

**Programa Institucional de Bolsas de Iniciação Científica do Instituto Chico
Mendes de Conservação da Biodiversidade PIBIC/ICMBio**



Relatório Final
(Ciclo 2022-2023)

**FILOGENIA DO GÊNERO *CALLICEBUS* UTILIZANDO
MARCADORES GENÔMICOS**

Letícia de Alcântara Oliveira Trajano

Orientador(a): Amely Branquinho Martins

João Pessoa

Setembro/2023

Resumo

O gênero *Callicebus* é endêmico do Brasil, e possui cinco espécies atualmente reconhecidas: *Callicebus personatus* Geoffroy 1812, *Callicebus melanochir* Wied Neuwied 1820, *Callicebus nigrifrons* Spix 1823, *Callicebus barbarabrownae* Hershkovitz 1990 e *Callicebus coimbrai* Kobayashi and Langguth 1999. Das cinco espécies atualmente reconhecidas para o gênero *Callicebus*, 4 foram consideradas ameaçadas de extinção na última avaliação do risco de extinção da fauna brasileira (MMA. 2014). Sendo *Callicebus barbarabrownae*, o guigó-da-Caatinga avaliado como Criticamente Em Perigo (CR); *Callicebus coimbrai* avaliado como Em Perigo (EN); e ambos *Callicebus melanochir* e *Callicebus personatus* avaliados na categoria Vulnerável (VU). Neste sentido, tanto o processo de avaliação do risco de extinção da fauna, quanto os Planos de Ação Nacional (PAN) que contemplam as espécies ameaçadas do gênero *Callicebus*, PAN para Conservação dos Primatas do Nordeste (PRINE) e PAN para a Conservação dos Primatas da Mata Atlântica e da Preguiça de Coleira (PPMA) poderão se beneficiar das informações sobre a diversidade genética do gênero *Callicebus*, as relações evolutivas das espécies do gênero, bem como dos marcadores genômicos identificados para futuros estudos populacionais. Este estudo pretende ampliar as análises filogenéticas pela implementação de análise Bayesiana e Análise de Quartetos além de realizar teste de delimitação de espécies pelo Fator de Bayes e identificar potenciais híbridos. Para caracterizar a estrutura da diversidade genética dos indivíduos com relação aos agrupamentos genéticos encontrados e avaliar a presença de hibridação, foram realizadas análises de estrutura populacional, usando o método de agrupamento baseado em modelo de análise Bayesiano com o software STRUCTURE para inferir o número ideal de populações ou grupos genéticos.

Palavras-chave: Callicebus, hibridização, populações, fologia, genética para conservação

Abstract

The genus *Callicebus* is endemic to Brazil, and has five species currently recognized: *Callicebus personatus* Geoffroy 1812, *Callicebus melanochir* Wied Neuwied 1820, *Callicebus nigrifrons* Spix 1823, *Callicebus barbarabrownae* Hershkovitz 1990 and *Callicebus coimbrai* Kobayashi and Langguth 1999. Between the five species in genus *Callicebus*, 4 were considered endangered in the last assessment of the risk of extinction of the Brazilian fauna (MMA. 2014). Being *Callicebus barbarabrownae*, the Caatinga guigó evaluated as Critically Endangered (CR); *Callicebus coimbrai* evaluated as Endangered (EN); and both *Callicebus melanochir* and *Callicebus personatus* evaluated in the Vulnerable (VU) category. In this sense, both the fauna extinction risk assessment process and the National Action Plans (PAN) that address endangered species of the genus *Callicebus*: PAN for the Conservation of Northeastern Primates (PRINE) and PAN for the Conservation of Primates da Mata Atlântica e da Preguiça de Coleira (PPMA) may benefit from information on the genetic diversity of the genus *Callicebus*, the evolutionary relationships of the species of the genus, as well as the genomic markers identified for future population studies. This study intends to expand the phylogenetic analyzes for the implementation of Bayesian analysis and Quartets Analysis in addition to carrying out a species delimitation test by the Bayes Factor and identifying potential hybrids. the clustering method based on Bayesian analysis model with the software STRUCTURE to infer the optimal number of populations.

Key words: Callicebus, hybridization, populations, phylogeny, conservation genetics

Sumário

1	INTRODUÇÃO.....	5
2	OBJETIVOS.....	7
3	MATERIAL E MÉTODOS	7
4	RESULTADOS.....	10
5	DISCUSSÃO E CONCLUSÕES	19
6	RECOMENDAÇÕES PARA O MANEJO.....	20
7	AGRADECIMENTOS.....	20
8	REFERÊNCIAS	21

1. Introdução

Os macacos zogues ou guigós pertencem a um grupo de primatas do Novo Mundo endêmicos da América do Sul, que foram recentemente reclassificados em três gêneros (*Cheracebus*, *Plecturocebus* e *Callicebus*) por Byrne e colaboradores (2016). O gênero *Callicebus*, que atualmente inclui cinco espécies, é endêmico do leste do Brasil, ocorrendo nos biomas Caatinga, Cerrado e Mata Atlântica. Em um estudo inédito (Carneiro *et al.* 2018) pesquisadores inferiram as relações filogenéticas destes três gêneros (Byrne *et al.* 2016, Carneiro *et al.* 2018), seus tempos de divergência e padrões biogeográficos - com base na análise molecular de uma sequência concatenada de 11 marcadores de DNA mitocondrial e nuclear, derivados de 13 espécimes. Seus resultados sugerem que o ancestral de todos os macacos zogues habitava uma área que se estendia da floresta amazônica ao sul da Mata Atlântica, há cerca de 12 milhões de anos. Ainda de acordo com este estudo, um primeiro evento de separação de linhagens originou o gênero *Cheracebus* no oeste da Amazônia constituindo-se, portanto, no gênero mais antigo dentre os guigós, enquanto o ancestral de *Callicebus* e *Plecturocebus* foram separados posteriormente. A diversificação de *Callicebus* deve ter ocorrido durante o Plioceno-Pleistoceno (começando em cerca de 5Ma) provavelmente influenciada por flutuações climáticas e eventos geológicos (Byrne *et al.* 2016).

O gênero *Callicebus* é endêmico do Brasil, e possui cinco espécies atualmente reconhecidas (Figura 1): *Callicebus personatus* Geoffroy 1812, *Callicebus melanochir* Wied-Neuwied 1820, *Callicebus nigrifrons* Spix 1823, *Callicebus barbarabrownae* Hershkovitz 1990 e *Callicebus coimbrai* Kobayashi and Langguth 1999. Entretanto, existem divergências quanto ao nível taxonômico das formas reconhecidas para o gênero enquanto espécies ou subespécies. Hershkovitz (1990) considerou as formas *personatus*, *melanochir*, *nigrifrons* e *barbarabrownae* como subespécies de *C. personatus*, sendo que *C. coimbrai* ainda não havia sido descrita. Por outro lado, Kobayashi (1995) e Kobayashi e Langguth (1999), principalmente considerando dados craniométricos, sugeriram o nível específico para todas as formas do gênero, o que foi aceito por van Roosmalen *et al.* (2002). Groves (2001), seguiu amplamente a proposta de Hershkovitz (1990), considerando as formas *personatus*, *nigrifrons*, *melanochir* e *barbarabrownae* como sub-espécies de *C. personatus*, mas aceitou o nível específico para *C. coimbrai*.

Os resultados do trabalho de Carneiro e colaboradores (2018) confirmaram a existência das cinco espécies que atualmente habitam áreas florestais. Entretanto, este estudo utilizou apenas um indivíduo de cada uma das formas de *Callicebus*, além disso, todas elas sendo amostras de cativeiro. Assim, mesmo tendo representado um importante avanço no conhecimento das

Comentado [UC1]: Leticia, como é um trabalho de continuação do primeiro, e de mesma temática, você pode utilizar parte da introdução do relatório do ciclo anterior. É essencial fornecer um arcabouço teórico do projeto na introdução!

relações evolutivas do gênero *Callicebus*, um novo estudo sobre a filogenia molecular das espécies deste gênero, com maior amostragem e diferentes marcadores genéticos é importante para o ampliar o entendimento sobre a evolução do grupo.

Das cinco espécies atualmente reconhecidas para o gênero *Callicebus* (Figura I), quatro foram consideradas ameaçadas de extinção na última avaliação do risco de extinção da fauna brasileira (MMA 2014). Sendo *Callicebus barbarabrownae*, o guigó-da-Caatinga, avaliado como Criticamente Em Perigo (CR); *Callicebus coimbrai* avaliado como Em Perigo (EN); e ambos *Callicebus melanochir* e *Callicebus personatus* avaliados na categoria Vulnerável (VU). Neste sentido, tanto o processo de avaliação do risco de extinção da fauna, quanto os Planos de Ação Nacional (PAN) que contemplam as espécies ameaçadas do gênero *Callicebus*: PAN para Conservação dos Primatas do Nordeste (PRINE) e PAN para a Conservação dos Primatas da Mata Atlântica e da Preguiça de Coleira (PPMA) poderão se beneficiar das informações sobre a diversidade genética do gênero *Callicebus*, as relações evolutivas das espécies do gênero, bem como dos marcadores genômicos identificados para futuros estudos populacionais.

No primeiro ano do projeto, foi possível reconstruir um cladograma a partir da análise de máxima verossimilhança que indicou e reiterou o estudo realizado anteriormente, mostrando que dentro do gênero, a espécie *C. nigrifrons* foi a primeira a divergir, seguido por *C. personatus*, e *C. melanochir*, tendo como últimas espécies a divergirem *C. barbarabrownae* e *C. coimbrai*. Entretanto, as amostras destas duas últimas espécies foram reconstruídas como parafiléticas dentro de um mesmo clado. Por isso, é importante e necessário a continuação do trabalho para procurar, por exemplo, evidências genômicas sobre a presença de animais híbridos entre as espécies estudadas (incluindo indivíduos cativos e de vida livre), além de aprofundar as análises filogenéticas com o uso de outras metodologias complementares, como a reconstrução filogenética usando modelo de coalescência. Assim, faz-se necessária a continuidade do presente estudo, com a avaliação da estrutura genética dos indivíduos amostrados, além de futura complementação das análises filogenéticas utilizando outros modelos estatísticos complementares, como o modelo de coalescência (Eaton and Overcast 2020).

2. Objetivos

2.1. Objetivo Geral

O principal objetivo do projeto é utilizar marcadores genômicos (SNPs) para dar continuidade às análises filogenéticas visando a caracterização genética dos indivíduos amostrados, com relação aos agrupamentos genéticos encontrados, além da potencial identificação de híbridos dentro do gênero *Callicebus*, [incluindo amostras de animais de cativeiro e de vida livre.]

2.2. Objetivos específicos

- a) Realizar a caracterização genética de indivíduos com relação aos agrupamentos genéticos encontrados, além da identificação de animais potencialmente híbridos;
- b) Comparar os resultados com estudos anteriores, gerando subsidiar processos de avaliação de risco de extinção e Planos de Ação Nacional que contemplem as espécies alvo do estudo.

3. Material e Métodos

As amostras (Tabela I) e material biológico utilizados nas análises genéticas foram coletadas previamente em atividades e pesquisas coordenadas pelo Centro Nacional de Pesquisa e Conservação de Primatas Brasileiros ou instituições parceiras. O material genômico (ADN total) dos animais amostrados foi extraído e enviado a uma empresa especializada, para construção de bibliotecas de marcadores genômicos amostrados por todo genoma, utilizando a técnica de digestão dupla (*SphI* e *MluCI*) por enzimas de restrição (ddRAD; Peterson *et al.* 2012, Valencia *et al.* 2018), que utiliza duas enzimas de restrição, com adaptadores específicos para cada enzima, e seleção de tamanho por corte de gel automatizado. Os fragmentos de cada amostra, gerados a partir da digestão pelas duas enzimas utilizadas, foram ligados a adaptadores P1, específicos para a plataforma de sequenciamento utilizada, além de marcadores (*barcodes*) específicos para cada amostra, e selecionados por tamanho de fragmento (~300±20pb).

O genoma de referência utilizado foi do *Plecturocebus donacopilus* ([GenBank: PVKP000000000.1,https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/004/027/715/GCA_004027715.1_CalDon_v1_BIUU/GCA_004027715.1_CalDon_v1_BIUU_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/004/027/715/GCA_004027715.1_CalDon_v1_BIUU/GCA_004027715.1_CalDon_v1_BIUU_genomic.fna.gz)), e utilizamos o pipeline *reference*, disponível no programa *ipyRAD* para realizar a genotipagem a partir de mapeamento dos fragmentos de ADN obtidos pelo ddRAD-seq ao genoma do

Comentado [UC2]: Não ampliamos o número de indivíduos, com relação ao ciclo passado. Ampliamos com relação a outros estudos da literatura. Se for este o caso, é necessário deixar mais explícito agora.

Plecturocebus donacopilus. A avaliação da qualidade, filtragem, alinhamento e genotipagem dos fragmentos de ADN sequenciados para identificação de polimorfismos de nucleotídeo único (SNP) foi conduzida com o programa *ipyrad* (v 0.9.81) (Eaton 2014).

Tabela I: Amostras utilizadas no presente estudo.

	Amostra	Espécie	Origem	Precedência*¹
1	CPB215	<i>Callicebus melanochir</i>	Vida livre	CPB/ICMBio
2	CNPIMENTA	<i>Callicebus nigrifrons</i>	Vida livre	CPB/ICMBio
3	CAQ2	<i>Callicebus barbarabrownae</i>	Vida livre	MZUSP
4	CAQ3	<i>Callicebus barbarabrownae</i>	Vida livre	MZUSP
5	CAQ4	<i>Callicebus barbarabrownae</i>	Vida livre	MZUSP
6	CAQ5	<i>Callicebus coimbrai</i>	Vida livre	MZUSP
7	CAQ6	<i>Callicebus coimbrai</i>	Vida livre	MZUSP
8	CAQ7	<i>Callicebus melanochir</i>	Vida livre	MZUSP
9	COIJUNCO	<i>Callicebus coimbrai</i>	Vida livre	CPB/ICMBio
10	COI2	<i>Callicebus coimbrai</i>	Vida livre	CPB/ICMBio
11	COIRAONI	<i>Callicebus coimbrai</i>	Vida livre	CPB/ICMBio
12	CPB242	<i>Callicebus nigrifrons</i>	Cativeiro	FZB-BH
13	CPB243	<i>Callicebus nigrifrons</i>	Cativeiro	FZB-BH
14	CPB244	<i>Callicebus nigrifrons</i>	Cativeiro	FZB-BH
15	CPB245	<i>Callicebus nigrifrons</i>	Cativeiro	FZB-BH
16	CPB246	<i>Callicebus nigrifrons</i>	Cativeiro	FZB-BH
17	CPB247	<i>Callicebus nigrifrons</i>	Cativeiro	FZB-BH
18	CPB266	<i>Plecturocebus cf. moloch</i>	Cativeiro	CPB/ICMBio
19	CPB332	<i>Callicebus nigrifrons</i>	Cativeiro	CPB/ICMBio
20	CPB333	<i>Callicebus nigrifrons</i>	Vida livre	CPB/ICMBio
21	CPB334	<i>Callicebus nigrifrons</i>	Vida livre	CPB/ICMBio
22	CPB335	<i>Callicebus nigrifrons</i>	Vida livre	CPB/ICMBio
23	CPB336	<i>Callicebus nigrifrons</i>	Vida livre	CPB/ICMBio
24	CPB337	<i>Callicebus nigrifrons</i>	Vida livre	CPB/ICMBio
25	CPRJ2144	<i>Callicebus barbarabrownae</i>	Cativeiro	CPRJ
26	CPRJ2329	<i>Callicebus melanochir</i>	Cativeiro	CPRJ
27	CPRJ2381	<i>Callicebus personatus</i>	Cativeiro	CPRJ
28	CPRJ2389	<i>Callicebus melanochir</i>	Cativeiro	CPRJ
29	CPRJ2465	<i>Plecturocebus donacophilu</i>	Cativeiro	CPRJ
30	CPRJ2466	<i>Callicebus personatus</i>	Cativeiro	CPRJ
31	CPRJ2594	<i>Plecturocebus donacophilus</i>	Cativeiro	CPRJ
32	CPRJ2764	<i>Plecturocebus cf. catequensis</i>	Cativeiro	CPRJ

¹CPB/ICMBio = Centro Nacional de Pesquisa e Conservação de Primatas Brasileiros, MZUSP = Museu de Zoologia da USP, FZB-BH = Fundação Zoo-Botânica de Belo Horizonte, CPRJ = Centro de Primatologia do Rio de Janeiro.

Neste segundo ano de projeto, pretendeu-se realizar a caracterização genética dos indivíduos amostrados, buscando elucidar como estes se relacionam com os agrupamentos genéticos encontrados.

Para preparação das matrizes de SNPs para as análises de caracterização da estrutura genética dos indivíduos amostrados foi realizada a geração de uma matriz reduzida, contendo apenas os SNPs que estão há pelo menos 5000 pb de distância, para evitar o efeito do Desequilíbrio de Ligação – LD (Linkage disequilibrium), que é a associação não-aleatória de alelos em dois ou mais loci. Esta etapa foi realizada utilizando o programa VcfTools (Danecek, 2011). Após esta etapa, foi realizada a identificação e remoção de marcadores (SNPs) potencialmente sob efeito de seleção natural (outliers significativos nas frequências alélicas), que foi realizada se utilizando o programa BayeScan v2.1 (Fischer, 2011). **O BayeScan visa identificar loci candidatos sob seleção natural a partir de dados genéticos, usando diferenças nas frequências alélicas entre as populações. (FOLL, 2011).**

Finalmente, foram geradas matrizes de dados com 95% e 90% de completude ou seja, apenas com marcadores presentes em pelo menos 95% e 90% das amostras, respectivamente.

Para caracterizar a estrutura da diversidade genética dos indivíduos e avaliar a presença ou não de hibridação, foram realizadas análises de estrutura populacional, usando o método de agrupamento baseado em modelo de análise Bayesiano com o software STRUCTURE (PRITCHARD et al., 2000) para inferir o número ideal de populações genéticas (K), utilizando-se a matriz de dados de SNPs. O modelo usa a abordagem Markov Chain Monte Carlo (MCMC) que agrupa os indivíduos em "K" populações com base em seu genótipo.

Para as análises de estrutura populacional com o programa STRUCTURE, foram inicialmente realizadas estimativas do valor de lambda, que é um parâmetro referente à frequência dos alelos nas populações. Para cada inferência foram realizadas 10 repetições para cada valor de K e a média dos valores inferidos foi utilizada nas análises subsequentes. Foram utilizados os modelos de ADMIXTURE (que considera como premissa a frequência de alelos correlacionada entre as diferentes espécies ou populações) e **NO-ADMIXTURE** (que tem como premissa a não

Comentado [UC3]: Acredito que no relatório parcial pediram para retirar as partes da metodologias que fossem tão descritivas, ou então resumir.

Comentado [UC4]: Você não chegou a fazer essa análise de No=Admixture! Eu cheguei a rodar, por isso pode deixar aqui. Mas o CLUMPAK está fora do ar desde a última sexta e não consegui fazer os gráficos. Não deu tempo de fazer no R também.

correlação de alelos entre diferentes espécies ou populações) e avaliamos o K variando de 1-10, com 10 repetições para cada K, com 10.000 iterações de burn-in seguidas por 100.000 iterações de MCMC, como proposto em estudos anteriores (Lima et al., 2018; Martins, 2021). O número mais provável de agrupamentos foi identificado usando o método delta K (1K) (EVANNO et al., 2005) utilizando o programa CLUMPAK (KOPELMAN et al., 2015).

O CLUMPAK identifica conjuntos de *runs* altamente semelhantes, separando assim grupos distintos que representam modos distintos no espaço de soluções possíveis. O software identifica um alinhamento ideal de *clusters* inferidos em diferentes valores de K, estendendo uma abordagem semelhante implementada para um K fixo em Clumpp e simplificando a comparação de resultados de agrupamento em diferentes valores de K. Além disso, incorpora recursos adicionais, como implementações de métodos para escolher K e comparar soluções obtidas por diferentes programas, modelos ou subconjuntos de dados (KOPELMAN *et al.*, 2015).

Após isto, utilizamos do software PGDSpider para converter o arquivo VCF gerado anteriormente para formatos específicos, como o formato ".str" (aceito pelo STRUCTURE).

O VCFtools (Variant Call Format tools) (Danecek, 2011) é um conjunto de programas de código aberto para manipular e analisar arquivos no formato Variant Call Format (VCF), um formato de arquivo padrão na genômica e frequentemente usado para armazenar informações sobre variantes genéticas, como SNPs (polimorfismos de nucleotídeo único).

4. Resultados

Primeiramente o ambiente computacional necessário foi organizado, com a instalação dos programas necessários. Além disto, foi gerada a matriz reduzida, contendo apenas marcadores moleculares de nucleotídeo único (SNPs) com distância genética maior que 5000pb, para evitar o efeito de associação não-aleatória de alelos em dois ou mais loci (LD, Desequilíbrio de Ligação), e atender à uma das premissas do programa STRUCTURE. Antes desta filtragem, a matriz de dados era composta por 483532 SNPs, após a filtragem, a matriz resultante possui 42052 SNPs. As análises do STRUCTURE foram realizadas com ambas as matrizes de 90% e 95% de completude. O número de SNPs variou entre as duas matrizes, sendo o total de 16764 para de 95% de completude e 22257 para a de 90% de completude.

Comentado [UC5]: Você não chegou a fazer com o R Adegnet

Comentado [UC6]: Precisa padronizar

Comentado [UC7]: Desnecessário

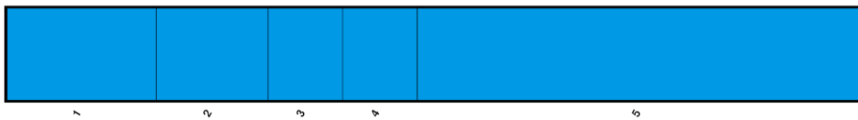
Comentado [UC8]: Esta informação está no pipeline que te enviei, achei que você iria entender que seria interessante incluir.

A escolha do melhor valor de "k" foi baseada em métodos de clustering (agrupamento) e determina o número ideal de grupos (ou populações) que melhor descreve a estrutura dos seus dados genéticos.

Identificamos o número mais provável de clusters usando o método delta K (ΔK , Evanno et al., 2005) utilizando o programa CLUMPAK (Kopelman et al. 2015) e o pacote R adegenet (Jombart 2008). O valores de lambda para cada matriz, foram de $\lambda = 3,0607$ para a de 95% de completude e $\lambda =$ para a de 90. As análises do STRUCTURE (K 1 a 10) estão demonstradas na Figura 2.

Comentado [UC9]: Você não chegou a fazer esta parte.

K=1 10/10, Mean(LnProb) = -206804.490, Mean(similarity score) = 1.000



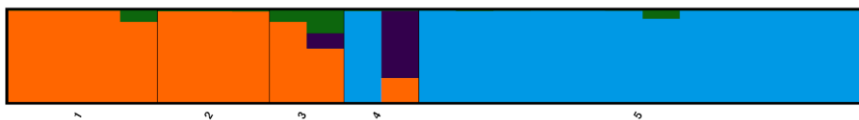
• K=2 10/10, Mean(LnProb) = -117667.850, Mean(similarity score) = 1.000



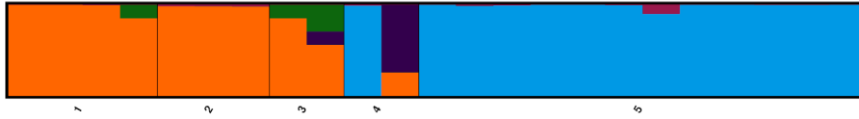
• K=3 10/10, Mean(LnProb) = -111941.660, Mean(similarity score) = 1.000



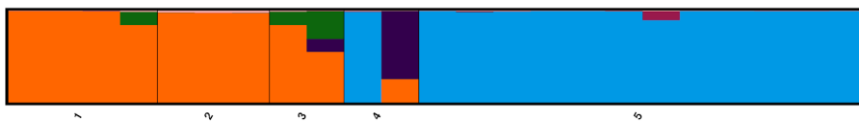
• K=4 10/10, Mean(LnProb) = -563239.060, Mean(similarity score) = 0.992



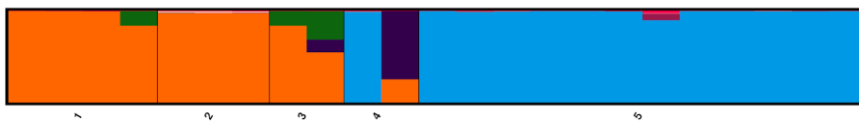
- K=5 10/10, Mean(LnProb) = -745414.440, Mean(similarity score) = 0.991



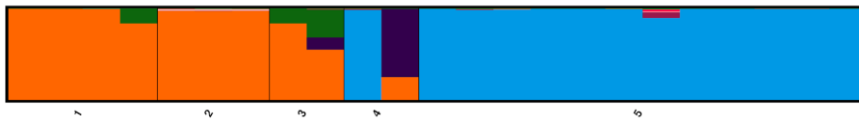
- K=6 9/10, Mean(LnProb) = -329986.822, Mean(similarity score) = 0.988



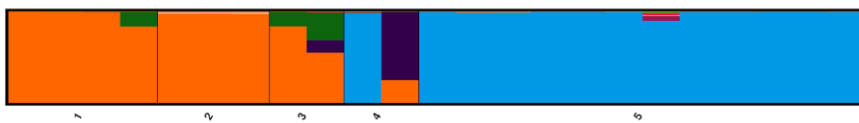
- K=7 10/10, Mean(LnProb) = -425832.610, Mean(similarity score) = 0.992



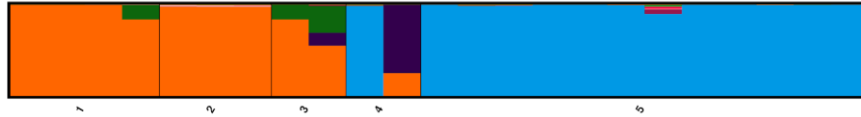
- K=8 10/10, Mean(LnProb) = -317098.790, Mean(similarity score) = 0.991



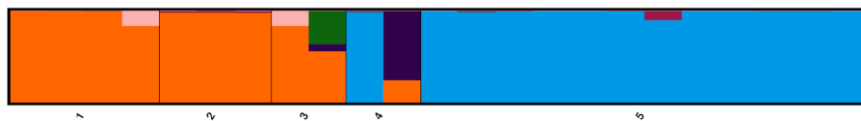
- K=9 8/10, Mean(LnProb) = -224713.200, Mean(similarity score) = 0.992



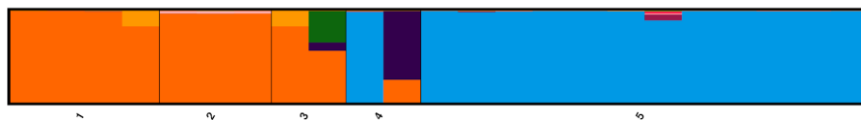
- **K=10 8/10, Mean(LnProb) = -312047.062, Mean(similarity score) = 0.991**



- **K=6 1/10, Mean(LnProb) = -773913.100, Mean(similarity score) = 0.000**



- **K=9 2/10, Mean(LnProb) = -1540970.700, Mean(similarity score) = 0.987**



- **K=10 2/10, Mean(LnProb) = -162571.100, Mean(similarity score) = 0.991**

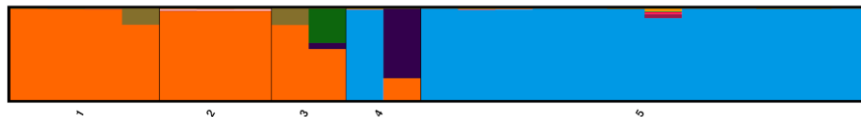


Figura 2. Gráficos da análise do STRUCTURE, modelo admixture, gerados pelo software CLUMPAK, para valores de K entre 1 e 10. Os números de 1 a 5 indicam as cinco espécies representativas de *Callicebus*, tendo 24 espécimes, divididas entre: 13 *Callicebus nigrifrons*, 1 *Callicebus personatus*, 2 *Callicebus melanochir*, 3 *Callicebus coimbrai* e 4 *Callicebus barbarabrownae*.

O valor de “K” representa o número de clusters que o algoritmo tenta identificar em seus dados. Em K=1 10/10 por exemplo, significa que ao tentar agrupar os dados em apenas 1 cluster, o algoritmo atribui todas as amostras a esse único cluster da mesma maneira em todas as 10 repetições. Em K=2 10/10: Ao tentar agrupar seus dados em 2 clusters, o algoritmo também atribui todas as amostras aos dois clusters. Isso sugere que o algoritmo encontrou duas

categorias distintas entre os dados. O mesmo padrão se repete para 3,4 e 5 clusters; todas as amostras são atribuídas a esses três, quatro e cinco grupos distintos (K=3, K=4 e K=5 10/10). Já com K=6 9/10, 1/10: Indica que ao tentar agrupar os dados em 6 clusters, a maioria dos agrupamentos (9 das 10 repetições) seguiu um mesmo padrão, mas em uma das repetições o programa encontrou uma possibilidade de agrupamento genético diferente. Para o K=7 10/10 indica que com 7 clusters, todas as repetições retornaram o mesmo padrão de agrupamento novamente e o mesmo ocorreu com 8 clusters. Com os valores de K9 e K10 a maioria dos agrupamentos genéticos encontrados seguiram o mesmo padrão (8 das 10).

Os resultados demonstram como o número de clusters (K) afeta a atribuição das amostras aos grupos. Quando o K é pequeno, todas as amostras podem ser agrupadas com alta precisão, mas à medida que K aumenta, algumas amostras podem ser atribuídas a clusters diferentes, o que pode indicar uma estrutura mais complexa nos dados, e significar a presença de animais híbridos, ou ainda que a distribuição das espécies não está correta como é atualmente dividida. De acordo com Vähä e Primmer (2005) o parâmetro para identificar o indivíduo como “puro” é o de no mínimo 90% de pertencimento ao cluster da respectiva espécie. O que vemos, porém, são na verdade dois grupos genéticos muito bem distintos, cdivergindo do que seria esperado para cinco espécies distintas.

Os valores indicam uma análise de agrupamento (clusterização) sólida e clara e são exatamente os mesmos para as duas matrizes que utilizamos.

A Mean(LnProb) representa a média dos logaritmos naturais das probabilidades. Em análises bayesianas, o logaritmo natural das probabilidades é usado para avaliar a adequação do modelo aos dados. Quanto menor for esse valor, maior será a probabilidade dos dados sob o modelo. Utilizando o valor de K=2, por exemplo, a probabilidade encontrada pode sugerir que o modelo se ajusta bem aos dados e que estes são muito consistentes com as expectativas do modelo.

Já o Mean(similarity score) indica a média das pontuações de similaridade entre os grupos, no exemplo de K=2 esta média é igual a 1.000, o que significa que os grupos são idênticos em termos de seus atributos ou características. Isso reforça a ideia de que a clusterização é muito precisa e que as amostras são muito semelhantes dentro dos seus respectivos clusters.

Em resumo, os resultados sugerem que os dados foram agrupados com sucesso em dois clusters bem definidos, com alta consistência entre as amostras dentro de cada cluster. Isso pode indicar que seus dados têm uma estrutura clara com duas categorias distintas, pelo menos no contexto da análise realizada

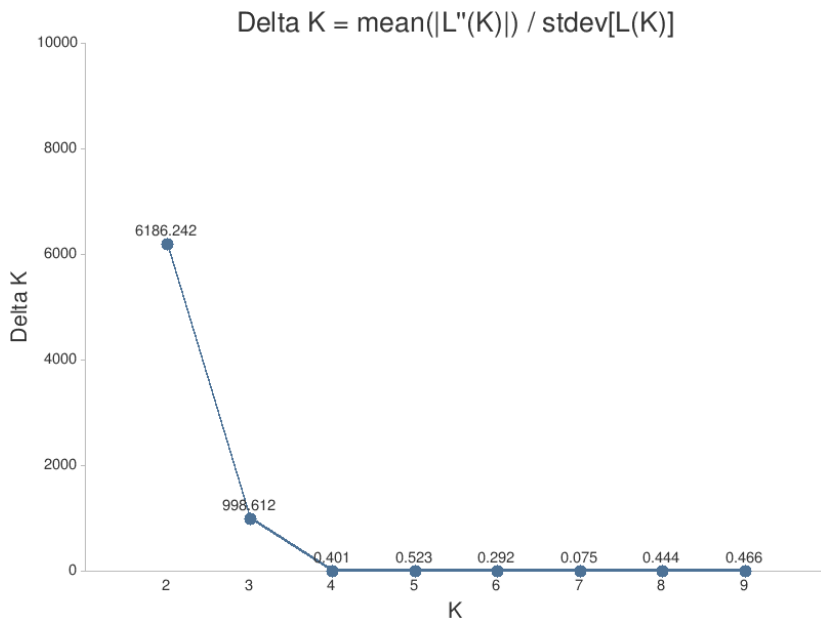


Figura 3. Gráfico do resultado do teste de Evanno, com valor de ΔK para a matriz de 95% de completude.

O ΔK é bastante utilizado na análise de estrutura populacional (clusterização) de dados genéticos, especialmente em análises que utilizam o programa STRUCTURE (Evanno *et al.* em 2005) como uma abordagem para identificar o número ótimo de clusters (K) em uma análise de estrutura populacional. O que é avaliado é a mudança na probabilidade logarítmica dos dados em relação ao número de clusters (K) considerados. Uma série de análises foram realizadas para cada valor de K (de 1 a 10) e em seguida a média da mudança na probabilidade para cada valor de K em todas as suas repetições ou execuções da análise (as análises são geralmente estocásticas e podem produzir resultados ligeiramente diferentes a cada vez). É por fim, calculada a variação entre as médias para diferentes valores de K.

O método de Evanno procura o valor de K que produz o maior ΔK (considerado o número ótimo de clusters).

Em resumo, o método de Evanno é uma técnica estatística usada para determinar o número ideal de clusters em análises de estrutura populacional, levando em consideração a variação nas probabilidades logarítmicas dos dados à medida que o número de clusters aumenta. É útil e eficaz para identificar a estrutura populacional mais provável em dados genéticos.

O resultado do valor “ótimo” de ΔK foi igual a 2 (Figura 3), indicando que, de acordo com o método de Evanno, o valor mais provável para o número de clusters (K) que melhor se ajusta aos dados é 2, o que sugere que os dados genéticos das espécimes provavelmente podem ser

melhor explicados por dois clusters. Cada cluster representaria uma subpopulação ou grupo genético distinto.

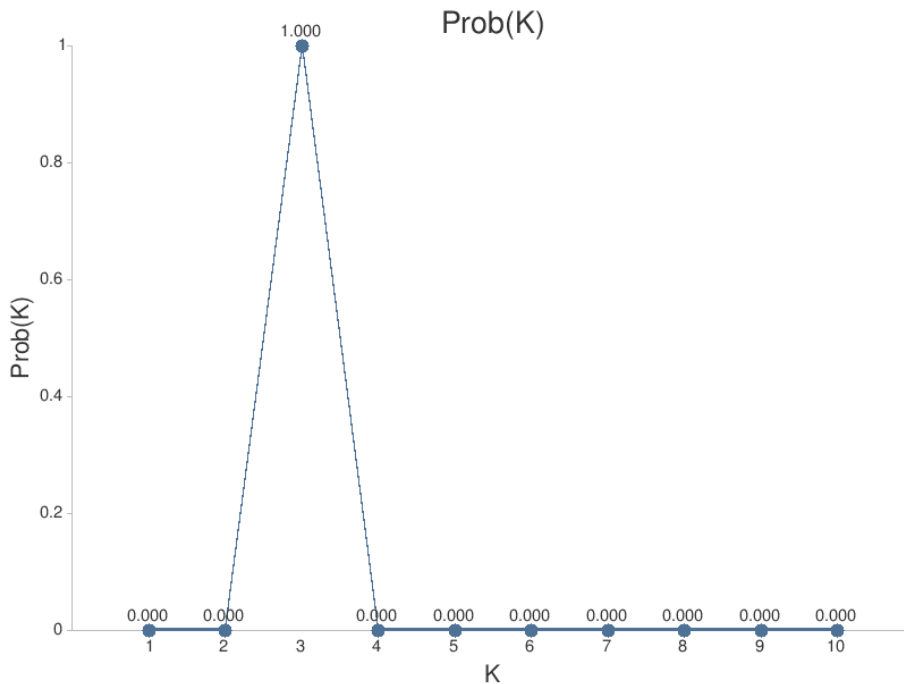


Figura 4. Gráfico da análise de probabilidade de K para a matriz de 95% de completude.

A análise de probabilidade por K é uma técnica comum usada em análises de estrutura populacional para ajudar a determinar o número ótimo de clusters. Nesse caso, o valor mais alto da probabilidade ($\Pr(K=k)$) ocorreu quando K é igual a 3, levando em conta os valores medianos de $\ln(\Pr \text{ Data})$, sugerindo que a estrutura genética ou de agrupamento dos dados é mais provável de ser dividida em três clusters genéticos distintos (ou três subpopulações diferentes). Levar em conta um $K=3$, sugere uma estrutura mais complexa subjacente aos dados, com a presença de três categorias distintas que explicam as variações genéticas, que é o mais provável, levando em consideração a complexidade das espécies. Também indica que atualmente as populações podem estar sendo agrupadas filogeneticamente de modo incorreto, revelando uma necessidade de análises estatísticas complementares.

Resultados da matriz com 90% de completude

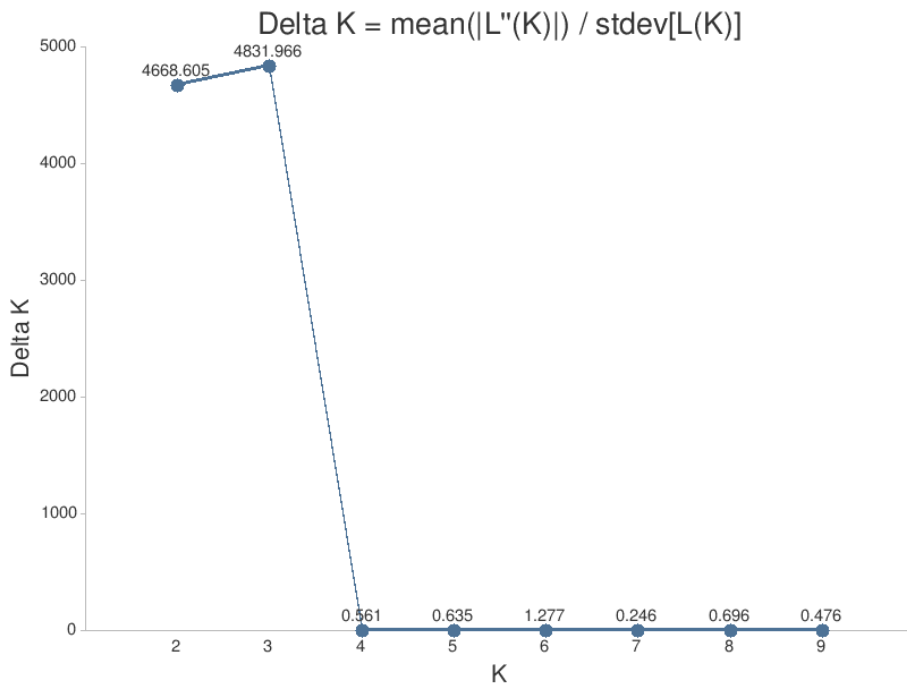


Figura 5.

Comentado [UC10]: Tem que adicionar legenda para cada figura incluída e mencionar a figura no texto!

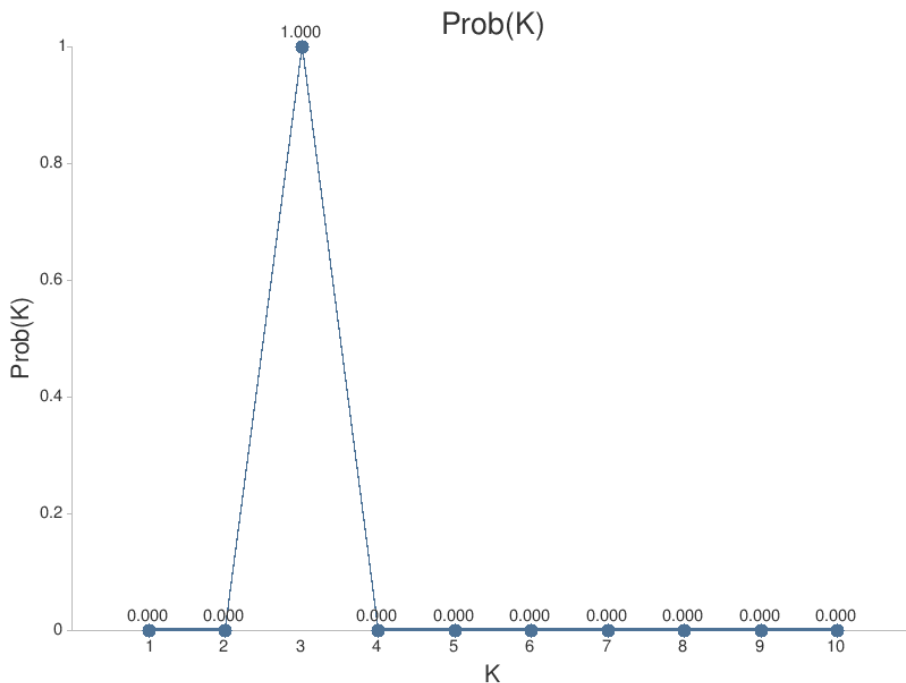


Figura 6. O resultado desta segunda matriz difere apenas no valor de ΔK , o resultado encontrado foi o valor mais provável para o número de clusters (K) que melhor se ajusta aos dados é 3.

Comentado [UC11]: Tem que adicionar legenda para cada figura incluída e mencionar a figura no texto!

De acordo com Vähä e Primmer (2005) o parâmetro para identificar o indivíduo como “puro” é o de no mínimo 90% de pertencimento ao cluster da respectiva espécie.

Comentado [UC12]: Esta frase está “perdida no texto”, precisa colocar em contexto.

5. Discussão e Conclusões

A identificação da estrutura genética de populações usando dados de genótipo de múltiplos loci tornou-se um componente central da análise genética populacional moderna. A aplicação de programas de agrupamento baseados em modelos frequentemente envolve várias etapas, nas quais o usuário considera diferentes pressupostos de modelagem, compara resultados em diferentes valores pré-determinados do número de agrupamentos assumidos (um parâmetro geralmente denotado como K), examina múltiplas execuções independentes para cada valor fixo de K e distingue entre execuções pertencentes a soluções de agrupamento substancialmente distintas.

Comentado [UC13]: Essa parte pode ir para a parte da discussão, quando for iniciar a discussão sobre as análises do STRUCTURE

A caracterização da estrutura genético-populacional das amostras incluídas no estudo não corroboram com a atual divisão do grupo em cinco espécies distintas (Carneiro et al. 2018).

Uma dúvida que surgiu após a criação da árvore filogenética de *Callicebus* durante o programa PIBIC 2021-2022 foi a identificação de *Callicebus barbarabrownae* e *Callicebus coimbrai* como grupos parafiléticos, ou sejam, como se fossem uma única espécie.

Após as análises do STRUCTURE e formação dos clusters os resultados apontam que eles pertencem a um único cluster em todas as amostras, o que aponta para uma dúvida da identificação específica para estas duas espécies.

O $k=3$ aponta que é possível que a antiga classificação estivesse correta.

6. Recomendações para o manejo

O desenvolvimento e resultados do presente estudo devem refinar as análises evolutivas e, principalmente, identificar potenciais híbridos. A identificação ou não de híbridos fornecem uma maior segurança e ajuda na tomada de decisão que subsidiem o manejo *in situ* e *ex situ* da espécie, se necessários.

7. Agradecimentos

Às agências fomentadoras ICMBIO/CNPQ que forneceu a bolsa de Iniciação Científica a LAOT e a Fundação Nacional de Ciências (National Science Foundation - NSF, EUA) que concedeu financiamento pelo programa DDRI Bio Anthro (Award#1650844) a ABM.

8. Cronograma de Conclusão do Plano de Trabalho

Etapa 3 - Realização de caracterização genética de indivíduos para verificar a presença de animais híbridos;

Etapa 4 - Realização das análises de Inferência Bayesiana e de delimitação de espécies por estatística de Fator de Bayes;

Etapa 5 - Redação de relatórios.

Comentado [UC14]: Na realidade as etapas estão ao contrário, o STRUCTURE vem antes, porque se tiver indivíduo híbrido, nós devemos retirar da matriz para rodar as análises filogenéticas.

Etapa	Mês 7	Mês 8	Mês 9	Mês 10	Mês 11	Mês 12
1						
2						
3	X	X	X			
4			X	X		
5					X	X

9. Citações e referências bibliográficas

Afonso Silva, A.C.; Santos, N.; Ogilvie, H.A.; Moritz, C. 2017. Validation and description of two new northwestern Australian Rainbow Skinks with multispecies coalescent methods and morphology. PeerJ : e3724.

Baker, R.J.; Bradley, R.D. 2006. Speciation in mammals and the genetic species concept. *Journal of Mammalogy*.87: 643-662.

Bloch, J.I.; Woodruff, E.D.; Wood, A.R.; Rincon, A.F.; Harrington, A.R.; Morgan, G.S.; et al. 2016. First North American fossil monkey and early Miocene tropical biotic interchange. *Nature* 533: 243-246.

Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; et al. 2019.

BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 15: e1006650.

Byrne, H.; Rylands, A.B.; Carneiro, J.C.; Alfaro, J.W.L.; Bertuol, F.; da Silva, M.N.F.; et al. 2016. Phylogenetic relationships of the New World titi monkeys (*Callicebus*): first appraisal of taxonomy based on molecular evidence. *Frontiers in Zoology* 13.

Cabrera, A. 1957. Catálogo de los mamíferos de América del Sur I (Metatheria, Unguiculata, Carnivora). In: Revista del Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" e Instituto Nacional de Investigación de las Ciencias Naturales. Ciencias zoológicas, t. 4, no. 1-2. Impr. y Casa Editora "Coni," Buenos Aires, 1-307p.

Carneiro, J.; Sampaio, I.; de Sousa e Silva-Júnior, J.; Farias, I.; Hrbek, T.; Pissinatti, A.; et al. 2018. Phylogeny, molecular dating and zoogeographic history of the titi monkeys (*Callicebus*, Pitheciidae) of eastern Brazil. *Molecular Phylogenetics and Evolution* 124: 10-15.

Defler, T.R. 2004. *Primates of Colombia*. Conservation International, Washington, DC.

Eaton, D.A.R.; Overcast, I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*

Elliot, D.G. 1913. *A review of primates*. New York: American Museum of Natural History.

Evanno, G.; Regnaut, S.; Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611-2620.

Groves, C.P. 2001. *Primate taxonomy*. In: *Smithsonian Series in Comparative Evolutionary Biology*. Smithsonian Institution Press, Washington, DC, 350p.

Grummer, J.A.; Bryson, R.W.; Reeder, T.W. 2014. Species Delimitation Using Bayes Factors: Simulations and Application to the *Sceloporus scalaris* Species Group (Squamata: Phrynosomatidae). *Systematic Biology* 63: 119-133.

Hershkovitz, P. 1963. A systematic and zoogeographic account of the monkeys of the genus *Callicebus* (Cebidae) of the Amazonas and Orinoco River basins *Mammalia* 27

Hershkovitz, Philip. 1990. Titis, new world monkeys of the genus *Callicebus* (Cebidae, Platyrrhini) : a preliminary taxonomic review / Philip Hershkovitz. Field Museum of Natural History, Chicago, III. :

Hill, W.C.O. 1960. Primates: comparative anatomy and taxonomy : a monograph. In: IV. Cebidae. Part A. Edinburgh: At the University Press.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405.

Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermin, L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587-589.

Kay, R.F. 2015. Biogeography in deep time - What do phylogenetics, geology, and paleoclimate tell us about early platyrrhine evolution? *Molecular Phylogenetics and Evolution* 82: 358-374.

Kobayashi, S. 1995. A phylogenetic study of titi monkeys, genus *Callicebus*, based on cranial measurements: I. Phyletic groups of *Callicebus*. *Primates* 36: 101-120.

Kobayashi, S.; Langguth, A. 1999. A new species of titi monkey, *Callicebus Thomas*, from north-eastern Brazil (Primates, Cebidae). *Revista Brasileira de Zoologia* 16: 531-551.

Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179-1191.

Lima, M.G.M.; Silva-Júnior, J. de S. e; Cerny, D.; Buckner, J.C.; Aleixo, A.; Chang, J.; et al. 2018. A phylogenomic perspective on the robust capuchin monkey (*Sapajus*) radiation: First evidence for extensive population admixture across South America. *Molecular Phylogenetic and Evolution* 124: 137-150.

Leaché, A.D.; Fujita, M.K.; Minin, V.N.; Bouckaert, R.R. 2014. Species Delimitation using Genome-Wide SN Data. *Systematic Biology* 63: 534-542.

Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; et al. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37. 1530_1531