern Argentina (Myers *et al*. 2000; Galindo-Leal & Câmara 2003). Currently less than 16% of the original area of the AF remains (Ribeiro *et al.* 2009). These high biodiversity and degradation levels lead the AF to be one of the top five conservation priority areas worldwide (Myers *et al.* 2000).

The current limits of the AF seem to have been established in the Pliocene due to global cooling, which lead to the expansion of the savanna-like and dry vegetation South American dry diagonal (Caatinga, Cerrado, and Chaco biomes), separating the AF from the other South American humid forests (Morley 2000). This isolation in conjunction with a relatively wide range of latitude, longitude, and altitude, seem to have contributed to the high heterogeneity of this biome (Silva *et al.* 2004). For example, average humidity, soil depth, air temperature change along the landscape, resulting in various phytophysionomies (Galindo-Leal & Câmara 2003). This results in various diverse macro and microhabitats in the AF (Myers *et al.* 2000).

Several hypotheses have been raised to explain the origin of diversity in subtropical and tropical forests (Moritz *et al.* 2000) and many of them focus in the Quaternary (Whitmore & Prance 1987). Predictions associated to potential geographic barriers, such as rivers (Wallace 1852; Pellegrino *et al.* 2005) and mountains (Silva & Straube 1996), are the presence of discontinuities of distribution of organisms close or at the barrier, absence of demographic change, and the estimated date of divergence of sister taxa separated by the barrier should match the date of origin of the barrier. Based on paleoclimatic data the refuge hypothesis (Haffer 1969; Vanzolini & Williams 1970) is also mentioned in studies of the AF biota (e.g. Cabanne *et al.* 2007, Costa *et al.* 2000, Carnaval *et al.* 2009, D'Horta *et al.* 2011, Batalha-Filho *et al.* 2012). It suggests that during glacial maxima, especially in the late Pleistocene, the rainforest possibly contracted originating refuges isolated by open areas, and this could have isolated populations that could have originated differentiated lineages. When the global

temperature increased during interglacial periods, the vegetation once restricted in refuges may have expanded and these lineages got into contact but depending on the level of differentiation, they would not present any gene flow (Brown & Ab'Saber 1979; Graham *et al.* 2006; Sant'Anna-Neto & Nery 2005; Carnaval & Moritz 2008).

To increase our understanding of AF historical dynamics, in the present study we used as models two AF endemic birds, *Myrmoderus loricatus* (Lichtenstein 1823) and *Myrmoderus squamosus* (Pelzeln 1868) also called as white-bibbed antbird and squamate antbird, respectively. They are sister species (Gómez *et al.* 2010) and belong to the family Thamnophilidae (order Passeriformes), one of the most numerically families of the Neotropical avifauna (Skutch 1996). They have parapatric distribution, with overlapping areas in the central AF region (Fig. 1, state of Rio de Janeiro) and no clear geographical and ecological boundaries separate their distributions (Ridgely & Tudor 2009). They forage close to the ground searching for arthropods, occur from sea level up to 1200 m or more, and the couples from both species apparently do not join mixed flocks, form long-term social pairs, and intensify territorial defense during breeding season (Ridgely & Tudor 2009). In the past *M. squamosus* was considered as a subspecies of *M. loricatus* (Sick 1997) given the similar male plumage, as well as habits and sounds.

Here, we present one of the few phylogeographic studies of AF taxa using next generation sequencing data (NGS), aiming to investigate the evolutionary history of these two bird species and to contribute to the understanding of the processes underlying the origin of the AF biodiversity. From the phylogeographic perspective (Edwards *et al.* 2015), this approach is proving to be exceptionally useful, being capable of producing millions of DNA sequences in a single run and rapidly changing the landscape of genetics (Mardis 2017), resulting in more detailed information than that based on individual markers (Davey & Blaxter 2010). Here, based on a subgenomic dataset, we explored the demographic history, the degree of structure, and tested the influence of the Quaternary events in the evolutionary history of *M. loricatus* and *M. squamosus*.

Material and Methods

Samples

We obtained 35 individuals of *M. loricatus* and 30 of *M. squamosus* from 28 locations (14 per species) within the Brazilian AF (Fig. 1, Table 1). Tissue samples (muscle and

blood) were preserved in absolute ethanol and stored at -20ºC. DNA was extracted with the DNeasy kit (Qiagen Inc.). DNA was quantified using Qubit (Thermo Fisher Scientific) and then concentrated to ~100 ng/µl. DNA library construction (Genotyping-by-sequencing digested with *Pst*I; Elshire *et al.* 2011) and sequencing (single-end in HiSeq 2500, Illumina) were outsourced at Ecomol consultoria, ESALQ/USP.

Data processing

Among the challenges faced in conducting a study with NGS, knowledge of the theoretical basis behind the methods and pipelines is required for making the appropriate choices of parameters (Fonseca *et al.* 2016). It is also necessary to have more than basic bioinformatics background (Nielsen *et al.* 2010). Many successful tools are carefully tailored to the specialized analysis demands of their users (Nielsen *et al.* 2010), as those described below. We initially used FastQC (Andrews 2010) to check the general quality of the reads. Then we used ipyrad (Eaton & Overcast 2016), which is a set of scripts intended to process reads of libraries constructed using restriction enzyme. This processing consists of seven steps and as the choice of parameters dramatically affects the resulting datasets (Eaton & Overcast 2016), different parameter values were tested in six independent ipyrad runs to evaluate our *Myrmoderus* spp. data. The final parameter values in bold were the selected ones for the analyses of reads from both species and were also applied for the analyses of each species separately. Sites with Phred values below **33** were transformed into Ns and reads with more than **four** Ns were discarded (see Table S1). Reads were clustered together with minimum **85%** of similarity between them and a strict filter was used for adapters. We discarded reads with more than **five** uncalled bases in consensus reads, more than **four** heterozygous sites (Hs), and depth per site lower than **eight**. We allowed a maximum of **50%** shared polymorphic sites in a locus, maximum of **five** indels, **five** SNPs per final locus, and **two** alleles. We also checked two quality indicative flags for all samples in VCFtools (Danecek *et al.* 2011): "--indv-freq-burden" (number of variants [e.g. singletons, doubletons] for each individual in comparison to the mean frequency of variants of the total sample, this allows to detect individuals whose sequences may not have the same quality as the others and they were excluded) and "--singletons" (identifies the loci and individuals that present singletons). After the filtering process and quality checking using fastQC, ipyrad, and VCFtools two samples were excluded, because their final datasets presented low quality. Then ipyrad was rerun with the same parameters

described above with the 65 remaining samples (35 *M. loricatus* and 30 *M. squamosus*, Table 1 and Fig. 1). In VCFtools we excluded indels and randomly selected one SNP per locus to obtain a matrix of unlinked SNPs. We obtained SNPs matrices with various percentages of completeness of data (percentage of samples represented per matrix, Table S2).

Data analyses

The analyses of population genetic structure followed those described in Chapter 1. Briefly, we inferred the population structure, the number of distinct populations, and assigned individuals to these populations using sNMF in the R package for Landscape and Ecological Associations studies (Frichot & François, 2015). Additionally a Principal Component Analysis (PCA) was performed using the 90% complete SNP matrix in ADEGENET 2.0 (Jombart & Collins, 2015). We estimated the effective migration surface (EEMS) and tested for isolation by distance (IBD) with 999 replications in ADEGENET 2.0.

Demographic inference

The evolutionary history of the species under a demographic approach was inferred using Momi2 (MOran Models for Inference 2, Kamm et al., 2019, see more details in Chapter 1). We simulated different demographic scenarios/models (Fig. S1) based on previous findings for the species (Amaral *et al.* 2013) and also our results. Amaral *et al.* (2013) observed three lineages: *M. squamosus*, *M. loricatus*, and *M. loricatus* from the Chapada Diamantina (*M. loricatus* CD), which is an upland area west to the Bahia refuge (see Carnaval & Moritz 2008). So our **Model 1** considered the presence of these three lineages with localities 1 and 2 (Fig. 1 and Table 1) corresponding to *M. loricatus* CD. Additionally, this Model 1 considered that the lineages presented constant size to estimate the divergence time among them. As our results suggested the presence of only two lineages that correspond to the two species, in the other models we considered two lineages. **Model 2** included migration (gene flow) from *M. squamosus* to *M. loricatus*. While **Model 3** tested migration in the opposite direction (from *M. loricatus* to *M. squamosus*). **Model 4** tested exponential population expansion. While **Model 5** considered exponential population contraction and **Model 6**, population with constant size. We set uniform priors for: current effective population size (Ne; 1), size of instantaneous population size (upper 1,500,000/lower 10,000), growth rate (model 4:

upper 0.01/lower 0.0001, model 5: upper -0.0001), pulse rate (model 3 and 4: upper 20%) and divergence time (model 1: 10 thousand years ago [kya] to 5 million years ago [Mya]. For Models 2 to 6 we fixed the divergence time at 2.5 Mya to optimize the analysis. To estimate divergence times we assumed an average mutation rate of $2.5x10^{-9}$ (Nadachowska-Brzyska *et al.* 2015) and we assumed one year as the generation time (Cabanne *et al.* 2008). We performed 1000 runs with 2000 interactions for model 1, 2, 4, 5 and 6 and 100 runs with 2000 interactions for model 3 and 4, due the computer processing capacity. In order to select the best-fit model to the observed data we calculated the Akaike information criterion (AIC) between the pairs of scenarios: Models 2, 3 and 6; Models 4, 5 and 6. The model with the lowest AIC score was considered as the best one. To obtain 95% confidence intervals for the parameters of the best-fit model we performed 100 parametric bootstraps (Kamm *et al.* 2019).

Results

The final data processing produced raw VCF files with 178,318 SNPs for the 65 individuals. The separate analyses resulted in 115,410 SNPs for *M. loricatus* (35 individuals) and 48,897 for *M. squamosus* (30 individuals). The average number of consensus of clustered reads (see "Reads cons", Table S1) was 64,645.78; the average size of concatenated contigs (see "N sites", Table S1) was 6,629,786.26 bp, and the average site depth was 28.33 (see "Avg depth", Table S1). After randomly selecting one SNP per loci we obtained 11,961 SNPs and 15,527 SNPs for matrices with 95% and 90% completeness of data for *Myrmoderus* spp., respectively; 14,413 SNPs and 17,294 SNPs for matrices with 95% and 90% data for *M. loricatus*, respectively; and 5,962 SNPs and 9,621 SNPs for matrices with 95% and 90% data for *M. squamosus*, respectively (Table S2). The population structure analyses in sNMF for *Myrmoderus* spp. presented the best value of cross-entropy (value = 0.2607) with alpha parameter of 100 (Fig. S1a, Table S3) with K= 2. In this result all *M. loricatus* individuals were represented in the red cluster (Fig. 2a) and the green cluster represented *M. squamosus* (Fig. 2a). The sNMF for *M. loricatus* alone showed K=2 with the best value of cross-entropy (value = 0.4764) with alpha parameter of 10 (Fig. S1b, Table S3). Individuals sampled in the north (BA) were associated with the red cluster (Fig. 2b), individuals sampled in the central region (MG) showed mixed ancestry, and individuals predominantly carrying the other genetic set (green, Fig. 2b) were sampled in the south of the species distribution (RJ), indicating a pattern of IBD. We also estimated the

sNMF for *M. squamosus* which presented one ancestral population (K= 1) in alpha parameter =10 and cross-entropy value = 0.5803 (Fig. S1c, Table S3), showing no genetic structure.

The results of the PCA of *Myrmodeurs* spp. (Fig. 3a) was congruent with those obtained by sNMF (Fig. 2a). Samples of the two species were separated from each other and indicated the same gradual change of the genetic diversity within each of them. The first component (X axis) showed a change of genetic diversity of *M. squamosus* from the north to the south (left to right). While the second component showed a more gradual change from the north to the south (from top to bottom) for *M. loricatus* indicating IBD (Fig. 3a). The PCA of each species separately (Figs. 3b, 3c) showed this gradual change in the genetic composition of individuals throughout the distribution with highest differences between the extremes. Also, we detected a horseshoe effect; an artifact in which the second axis is curved relative to the first axis (Podani & Miklos 2002), that is, an upside down U-shape curve (Fig. 3b.) and U-shaped curve (Fig. 3c) that indicate in our case that the genetic variability is distributed as a gradient. This is also an evidence of IBD pattern (Frichot et al. 2012). In *M. loricatus* the individuals from the northernmost localities 1 and 2 (Fig. 1, Table 1) were represented by the pronounced red (Fig. 3b), that is very different from the others indicating a larger genetic distance. The PCA of *M. squamosus* (Fig 2c) showed the northernmost localities 15 and 16 (Fig. 1, Table 1) in light green and those from the southernmost localities 27 and 28 (Fig. 1, Table 1) were in orange.

The results of EEMS (Fig. 4) based on 14,416 and 7,446 SNPs (singletons excluded) of *M. loricatus* and *M. squamosus*, respectively, indicated that the northern portion of both species' range presented the highest genetic variability (Fig. 4a, 4b). Interestingly for *M. loricatus*, the highest genetic variation (localities 3 to 5, Fig. 4a) is observed east to the Chapada Diamantina (localities 1 and 2, Fig. 4a), possibly caused by the contrasting genetic dissimilarity with the localities 1 and 2 (which in the PCA presented the largest genetic distance, Fig. 3b). This seems to be reinforced by the significantly lower migration rate between the Chapada Diamantina localities and these eastern localities (3 to 5) near to the cost (Fig. 4b). The spatial distribution of genetic variability in the effective migration maps showed no clear structure pattern associated with regions or localities, suggesting that gene flow is occurring along all the distribution area for both species (Fig. 4b, 4d). EEMS also indicated the presence of IBD pattern for both species, which is especially linear in *M. squamosus* (Fig S2b), as the dissimilarity between

individuals increases according to their geographic distance (Figs. S2a, S2b). The Mantel test for IBD showed a significant effect (*p-value* = 0.001) of geographic distance on the genetic distance for both species. Assuming that *r* values between 0.7 and 0.9 correspond to strong correlation and between 0.3 and 0.5 to weak correlation (Mukaka 2012), in *M. loricatus* a strong correlation was observed between the two matrices (*r* = 0.718; Fig. 5a). While a weak correlation was observed between the two matrices for *M. squamosus* (*r* = 0.446; Fig. 5b).

The results of the analysis of model 1 indicated that the divergence between *M. loricatus* and *M. squamosus* occurred around 2.6 Mya and the split between *M. loricatus* and *M. loricatus CD* was estimated to have occurred around 68,551 ya. These dates are similar to those estimated by Amaral et al. (2013) of 2.6 Mya (95% HPD: 1.5–4.4 mya) and 135,000 YBP (95% HPD: 42,000–1,388,000), respectively. We faced problems running the subsequent scenarios (Fig. S1, Table S4). First, the 100 runs in models 2 and 3 were not enough to estimate the parameters. Second, models 4 and 5 presents a pathology called runaway behavior, where parameter values (e.g., effective population sizes and estimated date of a given event such as divergence, population expansion and contraction or migration) degenerate to zero or diverge to infinity, caused by statistical noise that leads SFS to be incompatible with the models, or our models were misspecified (Rosen *et al.* 2018). Given these problems it was not possible to calculate the AIC values to identify the best model. These problems shall be solved before the submission of the manuscript.

Discussion

Population genetic structure

The analyses of AF taxa have been unveiling a complex effect of historical processes (Costa, 2003), and this includes phylogeographic studies of AF birds (e.g. Cabanne *et al.* 2007, D'Horta *et al.* 2011, Batalha-Filho *et al.* 2012). Until recently the majority of phylogeographic investigations have been based on relatively small number of genes and the limitations of such datasets has been long recognized (e.g., Cartens *et al.* 2012). One potential problem is that a single or a few loci represent just one or a few of many possible realizations of a given organismal demographic history (Knowles 2004). Also, as the number of substitutions of different loci of a given species and among species for a given gene varies according to stochastic nature of DNA change, estimations of population divergence times based on few loci can be more biased than those based on

many markers (Edwards & Beerli 2000). However, in the last few years, due to technological advances evolutionary biology is undergoing an exciting transition with the use of high-throughput sequencing (McCormack *et al.* 2012).

In the present study we used data from thousands of loci produced by reduced genome sequencing to investigate the evolutionary history of two AF bird species. The first difference between these species was the number of SNPs recovered using the same filtering parameters. *M. loricatus* presented more SNPs than *M. squamosus* (Table S2), indicating its higher genetic diversity and in agreement with previous findings for these species (Amaral *et al.* 2013). The lower diversity in *M. squamosus* could be associated to its presence in the southern AF that was more fragmented in the Last Glacial Maximum than the central and northern AF (Moritz *et al.* 2000, Carnaval & Moritz, 2008). And it is expected to observe higher genetic diversity in in the more stable areas (refugia, Hewitt 2000) and in the case of the AF these areas were in the north. While in the south, the stronger range changes could have resulted in lower population effective sizes and therefore lower genetic diversity (Cabanne et al. 2016). We also detected a decrease in genetic diversity from the north to the south in both species (Figs. 4a, 4c). This pattern was also detected in other studies of AF organisms (e.g. plants, Palma-Silva et al. 2009; birds, Maldonado-Coelho et al. 2012). This could indicate that for both species we studied, the southern area of distribution was colonized by individuals from the north.

Based on our sNMF and PCA results, we confirmed that *M. loricatus* and *M. squamosus* are distinct evolutionary units (Fig. 2a, 3a). It seems that these birds are under a scenario of complete reproductive isolation, as previously suggested by Amaral *et al.* (2013). Even though their songs sound strikingly similar, playback experiments showed that there is clear mate recognition with responses to conspecific songs in both species (Macedo *et al.* 2019). Their similar ecological niches could be leading to competitive exclusion (den Boer 1986; Weir & Price 2011), also sexual selection (Catchpole, 1987), and sexual discrimination based on multiple traits (Hohenlohe & Arnold, 2010), and this could be reinforcing the maintenance of these divergent lineages (Edwards *et al.* 2005). Based on the sNMF results (Fig. 2a) admixture appears to be very limited, and the low percentage of shared genetic ancestrality between species could be a sign of old introgression since these individuals present parapatric distribution, or retention of ancestral polymorphism. However, we noted that the three *M. squamosus* individuals with lowest number of consensus of clustered reads after filtering (individual 2452 with

16661 reads, 16502 with 19985 reads, and individual 18624 with 29355 reads, Table S1) showed the highest admixture levels with *M. loricatus*. This indicated that this admixture could be an artifact of poorer sequencing. In agreement with Amaral et al (2013), *M. squamosus* presented K=1 with the best value (Fig. S1c, Table S3). However, we found evidences of IBD in the results of PCA (Figs. 3c), Mantel test (Fig. 5b), and EEMS (Fig. S2b). As the correlation between geographic and genetic distances in the Mantel test was weak (r = 0.446), sNMF may not have been able to detect the IBD pattern.

M. loricatus showed a clear IBD pattern based on various results: PCA (Fig. 3b), a strong correlation in Mantel test (r = 0.718, Fig. 5a), and also sNMF that showed gradual change of ancestry from the north to the south (Fig. 2b). sNMF showed K=2 as the best population assignment (Fig. S1b, Table S3. A phylogeographic study of an AF bird from the same family (Thamnophilidae) as the one of the two species studied here and based on one mitochondrial marker did not reject IBD in the north of the AF in Bahia refuge but rejected IBD in southern populations (Maldonado-Coelho 2012). The same pattern was observed in the present study, with stronger IBD in the northern *M. loricatus* (Fig. 5a) than in *M. squamosus* (Fig. 5b).

In general, species with restricted dispersal abilities seem to present IBD (e.g. birds, Dixo *et al.* 2009; seahorse, Lourie *et al.* 2005; insects, Phillipsen *et al.* 2015). And in usually antbirds seem to present inability to disperse through or survive in habitats other than their own (Bravo, 2012). This characteristic could be associated to the gradual change in allele frequencies from one end of the distribution of the species to the other (Figs.3b, 3c). And local accumulation of genetic differences could also have been the result of restricted dispersal among populations or subgroups (Slatkin 1993). Also, most birds of tropical rainforest are sedentary (Karr 1971) and have long been suspected to have low mobility (Willis 1974). The limited dispersal ability of tropical forest birds is congruent with aspects of the species' ecology, as the necessity of travel extensively within continuous forest (Moore et al, 2008). It is also suggested that insectivores (as antbirds) in general present high habitat specificity, low mobility, and are more confined to the forest interior (Stouffer & Bierregaard Jr. 1995, Salisbury *et al.* 2012). The possible low dispersal ability and the territorial behavior in both species of the present study (Macedo *et al.* 2019, Zimmer & Isler, 2003) could be associated to the evidences of IBD observed. Another possible process that could have resulted in this

IBD pattern is if the past range expansion was rapid with a single ancestral population giving rise to all extant demes (Slatkin, 1993).

It is noteworthy that the use of thousands of markers allowed us to detect IBD that could not be detected based on a few markers (Amaral *et al*. 2013). This improvement in detection of population structure patterns based on thousands of markers produced by GBS was also observed in a switchgrass that was previously analyzed by random amplified polymorphic DNA (RAPD) markers (Lu *et al*., 2013). These more fine-tune results allow to better characterize mechanisms of dispersal and patterns of connectivity between individuals that help to infer how ecological and environmental heterogeneity shape the distribution of genetic variation in nature (Wang & Bradburd 2014).

Even though *M. loricatus* individuals from the Chapada Diamantina (localities 1 and 2, Fig. 1, Table 1) presented high genetic distance (Fig. 3b) compared to those from the coast of Bahia, this differentiation was not as significant as was observed for these birds by Amaral *et al*. (2013). It is noteworthy that the Chapada Diamantina is known for its unique composition of avian community (Parrini et al. 1999) and presence of differentiated lineages in several groups of organisms (e.g. plants, Valente et al. 2013; amphibian, Napoli & Juncá, 2006; fish, Melo & Espindola 2016; birds, Gonzaga et al. 2007). But, in our case, this differentiation was not high enough to disturb the IBD pattern. It is possible that gallery forests along the Paraguaçu River basin could act as corridors between the Chapada Diamantina and the coast (Silveira et al. 2019), keeping gene flow between these localities.

Demographic inferences

The divergence date between the two *Myrmoderus* species was estimated around the Miocene and Pliocene. This period is characterized by high tectonic activity, at the River of Paraíba do Sul Valley (Petri & Fúlfaro 1983, Valladares et al. 2017) around where these species meet (Fig. 1). This valley is also a landmark (not always a primary barrier) that limits other vertebrate taxa (lineages or species; e.g. Silva & Straube 1996; Pellegrino 2005; Pessoa 2007; Cabanne *et al.* 2007; D'Horta *et al.* 2011). Thus, it can be possible that past events that occurred around the Paraíba do Sul Valley may have influenced the divergence of *M. loricatus* (north of the valley) and *M. squamosus* (south).

Conclusion

Natural populations undergo evolutionary processes of migration, size changes, and divergence, and the history of these demographic events shaped their present genetic diversity. The present study suggests that is phylogeographic break correlate with a Paraiba valley in the Atlantic Forest, and also reinforce that the response of organisms to common histories may be idiosyncratic, and predictions about the history of the biome should take into account ecological characteristics and distribution of each specific taxa (D'Horta *et al.* 2011). Our results suggested that isolation by geographical distance was important in shaping the population genetic structure of *Myrmoderus* spp. and this was only detected based on the use of thousands of markers that were generated by next generation–sequencing (Gilad *et al.* 2009).

References

Amaral FR, Albers PK, Edwards SV, Miyaki CY (2013) Multilocus tests of Pleistocene refugia and ancient divergence in a pair of Atlantic Forest antbirds (*Myrmeciza*). *Molecular Ecology*, **22**, 3996-4013.

Andrews S (2010) *FastQC: A quality control tool for high throughput sequence data*. Brabaham Bioinformatics, Cambridge, England. http://www.bioinformatics.babraham.ac.uk/projects/fastqc .

Batalha-Filho H, Cabanne GS, Miyaki CY (2012) Phylogeography of an Atlantic Forest passerine reveals demographic stability through the last glacial maximum. *Molecular Phylogenetics and Evolution*, **65**, 892–902.

Bravo GA (2012) *Phenotypic and Niche Evolution in the Antbirds (Aves: Thamnophilidae)*. Ph. D dissertation. Louisiana State University, Baton Rouge.

Brown KS, Ab'Saber AN (1979) Ice-ages forest refuges and evolution in Neotropics: correlation of paleoclimatoligical, geomorphological and pedological data with biological endemism. *Paleoclimas*, **5**, 1-30.

Cabanne GS, Santos FS, Miyaki CY (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.

Cabanne GS, d'Horta FM, Sari EHR, Santos FR, Miyaki CY (2008) Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): biogeography and systematic implications. *Molecular Phylogenetics and Evolution*, **49**, 760–773.

Cabanne GS, Calderón L, Trujillo Arias, et al. (2016). Effects of Pleistocene climate changes on species ranges and evolutionary processes in the Neotropical Atlantic Forest. *Biological Journal of the Linnean Society*, **119**, 856-872.

Carnaval AC, Moritz C (2008) Historical climate modeling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, **35**, 1187–1201.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science*, **323**, 785–789.

Carstens B, Lemmon AR, Lemmon EM (2012). The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology*, **61**, 713-715.

Catchpole CK (1987) Bird song, sexual selection and female choice. *Trends in Ecology & Evolution*, **2**, 94-97.

Costa LP, Leite YL, Fonseca GAB, Fonseca, MT (2000) Biogeography of South American Forest mammals: endemism and diversity in the Atlantic forest. *Biotropica* **32**, 872-881.

Costa LP (2003) The historical bridge between the Amazon and the Atlantic forest of Brazil: a study of molecular phylogeography with small mammals. *Journal of Biogeography*, **30**, 71–86.

Cracraft J (1985) Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. In *Neotropical Ornithology* (eds Buckley PA, Foster MS, Morton CS, Ridgely RS, Buckley FG), pp. 49-84. American Ornithologists Union, Washington.

Danecek P, Auton A, Abecasis G *et al*. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416-423.

Den Boer PJ (1986) The present status of the competitive exclusion principle. *Trends in Ecology & Evolution*, **1**, 25-28.

D'Horta FM, Cabanne GS, Meyer D, Miyaki CY (2011) The genetic effects of Late Quaternary climatic changes over a tropical latitudinal gradient: diversification of an Atlantic Forest passerine. *Molecular Ecology*, **20**, 1923-1935.

Dixo M, Metzger JP, Morgante JS, Zamudio KR (2009) Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biological Conservation*, **142**, 1560-1569

Eaton DAR, Overcast I (2016) *Ipyrad: interactive assembly and analysis of RADseq data sets*. http://ipyrad.readthedocs.io/.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Edwards SV, Kingan SB, Calkins JD *et al*. (2005) Speciation in birds: genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6550-6557.

Edwards SV, Shultz AJ, Campbell-Staton SC (2015) Next-generation sequencing and the expanding domain of phylogeography. *Folia Zoologica*, **64**, 187-207.

Elshire RJ, Glaubitz JC, Sun Q *et al*. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS ONE*, **6**, e19379.

Fonseca RR, Albrechtsen A, Themudo GE *et al*. (2016) Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, **30**, 3-13.

Frichot E, Schoville SD, Bouchard G, François O (2012) Correcting principal component maps for effects of spatial autocorrelation in population genetic data. *Frontiers in Genetics*, **3**, 254.Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**, 925-929.

Galindo-Leal CG, Câmara IG (2003) The Atlantic Forest of South America: Biodiversity Status, Threats, and Outlook. Island Press, Washington.

Gilad Y, Pritchard JK, Thornton K (2009) Characterizing natural variation using next-generation sequencing technologies. *Trends in Genetics*, **25**, 463-471.

Gómez JP, Bravo GA, Brumfield RT, Tello JG, Cadena CD (2010) A phylogenetic approach to disentangling the role of competition and habitat filtering in community assembly of Neotropical forest birds. *Journal of Animal Ecology*, **79**, 1181–1192.

Gonzaga LP, Carvalhaes AM, Buzzetti DR (2007) A new species of *Formicivora* antwren from the Chapada Diamantina, eastern Brazil (Aves: Passeriformes: Thamnophilidae). *Zootaxa*, **1473**, 25-44.

Graham CH, Moritz C, Williams SE (2006). Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 632–636.

Haffer J (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hohenlohe PA, Arnold SJ (2010) Dimensionality of mate choice, sexual isolation, and speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16583-16588.

Jombart T, Collins C (2015) *An Introduction to Adegenet 2.0.0*. Imperial College London-MRC Centre for Outbreak Analysis and Modelling, 43.

Kamm JA, Terhorst J, Durbin R, Song YS (2019) Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, 1-42.

Karr JR (1971) Ecological, behavioral and distributional notes on some central Panama birds. *Condor*, **73**, 107–111.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Lourie SA, Green DM, Vincent ACJ (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: *Hippocampus*). *Molecular Ecology*, **14**, 1073-1094.

Lu F, Lipka AE, Glaubitz J *et al*. (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, **9**, e1003215.

Macedo G, Silva M, Amaral FRD, Maldonado-Coelho M (2019) Symmetrical discrimination despite weak song differentiation in 2 suboscine bird sister species. *Behavioral Ecology*, https://doi.org/10.1093/beheco/arz096.

Maldonado-Coelho M (2012) Climatic oscillations shape the phylogeographical structure of Atlantic Forest fire-eye antbirds (Aves: Thamnophilidae). *Biological Journal of the Linnean Society*, 105, 900–924.

Mardis ER (2017) DNA sequencing technologies: 2006-2016. *Nature Protocols*, **12**, 213-218.

McCormack JE, Maley JM, Hird SM *et al*. (2012) Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution*, **62**, 397-406.

Melo MR, Espindola VC (2016) Description of a new species of *Characidium* Reinhardt, 1867 (Characiformes: Crenuchidae) from the Chapada Diamantina, Bahia, and redescription of *Characidium bimaculatum* Fowler, 1941. *Zootaxa*, **4196**, 552-568.

Moore RP, Robinson WD, Lovette IJ, Robinson TR (2008) Experimental evidence for extreme dispersal limitation in tropical forest birds. *Ecology Letters*, **11**, 960-968.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, **31**, 533–563.

Morley RJ (2000) *Origin and Evolution of Tropical Rain Forests*. John Wiley & Sons Ltd, Chichester, UK.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GABD, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853-858.

Mukaka MM (2012) Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, **24**, 69-71.

Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H (2015) Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Current Biology*, **25**, 1375-1380.

Napoli M F, Juncá FA (2006) A new species of the *Bokermannohyla circumdata* group (Amphibia: Anura: Hylidae) from Chapada Diamantina, state of Bahia, Brazil. *Zootaxa*, **1244**, 57-68.

Nielsen CB, Cantor M, Dubchak I, Gordon D, Wang T (2010) Visualizing genomes: techniques and challenges. *Nature Methods*, **7**, S5-S15.

Paglia AP (2005) Panorama geral da fauna ameaçada de extinção no Brasil. In *Lista da fauna Brasileira ameaçada de extinção – incluindo a lista das quase ameaçadas e deficientes em dados* (eds. Machado ABM, Soares Martins C, Drumond GM). Fundação Biodiversitas, Belo Horizonte.

Palma-Silva C, Lexer C, Paggi GM, Barbará T, Bered F, & Bodanese-Zanettini MH. (2009). Range-wide patterns of nuclear and chloroplast DNA diversity in *Vriesea gigantea* (Bromeliaceae), a Neotropical forest species. *Heredity*, **103**, 503.

Parrini R, Raposo MA, Pacheco JF et al. (1999). Birds of the Chapada Diamantina, Bahia, Brazil. *Cotinga*, **11**, 86-95.

Pellegrino KCM, Rodrigues MT, Waite AN *et al*. (2005) Phylogeography and species limits in the *Gymnodactylus darwinii* complex (Gekkonidae, Squamata): genetic structure coincides with river systems in the Brazilian Atlantic Forest. *Biological Journal of the Linnean Society*, **85**, 13–26.

Pessoa RO (2007) *Sistemática e Biogeografia Histórica da Família Conopophagidae (Aves: Passeriformes): Especiação nas Florestas da América do Sul*. Ph. D dissertation, Universidade de São Paulo, São Paulo.

Petri S, Fúlfaro VJ (1983) *Geologia do Brasil – Fanerozóico*. Editora da Universidade de São Paulo, São Paulo.

Phillipsen IC, Kirk EH, Bogan MT *et al*. (2015) Dispersal ability and habitat requirements determine landscape level genetic patterns in desert aquatic insects. *Molecular Ecology*, **24**, 54-69.

Podani J, Miklos I (2002) Resemblance coefficients and the horseshoe effect in principal coordinates analysis. *Ecology*, **83**, 3331-3343.

Ribeiro CA, Metzger JP, Martensen AC, Ponzoni FJ, Hirota MM (2009) The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, **142**, 1141–1153.

Ridgely RS, Tudor G (2009) *Field Guide to the Songbirds of South America*, University of Texas Press, Austin.

Rosen Z, Bhaskar A, Roch S, & Song YS (2018) Geometry of the sample frequency spectrum and the perils of demographic inference. *Genetics*, **210**, 665-682.

Salisbury CL, Seddon N, Cooney CR, Tobias JA (2012) The latitudinal gradient in dispersal constraints: ecological specialization drives diversification in tropical birds. *Ecology Letters*, **15**, 847-855.

Sant'Anna-Neto JL, Nery JT (2005) Variabilidade e mudanças climáticas no Brasil e seus impactos. In *Quaternário do Brasil* (eds. Souza GRG, Suguio K, Oliveira MAS, Oliveira PL). Holos Editora, Ribeirão Preto.

Sick H (1997) *Ornitologia Brasileira*, Editora Nova Fronteira, Rio de Janeiro.

Silva JMC, Straube FC (1996) Systematics and biogeography of scaled woodcreepers (Aves: Dendrocolaptidae). *Studies on Neotropical Fauna and Environment*, **31**, 3–10.

Silva JMC, Sousa MC, Castelletti CHM (2004) Areas of endemism for passerine birds in the Atlantic forest, South America. *Global Ecology and Biogeography*, **13**, 85–92.

Silveira MHB, Mascarenhas R, Cardoso D, Batalha-Filho H (2019) Pleistocene climatic instability drove the historical distribution of forest islands in the northeastern Brazilian Atlantic Forest. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **527**, 67-76.

Skutch AF (1996) *Antbirds and Ovenbirds: their Lives and Homes*. University of Texas Press, Austin.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Stouffer PC, Bierregaard Jr RO (1995) Use of Amazonian forest fragments by understory insectivorous birds. *Ecology*, **76**, 2429-2445.

Stotz DF, Fitzpatrick JW, Parker III TA, Moskovits DK (1996) *Neotropical Birds: Ecology and Conservation*. University of Chicago Press, Chicago.

Tabarelli M, Vicente A, Barbosa DCA (2003) Variation of seed dispersal spectrum of woody plants across a rainfall gradient in north-eastern Brazil. *Journal of Arid Environments*, **53**, 197-210.

Valladares CS, Duarte BP, Heilbron M, Ragatky D (2017) The tectono-magmatic evolution of the occidental terrane and the Paraíba do Sul klippe within the Neoproterozoic Ribeira orogenic belt, southeastern Brazil. *Revista Brasileira de Geociências*, **30**, 001-006.

Valente EDB, Pôrto KC, Bastos CJP, & Ballejos-Loyola J (2013) Diversity and distribution of the bryophyte flora in montane forests in the Chapada Diamantina region of Brazil. *Acta Botanica Brasilica*, **27**, 506-518.

Vanzolini PE, Williams EE (1970) South American anoles: the geographic differentiation and evolution of the *Anolis chrysolepis* species group (Sauria, Iguanidae). *Arquivos de Zoologia do Estado de São Paulo*, **19**, 1–298.

Wallace AR (1852) On the monkeys of the Amazon. *Proceedings of the Zoological Society of London*, **20**, 107–110.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, **23**, 5649-5662.

Weir JT, Price TD (2011) Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *The American Naturalist*, **177**, 462–469.

Whitmore TC, Prance GT (1987) *Biogeography and Quaternary History in Tropical America*. Oxford University Press, Oxford.

Willis EO (1974) Populations and local extinctions of birds on Barro Colorado Island, Panama. *Ecological Monographs*, **44**, 153–169.

Wright S (1943) Isolation by distance. *Genetics,* **28,** 114-138.

Zimmer KJ, Isler ML (2003) Family Thamnophilidae. In: *Handbook of the Birds of the World. Vol. 8: Broadbills to Tapaculos* (eds. del Hoyo J, Elliot A, Christie DA). Lynx Editions, Barcelona.

Table 1. Sampling localities of *Myrmoderus loricatus* and *M. squamosus.* #- Locality number. * all deposited at the Laboratório de Genética e Evolução Molecular de Aves (LGEMA), Universidade de São Paulo.

Species	#	Locality	State	Longitude	Latitude	Samples*
M. loricatus	1	Bonito	BA	-41.2	-11.9	2670, 2671
M. loricatus	2	Lençois	BA	-41.37	-12.55	13743, 13750, 13756, 13759
M. loricatus	3	Serra da Jibóia (mata das torres); Elísio Medrado	BA	-39.47	-12.85	14963, 15024
M. loricatus	4	Serra do Timbó; Amargosa	BA	-39.63	-13.07	16736, 16758
M. loricatus	5	Mata do Pacangê, (Área da Michelin); Igrapiúna	BA	-39.23	-13.83	14917, 14918
M. loricatus	6	RPPN Serra Bonita; Camacan	BA	-39.57	-15.38	13724, 13725, 13728
M. loricatus	7	Rebio Mata Escura; Jequitinhonha	MG	-40.99	-16.0	13694, 13695, 13697
M. loricatus	8	Poté	MG	-41.82	-17.73	13539, 13551, 13553
M. loricatus	9	RPPN. Serra do Caraça	MG	-43.37	-20.17	15760, 15773, 15775
M. loricatus	10	Serra da Concórdia	RJ	-43.71	-22.32	2352
M. loricatus	11	Itatiaia	RJ	-44.52	-22.42	12068, 18630
M. loricatus	12	PN Serra dos Órgãos; Teresópolis	RJ	-42.9	-22.45	13012, 13014, 13020
M. loricatus	13	Rio Claro	RJ	-44.2	-22.85	18631, 18641, 18642
M. loricatus	14	Angra dos Reis	RJ	-44.35	-22.91	18634, 18635
M. squamosus	15	Paraty	RJ	-44.7	-23.04	18643, 18644
M. squamosus	16	Paraty	RJ	-44.82	-23.2	12008
M. squamosus	17	Catucaba; São Luiz Paraitinga	SP	-45.17	-23.23	1525
M. squamosus	18	Tremembé	SP	-45.67	-22.95	18622, 18623, 18624
M. squamosus	19	Estação Biológica de Boracéia	SP	-45.88	-23.65	16484, 16502, 16510, 15643
M. squamosus	20	Praias de Boracéia e Guaratuba; Bertioga	SP	-45.97	-23.73	10328, 10333
M. squamosus	21	Parque Estadual Serra do Mar, Nucleo Curucutu	SP	-46.72	-23.98	14713, 15220, 15066
M. squamosus	22	P. E. Morro Grande	SP	-46.96	-23.67	3162
M. squamosus	23	Peruíbe; Guaraú	SP	-47.02	-24.37	12065, 12066, 12067
M. squamosus	24	Piedade	SP	-47.45	-23.69	2452
M. squamosus	25	Estrada para Fazenda Santa Regina; Tapiraí	SP	-47.59	-23.94	17873
M. squamosus	26	Utinga; Guaraqueçaba	PR	-48.22	-25.2	14679, 14678, 14663
M. squamosus	27	Nova Trento	SC	-49.0	-27.35	14583, 14600, 14601
M. squamosus	28	Pousada Rancho Fundo, Morro Grande; Nova Roma	SC	-49.77	-28.7	14557, 14528

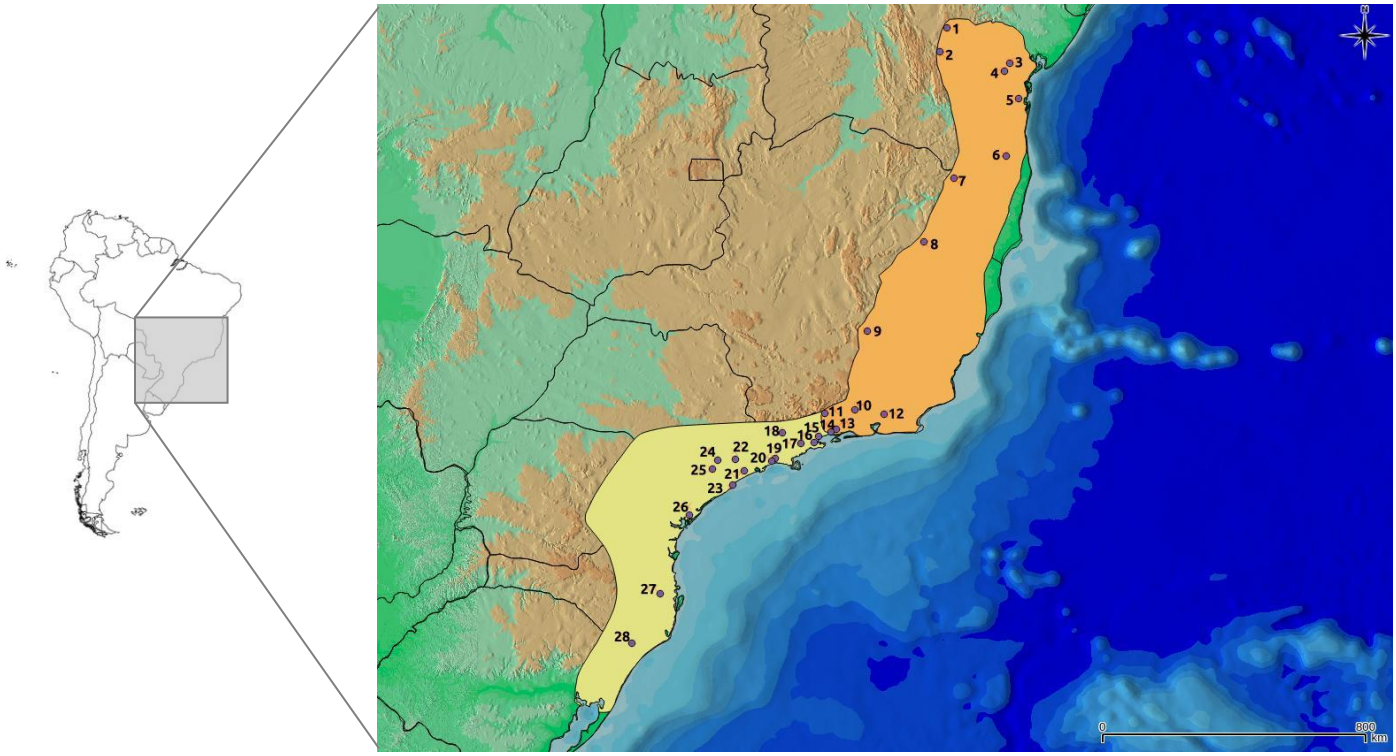


Figure 1. Map of South America and distribution of *Myrmoderus loricatus* (orange) and *Myrmoderus squamosus* (yellow). Numbers correspond to localities (see Table 1).

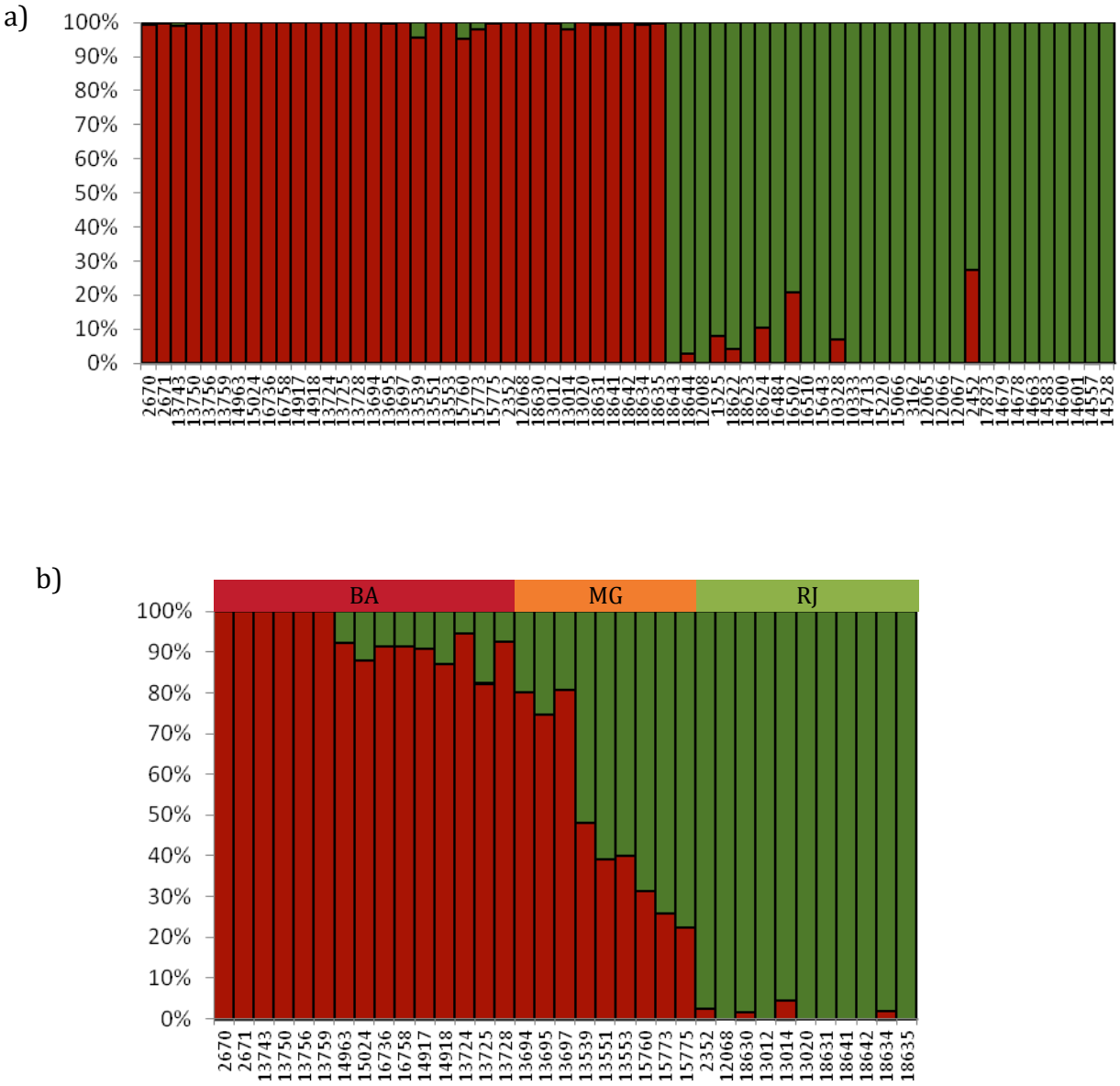


Figure 2. Results of sNMF analyses. a) K = 2 based on 15,527 SNPs of *Myrmoderus* spp. b) K = 2 based on 17,294 SNPs of *M. loricatus*. The colors of the vertical bars represent the two genetic clusters. The names below the graphs correspond to samples (see Table 1) arranged according to increasing latitude from left to right. BA – Bahia; MG – Minas Gerais; RJ – Rio de Janeiro.

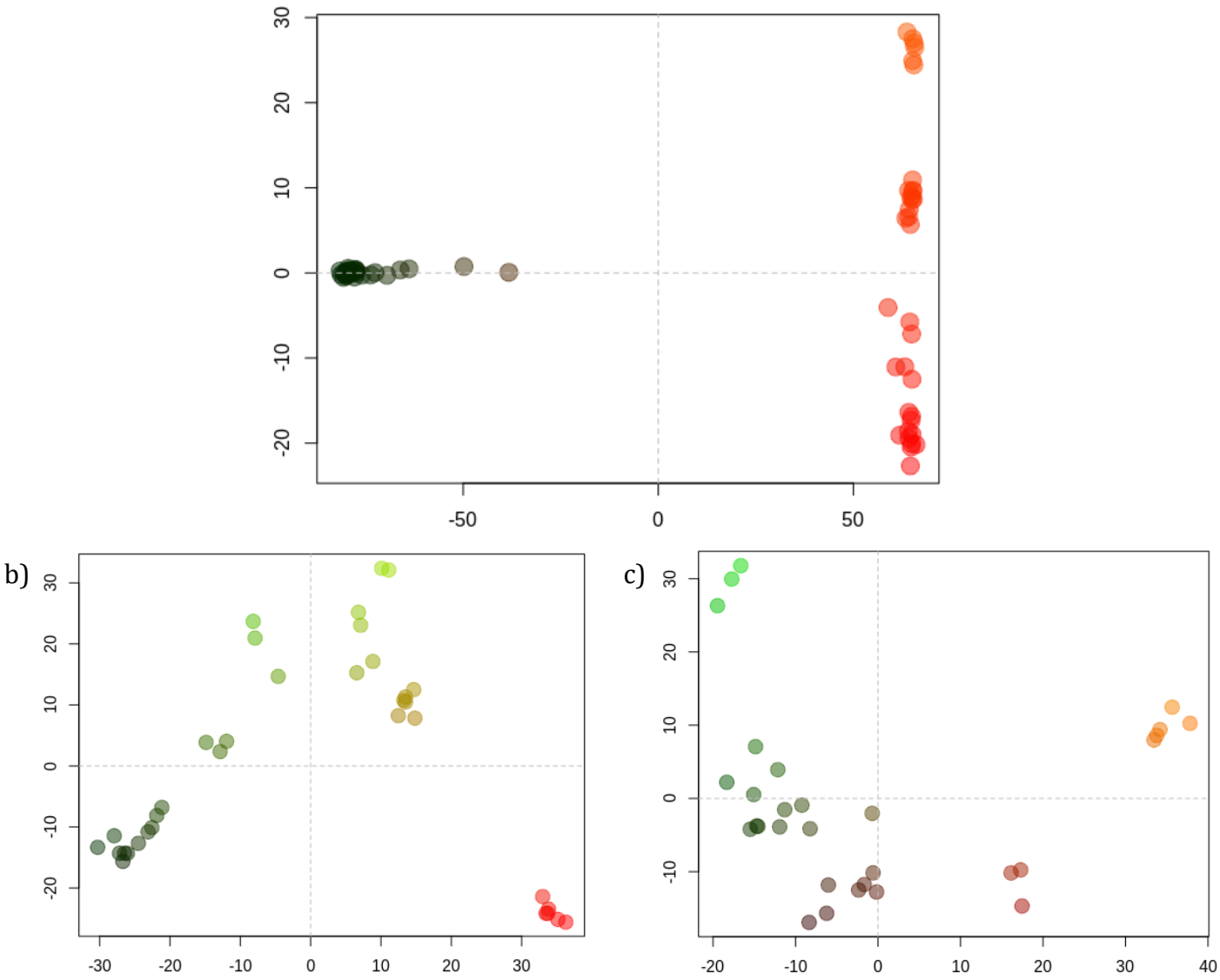


Figure 3. PCA results showing the first (X axis) and the second (Y axis) main components. a) *Myrmoderus* spp. based on 15,527 SNPs. b) *M. loricatus* based on 17,294 SNPs. c) *M. squamosus* based on 9,621 SNPs. Each circle represents one individual. Circles with similar colors correspond to genetically more similar individuals.

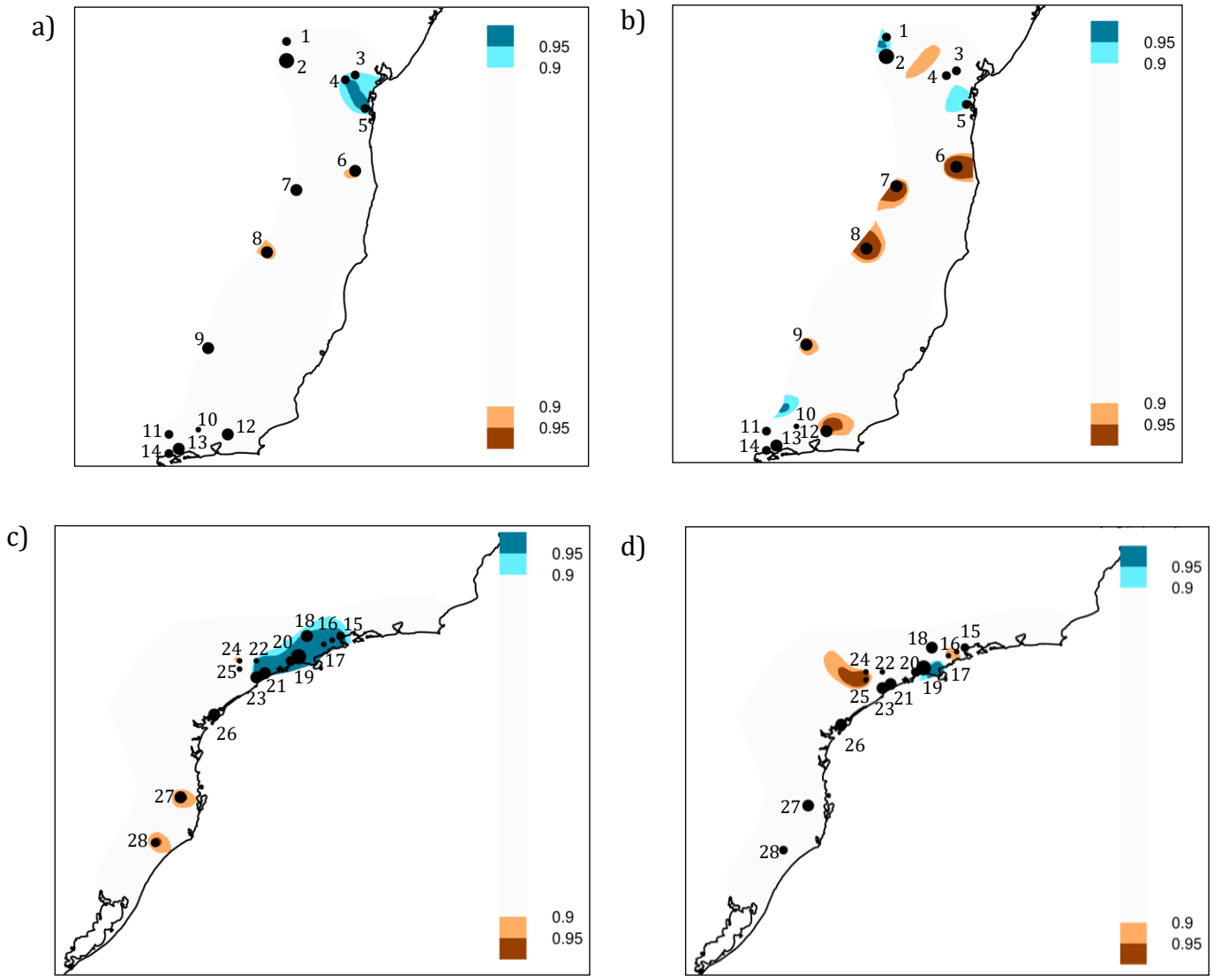


Figure 4. Spatial distribution of confidence intervals of posterior distributions of effective diversity and effective migration. a) Effective diversity of *M. loricatus* (35 individuals) based on 14,416 SNPs. b) Effective migration of *M. loricatus* (35 individuals) based on 14,416 SNPs. c) Effective diversity of *M. squamosus* (30 individuals) based on 7,446 SNPs. d) Effective migration of *M. squamosus* (30 individuals) based on 7,446 SNPs. The colors represent the higher (blue) and lower (brown) than average rates. Numbers correspond to localities (see Table 1).

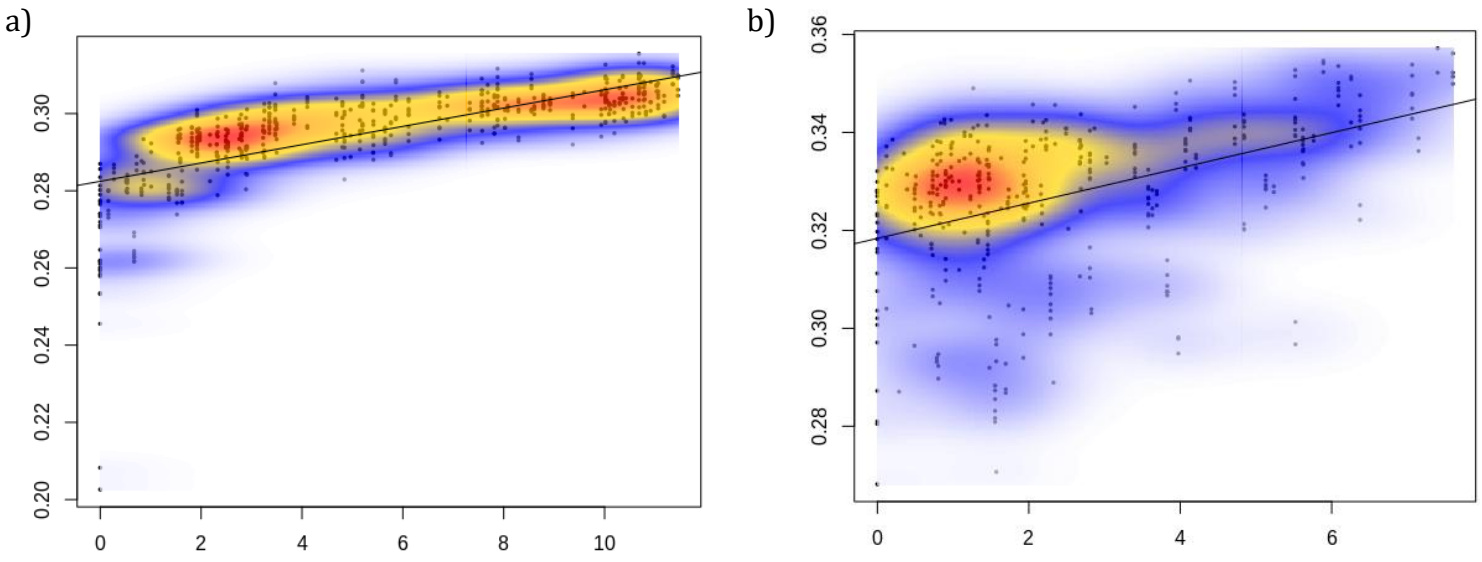


Figure 5. Plots of isolation by distance analysis showing the correlation between genetic distance (Y axis) and linear geographic distance (X axis, Euclidean distance). a) *M. loricatus* (35 individuals) based on 17,294 SNPs. b) *M. squamosus* (30 individuals) based on 9,621 SNPs

Table S1. Results of read filtering in ipyrad of 65 individuals *of M. squamosus* and *M. loricatus*. Steps 2 to 5 are shown. **N reads-** Number of reads; **Reads Ns-** Number of reads discarded due to number of Ns higher than four; **Reads min len-** Number of reads retained with minimum length of 75 pb; **Clusters total-** Total number of contigs; **Avg depth tot-** Average total depth; **Sd depth tot-** Total depth standard deviation; **Clusters hi depth-** Contigs with minimum depth of eight per site; **Avg depth-** Depth average using minimum of eight depth per site; **Sd depth-** Depth standard deviation using minimum of eight depth per site; **Hetero est-** Estimated heterozygosity per site; **Error est-** Estimated heterozygosity standard error; **Filt Max H-** Number of clusters removed with more than four heterozygous sites; **Filt Max N-** Clusters removed with more than five Ns per site; **Reads cons Total**- Total number of consensus of clustered reads; **N sites-** Size of all concatenated contigs; **N hetero-** Size of concatenated contigs of heterozygous sites.

	Step 2			Step 3						Step 4		Step 5				
Samples	N reads	Reads Ns	Reads min len	Clusters total	Avg depth tot	Sd depth tot	Clusters hi depth	Avg depth	Sd depth	Hetero est	Error est	Filt Max H	Filt Max N	Reads cons	N sites	N hetero
10328	3415153	858	1818261	269744	5.67	23.45	47532	22.31	52.65	0.006924	0.002039	1418	204	45910	4735880	13541
10333	3852882	1049	1329795	272559	8.79	39.23	63738	29.58	77.50	0.006708	0.001683	1818	171	61748	6467361	19883
12008	5742168	1559	2873273	341707	7.99	51.17	71240	29.46	109.39	0.007144	0.001938	2152	204	68883	7117257	23042
12065	5308646	1559	1783827	303114	10.95	50.59	82604	33.52	93.18	0.006229	0.001955	2199	185	80217	8334048	24959
12066	4049057	1164	1404872	290851	8.67	55.61	66297	29.70	113.94	0.006497	0.001883	1837	177	64281	6713474	19749
12067	4387034	1283	1415432	297202	9.52	53.55	72884	31.18	105.18	0.006210	0.001783	1890	178	70815	7427188	22084
12068	3487615	1159	939818	273295	8.71	54.38	66280	28.09	108.10	0.008267	0.002105	1989	197	64093	6515869	28418
13012	4324512	1456	1164770	282898	10.45	74.81	78731	30.73	139.75	0.007759	0.001902	2250	151	76328	7658379	31303
13014	1850751	614	457189	215436	6.18	27.03	42834	21.61	58.03	0.008543	0.002137	1367	139	41328	4254565	16979
13020	4776057	1402	1253085	295704	10.95	60.84	86453	31.16	109.88	0.007857	0.001871	2408	169	83874	8432883	36363
13539	1940048	590	425584	228734	6.31	25.53	45430	22.09	54.40	0.008251	0.002186	1332	168	43930	4500137	18532
13551	5024550	1577	1389376	317385	10.77	71.21	90344	31.36	131.19	0.008189	0.001894	2432	181	87728	8714354	40200
13553	5124613	1666	1396901	297516	11.65	55.37	90295	32.44	97.33	0.008022	0.002017	2516	178	87600	8779894	39558
13694	3654480	1210	977988	272724	9.15	50.56	68339	28.89	98.34	0.008072	0.002066	2001	149	66188	6760272	29059
13695	3225750	1073	883041	246649	8.84	58.02	61280	27.88	114.27	0.008485	0.002086	1903	160	59215	6108222	27015
13697	4458765	1348	1185601	284427	10.60	76.32	80424	30.88	141.48	0.008331	0.001942	2365	159	77899	7808008	36152
13724	6088315	1964	1563380	332078	12.62	93.51	102745	35.05	165.90	0.007934	0.002054	2780	168	99792	9824211	44510
13725	4997184	1544	1349269	293517	11.49	90.09	85908	32.95	164.53	0.008182	0.001937	2436	189	83278	8376232	38824
13728	2941192	928	758483	245086	8.31	35.48	59337	26.45	68.96	0.008344	0.002128	1877	180	57280	5853602	24541
13743	2104884	647	510769	225618	6.65	41.15	47355	22.59	87.95	0.008089	0.002210	1529	149	45676	4714796	17408
13750	3833071	1245	1057765	269857	9.54	59.05	70385	29.17	113.31	0.007851	0.001830	2087	157	68140	6901784	27542
13756	4406002	1449	1187301	279932	10.62	71.36	79070	30.99	132.07	0.007419	0.002078	2194	164	76710	7691800	30340
13759	3685666	1084	966533	282512	8.91	60.93	67172	29.41	122.69	0.007655	0.001933	1981	180	65010	6734277	25608
14528	4996257	1571	1659848	305592	10.38	68.26	80851	32.31	130.18	0.005472	0.001897	2086	192	78570	8173322	18841
14557	3448573	948	1463106	275919	6.87	47.23	54758	25.31	103.94	0.006293	0.001990	1697	225	52834	5545453	12326
14583	5572532	1682	1569107	399094	9.56	71.32	97371	31.56	142.13	0.005701	0.001744	2656	232	94479	9595354	22234
14600	3394868	1014	1110574	245401	8.93	40.42	60672	28.64	77.98	0.005485	0.001743	1496	167	59008	6196653	14248
14601	4892443	1554	1454451	326534	10.04	65.51	81020	32.90	128.81	0.005632	0.001897	2099	179	78741	8031388	19596
14663	4473285	1249	1694801	367490	7.25	37.23	81730	24.30	76.49	0.005110	0.001784	1989	256	79483	8319892	15594
14678	6971292	2165	2045744	384228	12.11	84.31	109083	36.35	155.59	0.005835	0.001587	2625	195	106259	10724614	29977
14679	3578758	989	1184395	285179	7.95	52.69	62549	27.55	110.26	0.006436	0.001724	1847	178	60521	6346343	16784
14713	2021165	671	510538	200495	7.08	53.55	47843	22.10	108.23	0.006817	0.002333	1470	201	46169	4957524	14278
14917	6316652	2092	1745178	305108	13.82	122.44	99978	36.93	212.00	0.008098	0.001761	2721	190	97061	9664106	45313
14918	3277681	1092	875923	275733	8.07	51.09	62848	26.87	104.81	0.008549	0.002207	1983	195	60669	6185115	27150

	Step 2			Step 3						Step 4		Step 5				
Samples	N reads	Reads Ns	Reads min len	Clusters total	Avg depth tot	Sd depth tot	Clusters hi depth	Avg depth	Sd depth	Hetero est	Error est	Filt Max H	Filt Max N	Reads cons	N sites	N hetero
14963	4837812	1554	1324789	318141	10.23	79.82	79305	33.34	157.60	0.008414	0.002176	2495	208	76598	7846997	35515
15024	3005653	991	776120	257642	7.95	51.72	57877	26.75	106.96	0.008565	0.001954	1885	172	55819	5782380	24708
15066	4231798	1312	1317399	263971	10.42	76.55	71383	31.72	145.05	0.006401	0.001976	1876	187	69317	7285215	22845
15220	3480692	1075	1010577	244384	9.57	66.28	64558	29.16	126.89	0.006504	0.002148	1785	190	62580	6630019	20157
15760	1956370	618	685768	176149	6.72	44.65	39637	22.51	92.36	0.007894	0.002515	1262	172	38202	3986674	13799
15773	2098282	622	649089	202980	6.65	41.18	43322	23.01	87.16	0.008525	0.002475	1430	233	41658	4377729	16961
15775	3591555	1127	1104124	298322	7.78	57.02	62021	28.12	122.91	0.007775	0.002373	1877	198	59944	6151803	23376
16736	5587152	1945	1492446	381499	9.95	87.22	92251	33.32	175.30	0.008202	0.002056	2677	225	89342	9047539	41351
16758	7114929	2391	1820873	360750	13.47	110.59	118093	35.88	191.32	0.007829	0.001892	2957	218	114912	11328041	52781
17873	2769202	846	750033	307239	6.23	45.23	52454	25.67	107.30	0.007129	0.002022	1724	182	50547	5370149	15801
18622	1562476	449	473794	173168	5.96	36.95	37091	19.73	78.26	0.006982	0.002071	1090	133	35868	3767349	10817
18623	3811908	1094	1352020	238821	9.67	71.59	58547	32.03	142.25	0.006512	0.001941	1671	190	56682	6038209	18155
18624	1248174	413	385301	153731	5.31	39.65	30460	18.10	87.87	0.007206	0.002538	932	173	29355	3006816	8735
18630	4675559	1508	1486910	344239	8.56	78.15	76312	30.17	164.14	0.007588	0.002368	2261	281	73768	7574688	28635
18631	4613187	1270	1526638	237642	11.56	86.05	63240	36.46	164.22	0.008463	0.002164	2188	164	60885	6264928	27231
18634	2847567	813	955800	208276	8.25	53.45	47359	28.56	109.64	0.008302	0.002441	1581	155	45622	4785039	18981
18635	3334849	1062	1002879	243534	8.73	66.29	54950	30.50	137.31	0.008674	0.002442	1967	203	52778	5555571	22557
18641	5689185	1757	1669179	392053	9.34	62.39	91183	31.98	126.72	0.007537	0.002042	2636	219	88325	8852300	34420
18642	5375055	1706	1576097	297960	11.62	75.05	82341	35.24	140.00	0.007917	0.001926	2503	171	79662	8136657	33949
18643	3838717	1107	1259283	258118	9.40	38.91	62913	30.92	74.76	0.006676	0.001746	1780	140	60993	6471915	20270
18644	1892383	551	513088	222970	5.87	23.61	41958	21.33	51.53	0.007152	0.002079	1275	146	40537	4178846	12395
15643	2165116	619	718832	199270	6.90	16.50	44824	22.81	29.58	0.006644	0.001723	1192	142	43490	4584334	13405
16484	2933958	854	1028231	226397	8.08	36.52	50678	28.22	73.67	0.006743	0.001828	1414	162	49101	5220428	15712
16502	812216	239	246958	130088	4.13	13.18	20787	14.10	30.87	0.007895	0.002492	705	97	19985	1960578	5946
16510	2203479	695	661376	199359	7.30	31.21	47619	23.19	61.14	0.006714	0.002034	1326	199	46094	4943600	14433
1525	1278604	434	348779	170076	5.19	31.53	32452	18.10	70.67	0.007414	0.002365	1065	168	31219	3236650	9508
2352	4935828	1543	1268497	305749	11.03	76.26	88843	31.68	139.31	0.008160	0.002006	2498	206	86137	8747927	39532
2452	767043	280	171971	154871	3.71	20.11	17394	16.76	58.23	0.007778	0.002948	614	119	16661	1684742	4792
2670	7135683	2170	1849386	358674	13.62	117.18	119041	35.95	201.53	0.007213	0.001953	2976	188	115871	11415082	46460
2671	3664980	1112	941394	275340	9.16	64.34	69767	28.55	125.80	0.007588	0.002002	1978	161	67626	6909473	26899
3162	2559889	743	672081	220287	8.15	44.88	54353	25.58	88.05	0.006473	0.001959	1520	151	52681	5630172	16121

Table S2. Number of SNPs (1 SNP per locus) obtained from 65 samples *of Myrmoderus* spp. (35 *M. loricatus* and 30 *M. squamosus*) according to the percentage of completeness of the matrix.

Matrix completeness	**SNPs** *Myrmoderus* spp.	**SNPs** *M. loricatus*	**SNPs** *M. squamosus*
50%	26246	28188	16153
85%	17574	18696	10837
90%	15527	17294	9621
95%	11961	14413	5962

Table S3. Lowest cross-entropy K values (1 to 10) obtained with four alpha values (100 replicates for each) of *Myrmoderus* spp., *M. loricatus* and *M. squamosus*. In bold the best values per dataset.

Myrmoderus spp.	**Alpha 1**	**Alpha 10**	**Alpha 100**	**Alpha 1000**
K=1	0.4026	0.4056	0.4069	0.4078
K=2	0.2590	0.2617	**0.2607**	0.2618
K=3	0.2608	0.2632	0.2623	0.2633
K=4	0.2654	0.2662	0.2681	0.2688
K=5	0.2671	0.2714	0.2737	0.2751
K=6	0.2762	0.2754	0.2791	0.2816
K=7	0.2839	0.2815	0.2823	0.2911
K=8	0.2908	0.2858	0.2958	0.2976
K=9	0.3008	0.3000	0.3029	0.3083
K=10	0.3072	0.3053	0.3088	0.3158
M. loricatus				
K=1	0,4831	0,4814	0,4805	0,4864
K=2	0,4794	**0,4764**	0,4778	0,4878
K=3	0,4876	0,4906	0,4899	0,5006
K=4	0,5012	0,4988	0,4969	0,5158
K=5	0,5178	0,5177	0,5188	0,5362
K=6	0,5384	0,5344	0,5356	0,5587
K=7	0,5535	0,5657	0,5651	0,5746
K=8	0,5828	0,5913	0,5814	0,6159
K=9	0,6026	0,6069	0,6095	0,6489
K=10	0,6334	0,6327	0,6399	0,6817
M. squamosus				
K=1	0,5826	**0,5803**	0,5803	0,5762
K=2	0,5873	0,5919	0,5818	0,5834
K=3	0,6217	0,6305	0,6214	0,6454
K=4	0,6686	0,6657	0,6729	0,7144
K=5	0,7119	0,7104	0,7193	0,7644
K=6	0,7540	0,7475	0,7634	0,8331
K=7	0,7840	0,7844	0,8104	0,8791
K=8	0,8290	0,8125	0,8082	0,9553
K=9	0,8587	0,8625	0,8852	0,9601
K=10	0,8986	0,9050	0,9214	1,0535

Table S4. Results obtained for each model (see Fig. S1) tested in Momi2.

Models	Ne Cur. MS	Ne Cur. ML	Ne Cur MLcd	Ne Anc. MS	Ne Anc. ML	T1 MS	T1 ML	Growth Rate- MS	Growth Rate- ML
Model 1	715,291	1,500,000	184,595	-	-	-	-	-	-
Model 2	unkown	unkown	-	unkown	unkown	unkown	-	-	-
Model 3	unkown	unkown	-	unkown	unkown	-	unkown	-	-
Model 4	1,500,000	1,500,000	-	750,818	1,539,040	10,000	10,000	0.0001	0.0001
Model 5	10,001	10,001	-	10,000	10,000	10,000	10,000	-0.006	-0.006
Model 6	1,500,000	754,939	-	-	-	-	-	-	-

MS: *Myrmoderus squamosus*, ML: *Myrmoderus loricatus*, MLcd: *Myrmoderus loricatus* Chapada Diamantina, Ne cur. - Effective population size of diploid individuals of the current population, Ne Anc. - Effective population size of diploid individuals of the ancestral population, T1 – Time 1 in years ago,(-) not tested in the corresponding model.

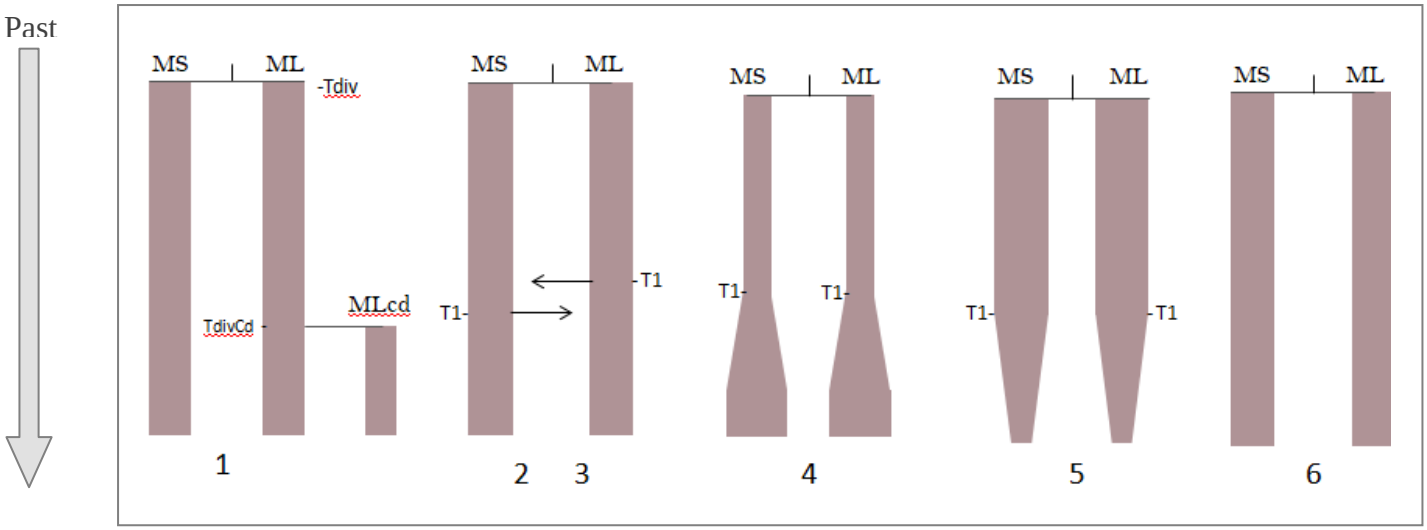


Figure S1. Six historical demographic models tested using SFS by simulations in Momi2. Model 1: used to estimate the divergence times among the three lineages with constant population sizes: *Myrmoderus squamosus* (MS), *Myrmoderus loricatus* (ML), and *Myrmoderus loricatus* Chapada Diamantina (MLcd). Models 2-5: populations were simulated with three types of historical events in T1: migration (models 2 and 3), population expansions (model 4), and bottlenecks (model 5). Model 6: after the divergence lineages (species) presented constant population sizes over time.

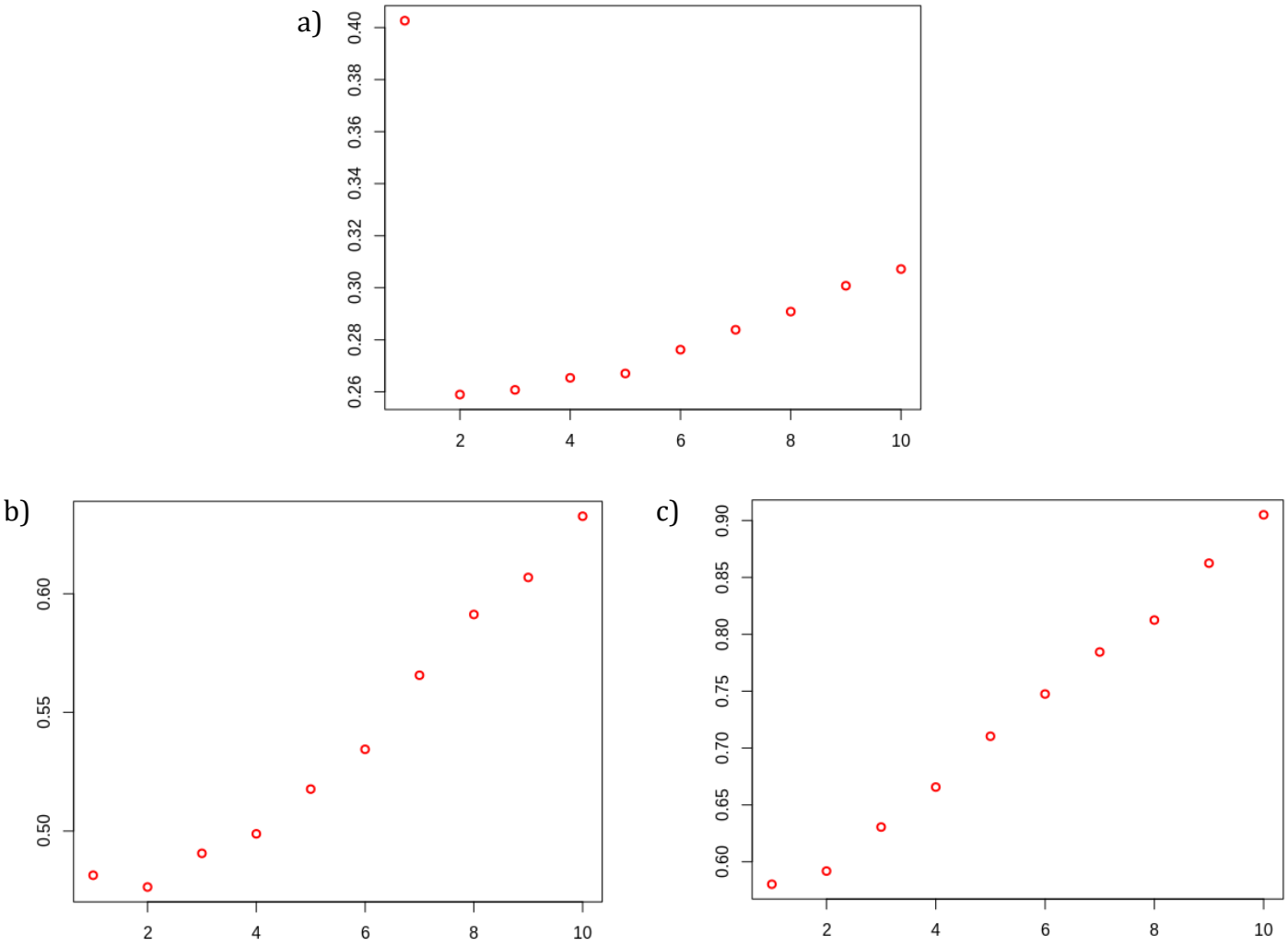


Figure S1. Cross-entropy values for K from 1 to 10 based on 100 replicates per K. a) *Myrmoderus* spp.with alpha = 100. b) *M. loricatus* with alpha = 10. c) *M. squamosus* with alpha = 10.

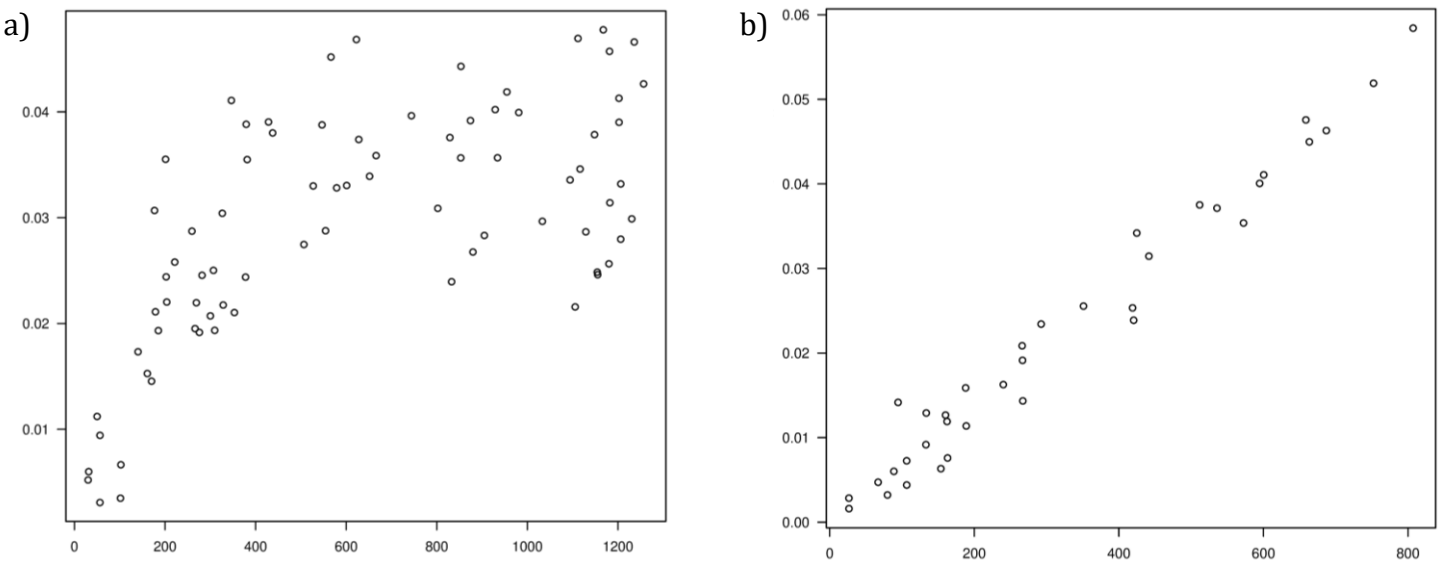


Figure S2. Observed dissimilarity between individuals computed between demes (Y axis) in relation to the geographic distance (X axis). a) *M. loricatus* (35 individuals) based on 14,416 SNPs. b) *M. squamosus* (30 individuals) based on 7,446 SNPs.

Na presente tese descrevemos padrões de diversificação de três espécies de aves endêmicas da Floresta Atlântica (FA), *Myiothlypis leucoblephara, Myrmoderus loricatus* e *Myrmoderus squamosus*, e inferimos potenciais processos históricos que moldaram a diversidade genética desses organismos. Para aumentar nossa compreensão da dinâmica histórica da FA e contribuir com o entendimento da biodiversidade da região Neotropical, foi essencial a testar modelos de diversificação e hipóteses relacionadas a esses modelos. As análises indicaram que a hipótese dos refúgios florestais parece ter sido importante na diversificação das espécies estudadas e isto corrobora outros estudos com organismos da FA (e.g. CABANNE *et al.* 2007, MARTINS *et al*. 2008, THOMÉ *et al*. 2010). Como exemplo da influência das variações climáticas detectada nas espécies de estudo, podemos citar: 1) sinal de expansão demográfica em *M. leucoblephara* no período correspondente ao último máximo glacial, resultado associado à instabilidade no sul da FA; 2) variação da diversidade genética ao longo da FA em *Myrmoderus* spp. com a maior diversidade genética associada a áreas mais estáveis, ao norte da FA em *M. loricatus*. Além da hipótese dos refúgios, pudemos inferir que as atividades tectônicas ocorridas no Quaternário, como a formação do Vale do Paraíba do Sul, possivelmente atuaram na diversificação das espécies irmãs do gênero *Myrmoderus.* Nesse contexto, o presente estudo revela um cenário complexo no qual os padrões de biodiversidade na FA resultaram de interações entre eventos geológicos antigos que moldaram a distribuição de linhagens e eventos climáticos mais recentes, que impulsionaram processos evolutivos. Além desses eventos que possivelmente atuaram na diversidade genética desses táxons, salientamos que atributos ecológicos intrínsecos, como dependência florestal, territorialismo e flexibilidade de habitat, conferem características únicas às espécies. Compreender a interação desses fatores ao longo do tempo e suas consequências na diversificação dos organismos é um grande desafio. Sugerimos que novas investigações multidisciplinares, como a associação entre dados genômicos e ambientais, incluindo modelagem de nicho ecológico, podem proporcionar uma melhor compreensão de como a distribuição da diversidade genética observada se moldou.

Observamos para as três espécies, fluxo gênico em curtas distâncias suficiente para impedir a diferenciação genética local. Esse resultado não havia sido detectado em estudos prévios baseados em poucos marcadores. Assim, possivelmente a detecção de isolamento por distância parece ser consequência do aumento substancial no número de marcadores. Isso reforça a importância da análise de grandes quantidades de dados

cujos resultados permitem identificar melhor mecanismos de dispersão e padrões de conectividade entre indivíduos, além de serem excelentes ferramentas para inferir como a heterogeneidade ecológica e ambiental molda a distribuição da variação genética na natureza (WANG; BRADBURD 2014).

Estudos filogeográficos têm amostrado progressivo uso de maior quantidade de locos, em parte motivado por trabalhos teóricos que mostraram que as estimativas dos principais parâmetros demográficos melhoram à medida que o número de locos aumenta (GARRICK *et al*. 2015). Isso aumenta a resolução e a precisão das investigações demográficas históricas e essa estratégia está apenas começando a ser aplicada aos organismos da FA. Assim, nosso trabalho é um dos primeiros estudos aplicando metodologia de sequenciamento de nova geração utilizando aves desse domínio como modelo, e podemos concluir que está de acordo com outros estudos que mostram que reanálises da história demográfica de organismos usando mais marcadores genéticos resultam em resultados mais robustos (MCCORMACK *et al.* 2012).

CABANNE, G. S.; SANTOS, F. R.; MIYAKI, C. Y. Phylogeography of *Xiphorhynchus fuscus* (Passeriformes, Dendrocolaptidae): vicariance and recent demographic expansion in southern Atlantic forest. **Biological Journal of the Linnean Society**. v. 91, p. 73–84, 2007.

GARRICK, R. C.; BONATELLI, I. A.; HYSENI, C.; MORALES, A.; PELLETIER, T. A.; PEREZ, M. F.; CARSTENS, B. C. The evolution of phylogeographic data sets. **Molecular Ecology**, v. 24. n. 6, p. 1164-1171, 2015.

MARTINS, F. M.; TEMPLETON, A. R.; PAVAN, A. C. O.; KOHLBACH, B. C.; MORGANTE, J. S. Phylogeography of the common vampire bat (*Desmodus rotundus*): marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. **BMC Evolutionary Biology**, v. 9, p. 294-XXX?, 2009.

MCCORMACK, J. E.; MALEY, J. M.; HIRD, S. M.; DERRYBERRY, E. P.; GRAVES, G. R.; BRUMFIELD, R. T. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. **Molecular Phylogenetics and Evolution**, v. 62, p. 397-406, 2012.

THOMÉ, M. T. C.; ZAMUDIO, K. R.; GIOVANELLI, J. G. R.; HADDAD, C. F. B.; BALDISSERA F. A. JR.; ALEXANDRINO, J. M. B. Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic forest. Molecular **Phylogenetics and Evolution,** v. 55, p. 1018–1031, 2010.

WANG, I. J.; BRADBURD, G. S. Isolation by environment. **Molecular Ecology**, v. 23, p. 5649-5662, 2014.