

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



**Relatório de Final**  
(Ciclo 2021-2022)

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

**Nome do(a) estudante de IC: Letícia de Alcântara Oliveira Trajano**

**Orientador(a): Amely B. Martins**

**João Pessoa**  
**Agosto/2022**

**Comentado [UC1]:** A árvore filogenética não foi colocada no relatório  
Este é o principal resultado!!!!

## 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. Porém, existem divergências quanto ao nível taxonômico das formas reconhecidas para o gênero enquanto espécies ou subespécies.

O presente estudo realizou nova inferência filogenética para o gênero, ampliando o número de indivíduos amostrados e incluindo animais de vida livre, garantindo, assim, maior robustez na reconstrução da história evolutiva de *Callicebus*. Utilizamos o programa *ipyrad* para realização da genotipagem dos fragmentos de ADN, obtidos pela técnica de ddRAD-seq, a partir de mapeamento ao genoma do *Plecturocebus donacopilus*. Esse é o primeiro delineamento de marcadores genômicos com indivíduos representativos de todas as espécies do gênero *Callicebus*, a genotipagem gerou uma matriz de dados com 483.532 SNPs (75% de completude). As análises filogenéticas foram realizadas por meio de análises de Máxima Verossimilhança, executadas no programa IQ-TREE e posterior construção gráfica da árvore no *Toytree*. A reconstrução filogenética foi alcançada com sucesso, tendo sido testados 286 modelos evolutivos diferentes, com o programa Model Finder, utilizando os critérios de Informação de Akaike (AIC), Akaike corrigido (AICc) e o Critério de Informação Bayesiana (BIC), que apontaram para o mesmo modelo evolutivo ideal: TVM+F. A árvore filogenética construída indica que dentro do gênero, a espécie *C. nigrifrons* foi a primeira a divergir, seguido por *C. personatus*, *C. melanochir*, tendo como últimas espécies a divergirem *C. barbarabrownae* e *C. coimbrai*.

Palavras-chave: Filogenia, Guigós, genética para conservação.

Comentado [UC2]: retirar parágrafo

## Abstract

The genus *Callicebus* is endemic to Brazil, and has five currently recognized species: *Callicebus personatus* Geoffroy 1812, *Callicebus melanochir* Wied Neuwied 1820, *Callicebus nigrifrons* Spix 1823, *Callicebus barbarabrownae* Hershkovitz 1990 and *Callicebus coimbrai* Kobayashi and Langguth 1999. However, there are divergences as to the level taxonomy of the forms recognized for the genus as species or subspecies. The present study carried out a new phylogenetic inference for the genus, expanding the number of individuals sampled and including free-living animals, thus ensuring greater robustness in the reconstruction of the evolutionary history of *Callicebus*. We used the reference pipeline (ipyrad) to carry out genotyping from the mapping of DNA fragments, obtained by the ddRAD-seq technique, to the genome of *Plecturocebus donacopilus*. This is the first design of genomic markers with representative individuals of all species of the genus *Callicebus*, the genotyping generated a data matrix with 483,532 SNPs (75% completeness). of the IQ-TREE program and subsequent graphic construction of the tree in Toytree. The phylogenetic reconstruction was successfully achieved, with 286 different evolutionary models being tested by the Model Finder, using the Akaike Information Criterion (AIC), Corrected Akaike (AICc) and the Bayesian Information Criterion (BIC), which pointed to the same ideal evolutionary model: TVM+F. The phylogenetic tree constructed indicates that within the genus, the species *C. nigrifrons* was the first to diverge, followed by *C. personatus*, *C. melanochir*, with *C. barbarabrownae* and *C. coimbrai* as the last species to diverge.

Keywords: Phylogeny, Guigós, Conservation genetics.

### Lista de Figuras

Figura I – Distribuição das espécies *Callicebus* atualmente conhecidas .....  
Figura II – Arvore reconstruída

Comentado [UC3]: É para ser o número da página aqui?

Comentado [UC4R3]: Precisa formatar corretamente

### Lista de Tabelas

Tabela I – Amostras coletadas .....11  
Tabela II - Sumário de dados das amostras .....13

### Lista de Abreviaturas e Siglas

SNP - Single Nucleotide Polymorphism; Polimorfismo de nucleotídeo único

RADseq - Restriction site-associated DNA sequencing; Sequenciamento de fragmentos de DNA associado a sítios de restrição

ddRADseq - *Double-digest Restriction site-associated DNA sequencing*; Sequenciamento de fragmentos de DNA associado a sítios de digestão dupla por enzimas de restrição

ADN - Ácido desoxirribonucleico

TACC- Texas Advanced Computing Center

Comentado [UC5]: Tem fontes diferentes. Precisa padronizar. Verificar ao longo do texto.

**Sumário**

<b>1 INTRODUÇÃO.....</b>	<b>6</b>
<b>2 OBJETIVOS.....</b>	<b>9</b>
<b>3 MATERIAL E MÉTODOS</b> .....	<b>9</b>
<b>4 RESULTADOS.....</b>	<b>12</b>
<b>5 DISCUSSÃO E CONCLUSÕES</b> .....	<b>14</b>
<b>6 RECOMENDAÇÕES PARA O MANEJO.....</b>	<b>15</b>
<b>7 AGRADECIMENTOS.....</b>	<b>15</b>
<b>8 REFERÊNCIAS</b> .....	<b>16</b>

Comentado [UC6]: Números de páginas precisam estar formatados corretamente!

## 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, 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.

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*.

Comentado [UC7]: Abaixo está como Figura 1  
Padronizar!

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 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 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.

Compreender a história evolutiva de grupos de animais, descrever a biodiversidade existente e delimitar espécies são temas de grande importância para a biologia evolutiva, mas também para a conservação da biodiversidade (Rylands *et al.* 200, Baker e Bradley 2006). O nosso estudo amplia o campo de observação e pode inferir novas descobertas, úteis para programas de manejo e conservação das espécies-alvo.

Figura 1: Distribuição das cinco espécies do gênero *Callicebus* atualmente reconhecidas.

## 2. Objetivos

### Geral

Realizar análise filogenética do gênero *Callicebus*, ampliando o número de indivíduos amostrados e incluindo amostras de animais de vida livre.

### Específicos

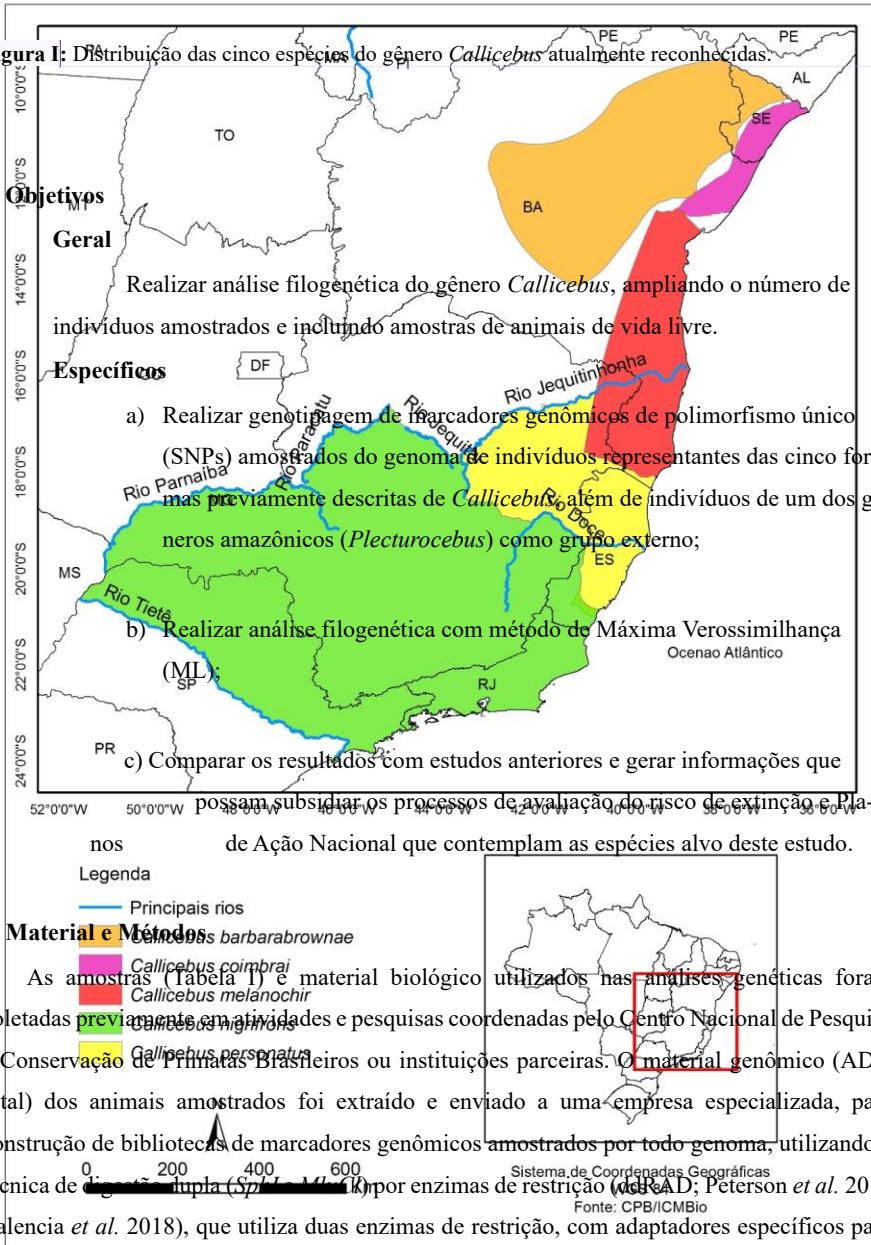
a) Realizar genotipagem de marcadores genômicos de polimorfismo único (SNPs) amostrados do genoma de indivíduos representantes das cinco formas previamente descritas de *Callicebus*, além de indivíduos de um dos gêneros amazônicos (*Plecturocebus*) como grupo externo;

b) Realizar análise filogenética com método de Máxima Verossimilhança (ML);

c) Comparar os resultados com estudos anteriores e gerar informações que possam subsidiar os processos de avaliação do risco de extinção e Planos de Ação Nacional que contemplam as espécies alvo deste estudo.

## 3. Material e Métodos

As amostras (Tabela 1) e material biológico utilizados nas análises genéticas foram coletadas previamente em atividades de 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 (*Spl*I-*Msp*II) por enzimas de restrição (deRAD; 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



Comentado [UC8]: No texto há referência a esta figura como I e 1 - padronizar.



P1, específicos para a plataforma de sequenciamento utilizada, além de marcadores (*barcodes*) específicos para cada amostra, e selecionados por tamanho de fragmento (~300±20pb). Posteriormente foram ligados adaptadores P2 com final divergente nos fragmentos com e sem adaptadores P1, esses fragmentos são amplificados por PCR com oligonucleotídeos específicos para P1 e P2, o que significa que apenas fragmentos com adaptadores P1 e P2 são amplificados, ou seja, os fragmentos que contém os sítios de restrição.

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](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 *Plecturocebus donacopilus*.

As bibliotecas de marcadores genômico já foram sequenciadas pelo método de sequenciamento de nova geração na plataforma de sequenciamento Illumina HiSeq 4000, com extensão de 2x150 pb e mínimo de 3 milhões de fragmentos para cada amostra.

O DNA não utilizado foi devolvido à pesquisadora Amely B. Martins e se encontra no laboratório onde a orientadora desenvolveu seu doutorado, na Universidade do Texas em Austin, aguardando oportunidade para repatriação. As bibliotecas de marcadores genômicos foram apenas utilizadas para sequenciamento, não havendo material restante para armazenamento.

Todas as sequências geradas se encontram em HDs da orientadora e ainda em backup em um servidor de armazenamento (RANCH – TACC - <https://portal.tacc.utexas.edu/user-guides/ranch>). As etapas de sequenciamento e análises genéticas foram conduzidas em laboratórios da Universidade do Texas em Austin, EUA e já estão disponíveis para execução deste plano de trabalho. Assim, todas as etapas de coleta de amostras, extração de ADN, preparação de bibliotecas genômicas e sequenciamento já foram realizadas previamente e os dados já estavam disponíveis para a realização deste estudo.

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). As análises filogenéticas foram conduzidas utilizando os marcadores moleculares de SNP concatenados através da análise de Máxima Verossimilhança, conduzida no programa *IQ-TREE* (v 2.2.0) (Nguyen *et al.* 2015, Minh *et al.* 2020), com réplicas de 200 e 1000 bootstraps. Para seleção do modelo evolutivo foi utilizado o

**Comentado [UC9]:** Esta parte é MATERIAL e MÉTODOS - deve ser colocada nesta parte.

**Comentado [UC10]:** Precisa colocar a versão do *ipyrad* que foi utilizada. Para saber, basta ir no TACC e rodar o comando:  
`ipyrad`

vai aparecer a versão do programa, que deve ser colocada no texto, após o nome do software.

**Comentado [UC11]:** Incluir a versão do programa aqui também!

aplicativo ModelFinder (Kalyaanamoorthy *et al.* 2017) implementado no programa IQ-TREE (Minh *et al.* 2020), considerando o Critério de Informação de Akaike corrigido (AICc), e complementarmente os critérios de Akaike(AIC) e Inferência Bayesiana. (BIC).

Comentado [UC12]: Pode acrescentar os demais, como está no resumo.

As etapas de bioinformática preparatórias (avaliação da qualidade, filtragem, alinhamento, genotipagem dos fragmentos de ADN e identificação de polimorfismos de nucleotídeo único) foram desenvolvidas no Cluster Lonestar 6 (<https://portal.tacc.utexas.edu/user-guides/lonestar6>), do Texas Advanced Computing Center (TACC). Por fim, as análises filogenéticas - usando análise de Máxima Verossimilhança e seleção do modelo evolutivo também foram desenvolvidas no TACC.

Comentado [UC13]: Você rodou alguma análise no CIPRES????  
As análises filogenéticas também foram feitas no TACC!

Inicialmente, a visualização gráfica da árvore filogenética gerada foi realizada no FigTree, (v. 1.4.4.), um visualizador gráfico de árvores filogenéticas e programa para produção de figuras prontas para publicação. A construção gráfica final da árvore filogenética foi realizada no Toytree, uma biblioteca de manipulação e plotagem de árvores em Python.

Ademais, foram realizadas reuniões semanais de acompanhamento e treinamento, com discussão de artigos e estudo bibliográfico e teórico, para aprendizado da metodologia. Além de artigos, estudamos também o livro *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (Philippe Lemey; Marco Salemi; Anne-Mieke Vandame).

**Tabela I:** Amostras utilizadas no presente estudo.

	<b>Amostra</b>	<b>Espécie</b>	<b>Origem</b>	<b>Precedência*1</b>
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
33	CPRJ2804	<i>Plecturocebus dubius</i>	Cativeiro	CPRJ

\*<sup>1</sup>CPB/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.

#### 4. Resultados

A construção da árvore filogenética envolveu as seguintes etapas: (1) Seleção dos marcadores genômicos (SNPs), (2) Alinhamento de múltiplas sequências, (3) Seleção do modelo de evolução, (4) Construção da árvore filogenética, (5) Avaliação da confiabilidade da árvore. Usamos sequências de DNA obtidas de trinta e três espécimes de todas as cinco espécies de *Callicebus* para confirmar a monofilia do gênero e inferir as relações filogenéticas entre as espécies. O desenvolvimento do projeto gerou, pela primeira vez, marcadores genômicos de indivíduos representativos de todas as espécies do gênero *Callicebus*, sendo utilizadas nas etapas das análises filogenéticas, e podendo ser utilizadas em futuros estudos populacionais ou outros de interesse para a conservação das espécies ameaçadas do gênero. É importante ressaltar

ainda, que durante as reuniões de acompanhamento e estudo do material teórico, discutimos acerca de algumas teorias e visitamos conceitos fundamentais para o andamento do projeto, como modelos utilizados evolutivos e análises estatísticas para sua escolha (hLRT'S, AIC e BIC), relógios moleculares, topologia de árvores, distâncias genéticas e modelos de substituição de nucleotídeos, inferência filogenética baseada em métodos de distância, inferência filogenética usando métodos de máxima verossimilhança e análise filogenética bayesiana.

Após rodar o *ipyrad* para realizar a avaliação da qualidade, filtragem, alinhamento e genotipagem dos fragmentos de ADN sequenciados para identificação de polimorfismos de nucleotídeo único (SNPs), foi recuperada uma matriz de dados com 483.532 SNPs no total, com apenas 24.89% de locus ou dados incompletos (*missing data*), que podem ser resultado de mutações em sítios de reconhecimento de enzimas de restrição ou de abandono alélico em certas amostras (Valencia *et al.* 2018). Como esperado para a técnica utilizada (Valencia *et al.* 2018, Peterson *et al.* 2012), houve grande variação no número de locus recuperados para cada amostra na matriz final de dados, variando entre <1000 e 43184 (CPB247). Os dados sobre o número de fragmentos de ADN retidos após as filtragens, número de *clusters* e locus finais obtidos para cada amostra estão discriminados na tabela II.

Para melhorar e refinar as análises filogenéticas, decidimos excluir as amostras “CPB215” (8 *loci* recuperados), “COI2” (158 *loci* recuperados) e “CPRJ2389” (328 *loci* recuperados) do presente estudo, pois obtiveram menos de 400 *loci* recuperados na matriz final de dados, o que julgamos ser insuficientes para uma primeira análise filogenética. É importante frisar que estes animais haviam sido encontrados já em estado de decomposição e, portanto, era esperado este resultado em virtude da alta degradação do ADN das amostras.

Polimorfismos de Nucleotídeo Único (SNPs) podem ser originados de mutações pontuais no DNA como as transições e transversões. Embora, em princípio, em cada posição da sequência de DNA seja possível ocorrer as quatro bases nucleotídicas, na prática os SNPs em organismos diploides são geralmente considerados bialélicos. Uma das razões para isso é a baixa frequência de substituições de nucleotídeo único que originaram os SNPs, estimada entre  $1 \times 10^{-9}$  e  $5 \times 10^{-9}$  por nucleotídeo por geração nas posições neutras em mamíferos (Li *et al.*, 1981; Vignal *et al.*, 2002). Dessa forma, a probabilidade de ocorrer duas mudanças independentes da base nucleotídica em uma única posição é muito baixa. (Vignal *et al.*, 2002). Entretanto, eles são abundantes e amplamente distribuídos nos genomas, podendo estar presentes em praticamente todos os loci gênicos, o que representa grande vantagem análises genéticas.

Comentado [UC14]: Fontes variadas ao longo do texto. Padronizar!

Comentado [UC15]: Estes não são modelos, são análises estatísticas para nos auxiliar a escolher o melhor modelo!

O uso de algoritmos de máxima verossimilhança (ML) no desenvolvimento de hipóteses filogenéticas estima os parâmetros desconhecidos de um modelo de probabilidade e atualmente é amplamente usado para a construção de árvores filogenéticas por sua maior capacidade computacional. Para ser utilizado, requer um modelo de evolução. Foram testados 286 modelos evolutivos diferentes, pelo *ModelFinder*. As estatísticas utilizando o Critério de Informação de Akaike (AIC), o Critério de Informação de Akaike corrigido (AICc) e o Critério de informação bayesiana (BIC) apontaram para o mesmo modelo ideal: TVM+F.

Essa escolha de modelo foi sustentada por três modelos estatísticos, o Critério de Akaike(AIC), o Critério de Akaike corrigido(AICc) e por Inferência Bayesiana (BIC). O Critério de Informação de Akaike (AIC) admite a existência de um modelo “real” que descreve os dados que é desconhecido, e tenta escolher dentre um grupo de modelos avaliados. O Critério Bayesiano (BIC) tem como pressuposto a existência de um “modelo verdadeiro” que descreve a relação entre a variável dependente e as diversas variáveis explanatórias entre os diversos modelos sob seleção. Assim o critério é definido como a estatística que maximiza a probabilidade de se identificar o verdadeiro modelo dentre os avaliados. O modelo com menor BIC é considerado o de melhor ajuste. (Gomes & Regazzi).

Comentado [UC16]: Inserir a referência utilizada!

A máxima verossimilhança é um método estatístico que estima os parâmetros desconhecidos de um modelo de probabilidade. O método de máxima verossimilhança é atualmente amplamente usado para a construção de árvores filogenéticas por maior capacidade computacional. A máxima verossimilhança avalia a probabilidade do modelo evolucionário selecionado preveja a as sequências observadas (Choudhuri,2014). Em outras palavras, a topologia das árvores filogenéticas construídas usando esse método deve obter a maior probabilidade de produzir as sequências observadas. O uso de algoritmos de máxima verossimilhança no desenvolvimento de hipóteses filogenéticas requer um modelo de evolução. A família de modelos GTR ( Xia X. & Xie Z. 2001) é a mais complexa e admite oito diferentes parâmetros, que são as seis possibilidades de substituição entre as bases (A↔G; A↔C; A↔T; G↔C; G↔T; C↔T), além de levar em conta as desigualdades na composição de nucleotídeos entre as sequências. O método assume diferentes taxas de substituição para cada par de nucleotídeos, além de assumir diferentes frequências de ocorrência de nucleotídeos. Para esses modelos, as frequências de nucleotídeo são estimadas pelas frequências observadas no alinhamento. O modelo escolhido para reconstrução da árvore foi TVM+F, que é da família dos

modelos General Time Reversible (GTR). A primeira parte (TVM) significa que é um modelo de transversão, isso indica que, bioquimicamente no processo de evolução da espécie, as purinas vêm sofrendo maior taxa de mutação e se tornando pirimidina, ou vice-versa. Ou seja, a Citosina (C) ou Timina (T) se tornam uma Adenina (A) ou Guanina (G), ou no sentido inverso, um A ou G se torna uma C ou T. A segunda (+F) indica que a frequência de bases é desigual (Kalyaanamoorthy,2017).

A família de modelos aninhados General Time Reversible (GTR) frequentemente usada engloba 64 modelos com diferentes combinações de parâmetros para substituição de sítios de DNA, onde as substituições são agrupadas hierarquicamente: simples, substituição geral de bases, transições e transversões, transições de purina para purina e pirimidina para pirimidina e transversões AC/GT e AT/CG. Um desses modelos é o de transversão (TVM) onde as frequências de base são variáveis, as taxas de transversão são variáveis, e as taxas de transição são iguais. A frequência de alinhamento indica que as bases foram contadas diretamente do alinhamento. (IQ-TREE)

Os resultados do estudo confirmaram a validade de todas as espécies atualmente reconhecidas. Entretanto, as amostras de *C. barbarabrownae* e *C. coimbrai* foram reconstruídas como parafiléticas. As amostras COINJUCO (*Callicebus coimbrai*), CPRJ 2381 (*Callicebus personatus*) e CPRJ 2144 (*Callicebus barbarabrownie*) foram excluídas após a primeira leitura da árvore gerada e uma nova análise de máxima verossimilhança foi rodada, novamente com 200 e 1000 réplica de bootstraps, a árvore foi sustentada pelos dois valores, se mantendo igual. A árvore encontrada na Figura II equivale a de 1000 réplicas de bootstrap.

De acordo com Felsenstein (1985), o bootstrap é o teste mais utilizado para se avaliar a confiança numa filogenia molecular. Ele permite avaliar a confiança no suporte de cada nó da topologia escolhida, mediante repetição da análise filogenética sobre pseudorréplicas do alinhamento original. As pseudorréplicas têm o mesmo tamanho do alinhamento original e suas colunas são sorteadas dentre as colunas do alinhamento original, com possibilidade de repetição, ou seja, uma coluna do alinhamento original pode ser sorteada mais de uma vez. O número de sorteios é igual ao número de sítios. Desta forma, alguns sítios serão amostrados mais de uma vez, enquanto outros estarão fora

pseudo-alinhamento. Portanto, árvores-réplicas são construídas, cada uma baseada em conjunto de dados diferentes. No final, o teste compara cada uma das árvores-réplicas com a árvore

Comentado [UC17]: REFERÊNCIA!!

Comentado [UC18]: Referência!!

Comentado [CL19]: Não entendi, a referência é o IQTREE?  
O IQTREE deve usar alguma referência teórica, não me parece correto referenciar o IQTREE.

original e o valor do bootstrap obtido corresponde ao percentual de vezes que o grupamento foi recuperado nas árvores-réplicas (Matioli, 2001). Geralmente, esse valor é dado em porcentagens sobre os respectivos nós das árvores filogenéticas, como é possível observar na figura III. Os valores obtidos são considerados confiáveis, estando acima de 70% (Hillis D.M. & Bull J.J. 1993.).

**Tabela II:** Sumário dos dados obtidos na etapa de avaliação da qualidade, filtragem, alinhamento e genotipagem dos fragmentos de ADN sequenciados realizada no *ipyrad*.

Amostras	Num. Fragmentos	Num. fragmentos após filtragem	% fragmentos após filtragem	Fragmentos mapeados no genoma de referência	Total de clusters	Clusters com mínimo de 6x	% de clusters com mínimo de 6x	Locus recuperados
CAQ2	3196528	3159081	98.83	2625011	174024	73703	42.35	42679
CAQ3	2178004	2156370	99.01	1404905	201168	56593	28.13	39603
CAQ4	1329724	1313940	98.81	846614	152979	44773	29.27	30713
CAQ5	6903012	6825499	98.88	6020448	204396	93534	45.76	42914
CAQ6	3635319	3597130	98.95	2294050	249974	75295	30.12	42061
CAQ7	3194286	3149589	98.60	2707348	175869	76469	43.48	42020
CNPIMENTA	1981326	1961149	98.98	1257441	191136	52422	27.43	38551
COIJUNCO	257340	253150	98.37	201760	62306	4254	6.83	1553
COIRAONI	2399381	2369398	98.75	1569931	203067	59192	29.15	40352
CPB242	4065630	4007050	98.56	2291848	287556	85563	29.76	42544
CPB243	3904349	3848691	98.57	2076079	289371	83306	28.79	41726
CPB244	3479133	3434618	98.72	2227405	251702	75186	29.87	42564
CPB245	4024458	3982799	98.96	2599927	260905	77904	29.86	43148
CPB246	2369161	2337126	98.65	1840171	196197	69764	35.56	39441
CPB247	4216213	4163162	98.74	2541459	292997	83974	28.66	43184
CPB266	2031452	2011272	99.01	1295135	196250	55179	28.12	32431
CPB332	1147678	1133733	98.78	699023	162325	37273	22.96	24242
CPB333	3274821	3226332	98.52	1982115	235761	83601	35.46	42199
CPB334	3417786	3374520	98.73	2046632	258934	73127	28.24	42496
CPB335	2639634	2606048	98.73	1705744	240885	64717	26.87	41155
CPB336	3531312	3482224	98.61	2080159	381976	83296	21.81	41469
CPB337	1972674	1943042	98.50	1256578	222666	59471	26.71	37275
CPRJ2144	2707697	2680350	98.99	1730367	233849	63751	27.26	40739



(Tabela II – Continuação)

<b>Amostras</b>	<b>Num. Fragmentos</b>	<b>Num. fragmentos após filtragem</b>	<b>% fragmentos após filtragem</b>	<b>Fragmentos mapeados no genoma de referência</b>	<b>Total de clusters</b>	<b>Clusters com mínimo de 6x</b>	<b>% de clusters com mínimo de 6x</b>	<b>Locus recuperados</b>
CPRJ2329	2408258	2377496	98.72	1311935	237727	58871	24.76	36270
CPRJ2381	455526	448983	98.56	293215	123518	9716	7.87	6237
CPRJ2465	2402013	2372217	98.76	1409112	280551	61833	22.04	31835
CPRJ2466	3495274	3444732	98.55	2009155	266647	77468	29.05	40937
CPRJ2594	581134	573493	98.69	331949	118483	14853	12.54	9052
CPRJ2764	2506500	2475351	98.76	1462924	247898	63551	25.64	31773
CPRJ2804	3922362	3882203	98.98	2485919	263862	79015	29.95	35253



Realizamos a genotipagem de marcadores genômicos de polimorfismo único (SNPs) obtidas do genoma de indivíduos representantes das cinco formas previamente descritas de *Callicebus*, além de indivíduos de um dos gêneros amazônicos (*Plecturocebus*) como grupo externo. Além disto, realizamos a análise e seleção do melhor modelo evolutivo que se aplica diretamente às matrizes de dados genômicos geradas. A seleção do modelo evolutivo foi a primeira etapa de análise para realização de reconstruções filogenéticas usando modelos de Máxima Verossimilhança. É importante também frisar que foi a primeira vez que foram feitas análises filogenéticas com animais de vida livre do gênero, e que foi a primeira vez que este tipo de marcador foi utilizado em uma pesquisa com guigós. A sequência de DNA não é só um registro de relações filogenéticas e tempos de divergência, mas também guarda marcas de como o processo evolutivo potencialmente tenha ocorrido e também o tempo decorrido no processo evolutivo das populações (Page R.D.M. & Holmes E.C. ,1998). O objetivo de filogenias moleculares é converter informações inicialmente dadas em sequências de DNA ou aminoácidos em uma árvore evolutiva. (Caldart,2016).

Árvores filogenéticas são hipóteses de relações evolutivas que podem ser representadas de diversas formas (como cladogramas - demonstram o ancestral mais próximo, filogramas - demonstram o número de substituições que ocorreram por meio de um coeficiente de similaridade, e ultramétricas - utilizadas para demonstrar o tempo evolutivo). O cladograma construído a partir da análise de máxima verossimilhança indica e reitera 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. É importante e necessário a continuação do trabalho para procurar, por exemplo, evidências genômicas sobre a história evolutiva de potenciais híbridos com outras espécies (especialmente entre indivíduos cativos) e 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, para complementação das análises filogenéticas utilizando modelo de Coalescência (MSC), através da e Análise de Quartetos (*SVDquartets*) (Eaton and Overcast

2020).

Os SNPs gerados neste estudo poderão, ainda, ser utilizados como marcadores moleculares em outros estudos em diferentes áreas, como estudos ecológicos, de mapeamento genético, análises populacionais, filogeografia (diversidade genética e estrutura populacional), estudos de associação fenótipo/genótipo e mapeamento genético.

Além dos resultados já alcançados, espera-se contribuir para o conhecimento da história evolutiva deste gênero de primatas que é endêmico do Brasil e auxiliar na delimitação do número de espécies do gênero, contribuindo para a determinação das espécies que deverão ser avaliadas no próximo ciclo de avaliação do risco de extinção dos primatas brasileiros, que será coordenada pelo Centro Nacional de Pesquisa e Conservação de Primatas Brasileiros – CPB/ICMBio.

Ademais e tendo em vista tudo que já foi apresentado, a presente pesquisa deverá contribuir para a ampliação do conhecimento sobre as relações filogenéticas do gênero *Callicebus*, que é endêmico do Brasil e possui quatro das cinco espécies reconhecidas para o gênero como ameaçadas de extinção. Primeiro, a pesquisa auxilia na geração de subsídios para a definição da lista de espécies de primatas com ocorrência no Brasil, auxiliando no processo de avaliação do risco de extinção da fauna brasileira. Segundo, os marcadores genômicos gerados nesta pesquisa poderão ser utilizados em futuros estudos que sejam de interesse dos Planos de Ação Nacional que contemplam as espécies alvo desta pesquisa (PAN-PRINE e PAN-PPMA) e tornando assim não apenas importante, mas também necessário.

## 6. Recomendações para o manejo

Recomendamos a continuação do presente estudo visando refinar as análises evolutivas e, principalmente, identificar potenciais híbridos. Ademais, recomendamos ampliar a coleta de amostras biológicas para futuros estudos de genética de populações e de genética de paisagem, que se complementem aos dados gerados no presente estudo e assim, sejam geradas informações que subsidiem o manejo *in situ* e *ex situ* da espécie, se necessários.

Comentado [UC21]: Este texto parece mais com discussão. Sugiro levar para a discussão.

Comentado [LT22]: não sei ao certo o que colocar aqui

## 7. Agradecimentos

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

## 8. Citações e referências bibliográficas

BRASIL. Portaria N° 444 , de 17 de dezembro de 2014.Ministério do Meio Ambiente, Reconhece como espécies da fauna brasileira ameaçadas de extinção aquelas constantes da "Lista Nacional Oficial de Espécies da Fauna Ameaçadas de Extinção".Diário oficial da União.Brasília, DF, n. 245, p.121. 18 de dezembro de 2014. Seção 1

Biologia molecular e evolução. . Ribeirão Preto: **Holos**.2001 Acesso em: 31 ago. 2022.

BYRNE, H.; RYLANDS, A.B.; CARNEIRO, J.C.; ALFARO, J.W.L.; BERTUOL, F.; DA SILVA, M.N.F.; **Frontiers in Zoology**.Phylogenetic relationships of the New World titi monkeys (*Callicebus*): first appraisal of taxonomy based on molecular evidence.,v.13, mar., 2016.

CAETANO,A,R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v.38, p.64-71, 2009.

CALDART,E;MATA,H;WAGECK,C;RAVAZZOLO,A. Análise filogenética: conceitos básicos e suas utilizações como ferramenta para virologia e epidemiologia molecular. **Acta Scientiae Veterinariae**, 2016.v. 44, p.1-20, ago, 2016.

CARNEIRO, J.; SAMPAIO, I.; DE SOUSA E SILVA-JÚNIOR, J.; FARIAS, I.; HRBEK, T.; PISSINATTI, A.; et al. Phylogeny, molecular dating and zoogeographic history of the

titi monkeys (*Callicebus*, Pitheciidae) of eastern Brazil. **Molecular Phylogenetics and Evolution**, v.124, p. 10–15, mar.201

CHOUDHURI, S. **Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tool**.1ª ed. Maryland,; Elsevier Inc,2014.

DAVID M. HILLIS, JAMES J. Bull. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. **Systematic Biology** , v. 42, p 182–192, jun, 1993.

DAVID POSADA, D.;CRANDALL, K.A.; Selecting Models of Nucleotide Substitution: An Application to Human Immunodeficiency Virus 1 (HIV-1), **Molecular Biology and Evolution**,V.18,,P.897–906,jun.,2001.

EATON, D.A.R.; OVERCAST, I. ipyrad: Interactive assembly and analysis of RADseq datasets. **Bioinformatics**, v. 36, p. 2592–2594, abr., 2020.

**Evolution and Genomics,Intensive and immersive training opportunities**, Nucleotide Substitution Models. Disponível em: <http://evomics.org/resources/substitution-models/nucleotide-substitution-models/>. Acesso em :01/08/22

FELSENSTEIN, JOSEPH. “Phylogenies and the Comparative Method.” **The American Naturalist**, vol. 125, no. 1, pp. 1–15,1985.

GOMES,C.;REGAZZI,A, Critérios para seleção de modelos baseados na razão de verossimilhança. Disponível em: <http://arquivo.ufv.br/dbg/resumos2008b/Resumo%20Claudomiro.htm>. Acesso em: 28/07/22

GROVES, C.P. **Primate taxonomy**. Smithsonian Series in Comparative Evolutionary Biology. Washington, DC: Smithsonian Institution Press, 2001.

- HERSHKOVITZ, PHILIP. 1990. **Titís, new world monkeys of the genus Callicebus (Cebidae, Platyrrhini)** : a preliminary taxonomic review. Chicago, Ill: Field Museum of Natural History, 1990.
- KALYAANAMOORTHY, S.; MINH, B.Q.; WONG, T.K.F.; VON HAESLER, A.; JERMIIN, L.S..ModelFinder: fast model selection for accurate phylogenetic estimates. **Nature Methods**, v.14, p.587–589, maio., 2017.
- KOBAYASHI, S. A phylogenetic study of titi monkeys, genus Callicebus, based on cranial measurements: I. Phyletic groups of Callicebus. **Primates**, v.36, p.101–120, jan., 1995.
- KOBAYASHI, S.; LANGGUTH, A. A new species of titi monkey, Callicebus Thomas, from north-eastern Brazil (Primates, Cebidae). **Revista Brasileira de Zoologia**, v.16, p. 531–551, abr., 1999.
- LI, WH., GOJOBORI, T. & NEI, M. Pseudogenes as a paradigm of neutral evolution. **Nature**, v. 292, p. 237–239, jul, 1981
- MATIOLI, S.R. **Biologia molecular e evolução**. . Ribeirão Preto: Holos. 2001
- MINH, B.Q.; SCHMIDT, H.A.; CHERNOMOR, O.; SCHREMPF, D.; WOODHAMS, M.D.; VON HAESLER, A.; et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. **Molecular Biology and Evolution** v.37,p. 1530–1534, maio., 2020.
- MOURA,A,R. **Crítérios de seleção de modelos: um estudo comparativo**.Tese(Mestrado em Modelagem Matemática e Computacional) -Universidade Federal da Paraíba.João Pessoa, p.82.2021
- NGUYEN, L.-T.; SCHMIDT, H.A.; VON HAESLER, A.; MINH, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. **Molecular Biology and Evolution**, v. 32, p.268–274, jan., 2015.

RYLANDS, A.B.; SCHNEIDER, H.; LANGGUTH, A.; MITTERMEIER, R.A.; GROVES, C.P.; RODRÍGUEZ-LUNA, E. An assessment of the diversity of New World primates. **Neotropical Primates**, n.2, v.8, p.61–83, jun.; 2000.

S. KALYAANAMOORTHY, B.Q. MINH, T.K.F. WONG, A. VON HAESLER, AND L.S. JERMIIN. ModelFinder: fast model selection for accurate phylogenetic estimates. **Nat. Methods**, v.14, p. 587–589, maio, 2017.

PAGE, R. D. M. & HOLMES, E. C.. **Molecular evolution: a phylogenetic approach**. Oxford: Blackell Science. Sunderland.1998

PETERSON, B.K.; WEBER, J.N.; KAY, E.H.; FISHER, H.S.; HOEKSTRA, H.E. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. **PLOS ONE**, v.7, maio, 2012.

VAN ROOSMALEN, M.; VAN ROOSMALEN, T.; MITTERMEIER, R. A Taxonomic Review of the Titi Monkeys, Genus *Callicebus* Thomas, 1903, With the Description of Two New Species, *Callicebus bernhardi* and *Callicebus stephennashi*, from Brazilian Amazonia. **Neotropical Primates** ,v.10, p.1–52, jun., 2002.

VIGNAL, A., MILAN, D., SANCRISTOBAL, M. *ET AL*. A review on SNP and other types of molecular markers and their use in animal genetics. **Genet Sel Evol** , v.34, p.275 - 305, fev, 2002.

VALENCIA, L.M.; MARTINS, A.; ORTIZ, E.M.; DI FIORE, A. A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. **PLOS ONE**, v.13, p.1-34, ago, 2018.