

Article

# A New Assessment of Robust Capuchin Monkey (*Sapajus*) Evolutionary History Using Genome-Wide SNP Marker Data and a Bayesian Approach to Species Delimitation

Amely Branquinho Martins <sup>1,2,\*</sup> , Mônica Mafra Valença-Montenegro <sup>1</sup>, Marcela Guimarães Moreira Lima <sup>3</sup> , Jessica W. Lynch <sup>4</sup> , Walfrido Kühn Svoboda <sup>5</sup>, José de Sousa e Silva-Júnior <sup>6</sup>, Fábio Röhe <sup>7</sup>, Jean Philippe Boubli <sup>8</sup> , and Anthony Di Fiore <sup>2,9</sup> 

- <sup>1</sup> Centro Nacional de Pesquisa e Conservação de Primatas Brasileiros, Instituto Chico Mendes de Conservação da Biodiversidade, Cabedelo 58310-000, PB, Brazil
- <sup>2</sup> Primate Molecular Ecology and Evolution Laboratory, Department of Anthropology, The University of Texas at Austin, Austin, TX 78712, USA
- <sup>3</sup> Laboratório de Biogeografia da Conservação e Macroecologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém 66077-530, PA, Brazil
- <sup>4</sup> Institute for Society and Genetics, Department of Anthropology, University of California-Los Angeles, Los Angeles, CA 90095, USA
- <sup>5</sup> Instituto Latino-Americano de Ciências da Vida e da Natureza, Centro Interdisciplinar de Ciências da Vida, Universidade Federal da Integração Latino-Americana, Foz do Iguaçu 85870-650, PR, Brazil
- <sup>6</sup> Museu Paraense Emílio Goeldi, Ministério da Ciência, Tecnologia, Inovações e Comunicações, Coordenação de Zoologia, Campus de Pesquisa, Setor de Mastozoologia, Belém 66077-830, PA, Brazil
- <sup>7</sup> Laboratório de Evolução e Genética Animal, Universidade Federal do Amazonas, Manaus 69067-005, AM, Brazil
- <sup>8</sup> School of Science, Engineering and the Environment, University of Salford, Salford M5 4WT, UK
- <sup>9</sup> Tiputini Biodiversity Station, Universidad San Francisco de Quito, Quito 170901, Ecuador
- \* Correspondence: amely.martins@icmbio.gov.br



**Citation:** Martins, A.B.;

Valença-Montenegro, M.M.; Lima, M.G.M.; Lynch, J.W.; Svoboda, W.K.; Silva-Júnior, J.d.S.e.; Röhe, F.; Boubli, J.P.; Fiore, A.D. A New Assessment of Robust Capuchin Monkey (*Sapajus*) Evolutionary History Using Genome-Wide SNP Marker Data and a Bayesian Approach to Species Delimitation. *Genes* **2023**, *14*, 970. <https://doi.org/10.3390/genes14050970>

Academic Editor: Ute Radespiel

Received: 6 March 2023

Revised: 11 April 2023

Accepted: 12 April 2023

Published: 25 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Robust capuchin monkeys, *Sapajus* genus, are among the most phenotypically diverse and widespread groups of primates in South America, with one of the most confusing and often shifting taxonomies. We used a ddRADseq approach to generate genome-wide SNP markers for 171 individuals from all putative extant species of *Sapajus* to access their evolutionary history. Using maximum likelihood, multispecies coalescent phylogenetic inference, and a Bayes Factor method to test for alternative hypotheses of species delimitation, we inferred the phylogenetic history of the *Sapajus* radiation, evaluating the number of discrete species supported. Our results support the recognition of three species from the Atlantic Forest south of the São Francisco River, with these species being the first splits in the robust capuchin radiation. Our results were congruent in recovering the Pantanal and Amazonian *Sapajus* as structured into three monophyletic clades, though new morphological assessments are necessary, as the Amazonian clades do not agree with previous morphology-based taxonomic distributions. Phylogenetic reconstructions for *Sapajus* occurring in the Cerrado, Caatinga, and northeastern Atlantic Forest were less congruent with morphology-based phylogenetic reconstructions, as the bearded capuchin was recovered as a paraphyletic clade, with samples from the Caatinga biome being either a monophyletic clade or nested with the blond capuchin monkey.

**Keywords:** Neotropical primates; phylogenomic; ddRADseq; evolutionary history

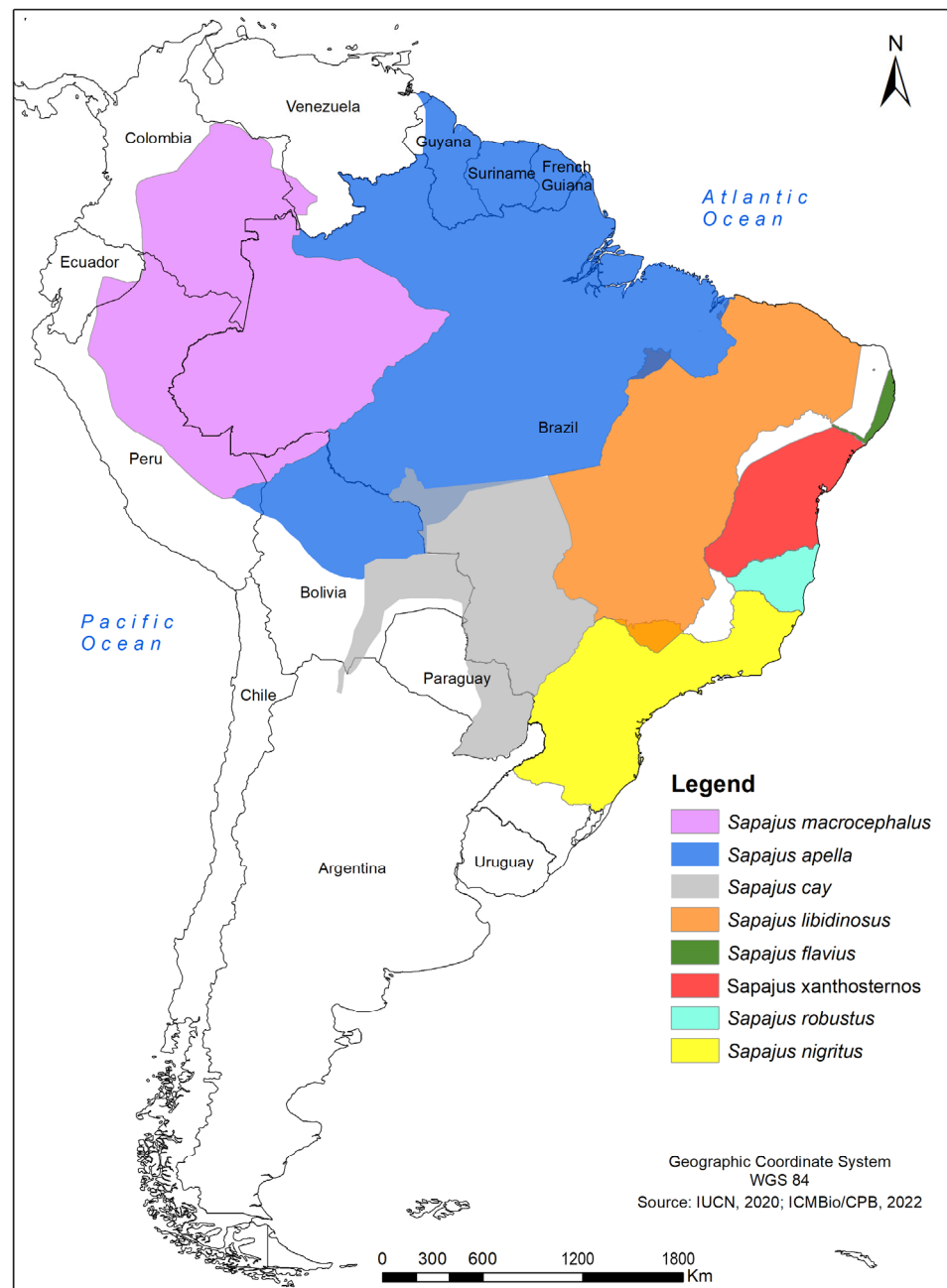
## 1. Introduction

Capuchin monkeys (genus *Cebus sensu lato*) (Erxleben, 1777) are a lineage of primates that are considered one of the groups with the most confusing taxonomy among Neotropical mammals [1,2]. The group has been divided into two genera—*Cebus* (the gracile capuchins) and *Sapajus* (the robust capuchins)—based on several traits, including genetic data [3–5].

Robust capuchin monkeys (*Sapajus*) have a widespread distribution found throughout South America, from the Colombian Llanos, through the Amazon basin, to the Cerrado, Caatinga, Atlantic Forest, and Pantanal of Brazil, and in countries of the Southern Cone [6,7].

Over the past quarter century, taxonomists and phylogeneticists have suggested a host of arrangements for robust capuchin taxonomy (Table S1) [1,3,5,7–13], with different studies dividing the genus into either four [5,11], six [13], seven [3], or eight [7,14] species. Under the broad “eight-species” classification, the taxonomic arrangement includes the following: *Sapajus xanthosternos*, the yellow-breasted capuchin, which is endemic to Brazil and found in the Atlantic Forest and Caatinga, from the south and east of the São Francisco River to the north of the Jequitinhonha River; *Sapajus robustus*, the robust tufted capuchin, which is endemic to the Brazilian Atlantic Forest, occurring at the south of the Jequitinhonha River and extending as far as the Doce River at the south and the Serra do Espinhaço mountains to the southwest; *Sapajus nigritus*, the black-horned capuchin, which occurs from the Doce River in Minas Gerais, Brazil, extending through the southern region of Brazil and the extreme northeastern tip portion of Argentina provinces of Iguazú and Misiones, at the east of the Paraná River; *Sapajus apella*, the brown capuchin, and *Sapajus macrocephalus*, the large-headed capuchin, both of which are Amazonian species; *Sapajus cay*, Azara’s capuchin, which ranges through the northern tip of Argentina, southern Bolivia, the eastern half of Paraguay, and Brazil, extending close to the Amazon in the forests of the Pantanal and the intersection of this biome and the Cerrado; *Sapajus libidinosus*, the bearded capuchin, which occurs throughout the Cerrado and Caatinga biomes of Brazil and has a wide distribution in these dry, savanna-like habitats; and *Sapajus flavius*, the blond capuchin, which has the smallest geographic range and is mostly limited to sparse remnants of the Atlantic Forest in the northeast of Brazil [7] (Figure 1). To the north and east, the Atlantic Ocean limits the distribution of this species. The western limit of its distribution is undefined, but supposedly coincides with areas of transition between the Atlantic Forest and the drier Caatinga biome [15–17], where it potentially overlaps with the distribution of the bearded capuchin (*S. libidinosus*). A lack of information about populations of wild capuchins in the Atlantic Forest–Caatinga transition has precluded researchers from clearly defining the geographic limits between *S. libidinosus* and *S. flavius*, generating uncertainties about their taxonomic identity.

As the blond capuchin monkey occurs mostly in the coastal Atlantic Forest, which is the biome where the robust capuchins are believed to have first speciated, it would be rather easy to hypothesize the blond capuchins should be more closely related, phylogenetically, to other capuchin species from the same biome, such as *S. xanthosternos*, the yellow-breasted capuchin. However, recent studies have suggested that blond capuchin monkeys either belong to the same species as, or are a sister taxon to, the bearded capuchin monkey, *S. libidinosus* from the Cerrado, and these two species are both more closely related to robust capuchin monkeys from the Amazon than to other species occurring in the Atlantic Forest [5,13]. Thus, several studies have recovered *S. flavius* as either a monophyletic clade within the widespread Amazonian group of robust capuchins (along with *S. apella*, *S. macrocephalus*, *S. cay*, and *S. libidinosus*) or as a sister taxon to *S. libidinosus*, the bearded capuchin, to the exclusion of the Amazonian clade. These studies have used sequence data from mitochondrial and certain nuclear markers (ultraconserved elements, UCEs) to infer the robust capuchin monkey phylogeny, but they only included three samples of the blond capuchin monkeys, all of which came from populations inhabiting the extreme eastern portion of the species’ distribution in the Atlantic Forest [5,13]. Additionally, limitations with the utility of both these genetic markers have been reported, especially for recently evolving lineages [18,19], and increasing taxon sampling, especially from populations living closer to the species boundaries, is sorely needed to provide a more complete picture of robust capuchin phylogenetic history [20–22]. Thus, one of the key goals of this study is to reconstruct the phylogenetic relationships and phylogeographic history of the capuchin monkeys of the northern Atlantic Forest and Caatinga biomes of Brazil and describe how they fit within the *Sapajus* phylogeny.



**Figure 1.** Map with the potential distributions of the *Sapajus* species according taxonomic arrangement with eight species [3,15].

Understanding evolutionary history and the underlying processes that drive rapid radiations is an important goal in evolutionary biology [23–26]. Therefore, given the uncertainties in the *Sapajus* phylogeny (especially within the recently evolving widespread lineages in the Amazon, Cerrado, Caatinga, and extreme northeastern Atlantic Forest of Brazil), better characterization of the evolutionary history of the *Sapajus* radiation is of great interest. However, establishing phylogenetic relationships among lineages that have undergone rapid and recent diversification is a challenge due to incomplete lineage sorting, which may be due to genes that have evolved slowly relative to the rate of speciation [27–30]. Additionally, it is especially difficult to reconstruct evolutionary history for geographically widespread lineages where there are few barriers to dispersal, where multiple zones of contact between lineages may exist, and where reproductive isolation is incomplete [31–34].

Fortunately, utilizing information from hundreds or thousands of genomic markers, such as single nucleotide polymorphisms (SNPs), modern phylogenomic analysis has the potential to decrease the impact of such difficulties. Considering the increasing availability of high-throughput sequencing technologies and their rapidly decreasing costs, it is now feasible to study the evolutionary history of lineages at the genome-wide scale for many taxa of interest, including non-model organisms [35–37]. Several methods have been developed to discover and screen a set of genome-wide markers by subsampling and sequencing just a small fraction of the genome, which, nonetheless, can include tens of thousands of variable sites [38–41]. Restriction-Site-Associated DNA sequencing (RADseq) is a term applied to a group of next-generation sequencing methods that rely on the use of one or more restriction enzymes to cleave the genome into a set of short DNA fragments flanked by the restriction sites [39], just a fraction of which are then isolated (e.g., by size) and sequenced. This process allows for genome-wide marker discovery and typing at a high coverage and low cost, favoring markers to be genotyped accurately across individuals at different population or taxonomic levels [42–44]. RADseq-based methods have been successfully used to discover thousands of SNPs in phylogenetically diverse organisms including fish [45–47], insects [48,49], birds [50], and mammals [51,52], including primates [43,53–55].

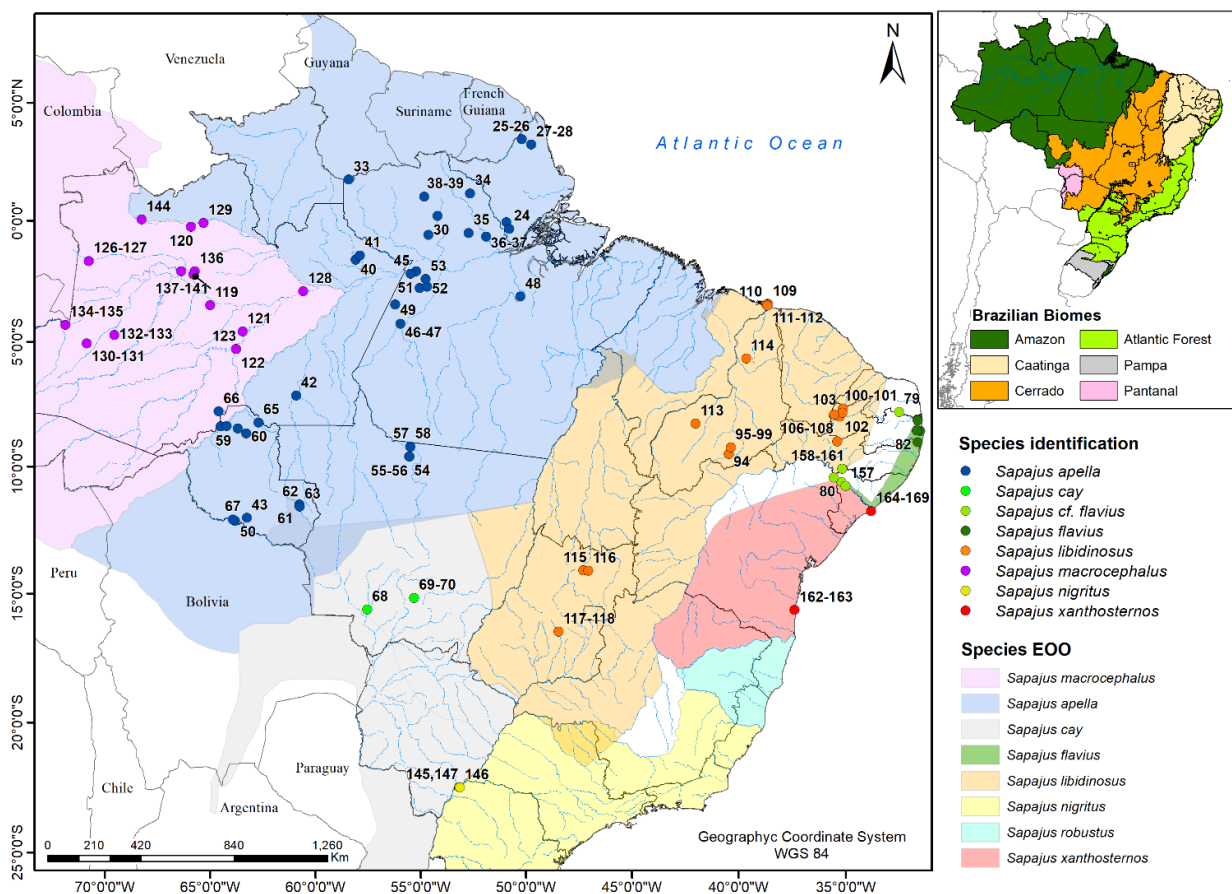
For this study, we constructed reduced representation genomic libraries using a variant of the RADseq approach known as the “double-digest Restriction-Site-Associated DNA” (ddRAD) method, coupled with Illumina high-throughput sequencing, which allowed us to genotype individual samples by sequencing short fragments of DNA flanked by a specific combination of two restriction endonuclease recognition sites [56]. This approach allowed us to genotype thousands of informative single nucleotide polymorphisms (SNPs) sampled from across the capuchin genome to infer the phylogenetic relationships among taxa in the *Sapajus* radiation and to assess the congruence of our results with previous phylogenetic studies of the genus. Our study used a wide range of samples from individuals representing all putative species of *Sapajus*, most with known provenance. In addition to a set of individuals also used in previous studies (e.g., [13]), we acquired dozens of new samples from the yellow-breasted, the bearded, and the blond capuchin monkeys, including samples from deep within the geographic ranges of each of these two later species, as well as from populations of capuchins occupying areas of the Atlantic Forest–Caatinga transition, whose taxonomic assignment is uncertain. Therefore, this study includes both a larger number of samples and samples from more localities than any previous research on *Sapajus* phylogenetics. To infer the phylogenetic relationships among our *Sapajus* samples, we used both a Maximum Likelihood (ML) and a multispecies coalescent approach grounded in quartet-based phylogenetic inference that combines information from multiple gene trees. In addition, we used a Bayes Factor (BF) validation method to test among seven alternative hypotheses of species delimitation within the *Sapajus* radiation. Finally, due to our improved sampling of the blond capuchin monkey, we also give special attention to this species and its phylogenetic relationships with its closest congener by testing alternative hypotheses regarding the species delimitation of both blond and bearded capuchin monkey lineages and by evaluating whether the presence of the blond capuchin monkey in the Caatinga biome is corroborated by our phylogenomic analyses.

## 2. Materials and Methods

### 2.1. Samples

For this study, we used samples from a total of 171 individuals (Figure 2): 149 individuals from the 8 putative species of the genus *Sapajus*, 19 individuals from 3 species of the genus *Cebus*, and 3 individuals from the genus *Saimiri* (Table S2). Provenance is known for 159 of these samples. Overall, 109 of these samples were collected by collaborators and used in previous studies [5,13,57], while 62 were newly collected for this study. For the latter set of samples, we collected blood or tissue samples from seven populations of wild blond capuchins, six populations of bearded capuchins, and one population of yellow-breasted

capuchins in a series of field seasons between 2005 and 2016. Animals were captured in tomahawk traps baited with corn and sedated with ketamine HCl (~30 mg/kg, IM) in consultation with a veterinarian. Traps were kept open during the day and monitored regularly (~every 2 h), and captured animals were processed immediately to minimize holding time. Individuals were released as soon as the effects of the sedation wore off. During processing, a wildlife veterinarian collected 3 to 5 mL of blood (depending on body mass) from the femoral vein of each individual using a 23-gauge needle. Samples were collected into color-coded vacutainers containing EDTA as an anticoagulant and then preserved on ice. At the end of each day's trapping, blood and tissue samples were transferred to  $-20\text{ }^{\circ}\text{C}$  for long-term storage. Additionally, health conditions were examined, and photographs, morphometric measurements, weight, and dental casts were collected while animals were sedated.



**Figure 2.** Map showing the sampled localities. Sample numbers correspond to those in Table S2. *Sapajus cf. flavius*—samples from populations of capuchins occupying areas of the Atlantic Forest–Caatinga transition whose taxonomic assignment as *S. flavius* is uncertain.

During each capture event, all team members involved in handling animals used gloves and face masks as precautions against disease transmission from researchers to animals and zoonotic infection. Animal immobilization procedures were conducted by Brazilian wildlife veterinarians with broad experience in primate fieldwork. Capture procedures, measurements, and sample collection used were approved by the Brazilian SISBIO/ICMBio (Research Permit Number 19927), and all protocols used were approved by the Institutional Care and Use Committee (IACUC) of the University of Texas at Austin (Protocol ID: AUP-2016-00077). Additionally, DNA, blood, or tissue samples for other *Sapajus*, *Cebus* (outgroup), and *Saimiri* (outgroup) species were obtained from collaborators (Table S2).

## 2.2. DNA Extraction and Quantification

For most samples newly collected in this study, DNA extraction was performed in the Primate Molecular Ecology and Evolution Laboratory at the University of Texas at Austin, although, for some individuals, genomic DNA was provided by colleagues. Fresh genomic DNA was extracted from the tissue and blood samples using the DNeasy Blood & Tissue Kit<sup>®</sup> (Qiagen, Germantown, MD, USA) as per the manufacturer's instructions. The nucleic acid concentration of all samples was quantified in the Institute for Cellular and Molecular Biology's DNA sequencing facility at the University of Texas at Austin using a Qubit 2.0 fluorometer (LifeTechnologies, Carlsbad, CA, USA). Most of the samples yielded sufficient genomic DNA for normalization (via dilution) to ~10 ng/ $\mu$ L for digestion and subsequent library construction.

## 2.3. Whole Genome Markers

We constructed reduced representation genomic libraries using the double-digestion Restriction-Site-Associated DNA method (ddRADseq) [56]. After normalization, a total of 100 ng of genomic DNA for each sample was first digested with two restriction enzymes, *SphI* and *MluCI*, that have been previously demonstrated as useful for discovering large numbers of variable SNP markers across platyrrhines [58]. The resultant DNA fragment libraries were then purified using AMPure Bead XP and had P1 and P2 Illumina adapters and barcodes ligated to the fragments. Samples were pooled, size selected using Sage Science Pippin-Prep to focus only on fragments of  $300 \pm 30$  bp, and re-amplified using Phusion High Fidelity PCR. Library quality was assessed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA), and samples were sequenced using the Illumina HiSeq 4000 (Illumina, Inc., San Diego, CA, USA) platform with paired-end reads of  $2 \times 150$  bp length and a minimum of 2–3 million reads for each sample. All library preparation and sequencing were performed at the University of Texas at Austin's core Genome Sequencing and Analysis Facility.

## 2.4. Quality Control

Raw sequencing reads were first quality checked using FASTQC 0.11.9 [59], which is a quality control application specifically for high-throughput sequence data. Reads were then filtered using the BBDuk software from the BBTools suite of bioinformatic tools, version 38.76 [60]. With this tool, reads were first adapter-trimmed at the 3' end using a *kmer* length of 22, allowing a maximum of 3 mismatches, and discarding any reads smaller than 30 bp. Trimming was performed with the "tbo" and "tpe" options in order to trim adapters based on pair overlap detection and to trim all reads to the same length when an adapter sequence was only detected in one read of a pair. Then, as PhiX DNA is commonly used as a spike-in control during library preparation for Illumina sequencing, all reads that mapped to the PhiX genome were filtered out. After verification of the correct pairing of R1 and R2 reads, all unpaired sequences were discarded from further analysis. Because read quality often decreased at the end of a read, reads were then trimmed from the terminal end back to the first base that had an average FASTQC quality score of  $Q < 30$ . Lastly, we discarded all reads with an average quality score of  $Q < 30$ .

All remaining raw reads were assigned to individual samples using their barcode through the deML software [61], allowing for up to one mismatch in the barcode sequence. The final set of filtered, trimmed, and assigned reads consisted of, at most, 145 bp reads, beginning with either the 4 bp *MluCI* (for the R1 read) or the 6 bp *SphI* (for the R2 read) restriction enzyme recognition sites. We then ran a second round of adapter trimming at the 5' end of each read to remove the additional 5 or 4 bases corresponding to the restriction enzyme recognition sites for the R1 and R2 reads, respectively. Finally, all bases with Phred quality scores of less than 20 were replaced with the uncalled base symbol (i.e., Ns), and reads with more than 5% Ns were discarded from further analysis. Overall, ~95% of the raw reads were kept after all trimming and filtering steps.

### 2.5. Assembly and SNP Calling

Identification of orthologous ddRAD sequences and SNP loci was performed using the software ipyrad [v.0.9.19] [62], which is a free open source tool for assembling restriction site-associated DNA sequence datasets. Using ipyrad, paired-end reads were mapped to the *Sapajus apella* reference genome [63] (NCBI, BioProjects accession PRJNA717806), identifying read copies from the same locus within samples and producing gapped alignments. Paired reads that mapped with incorrect orientation or to multiple locations (paralogous sequences) were discarded. The final set of putative loci for each sample was generated from those clusters of reads—i.e., groups of highly similar sequences mapping to the same genome location—with a sequencing depth of at least six reads ( $\geq 6\times$ ) [64]. We set the maximum number of heterozygous sites (Hs) and Ns within the consensus sequence at each locus as the upper bound of the 95% CI of these two variables estimated from the array of consensus sequences. To cluster both within and across samples, we set the clustering threshold to 85% according to previous studies (e.g., [58,65]) that have demonstrated that over-splitting tends to occur when using more stringent clustering thresholds. Such over-splitting in the identification of putative loci can be detrimental for phylogenetic inference [66]. Consensus loci found within each sample were then aligned across samples using Muscle v3.8.31 (EMBL-EBI, Hinxton, Cambridgeshire, UK) [67] to generate an initial data matrix of all putative homologous loci that were recovered in at least four individuals. We applied two additional filters to generate the final dataset, while avoiding ambiguous genotypes for each sample. First, we discarded putative loci that were heterozygous in more than 50% of individual samples to avoid potentially clustering paralogous loci rather than true heterozygous sites. Second, we excluded loci containing more than a specified maximum number of SNP sites across the entire set of samples to avoid potential effects of poor alignments in repetitive regions. The threshold for this maximum number of SNPs was set as the upper bound of the 95% CI of the distribution of the number of SNPs per locus. Bioinformatic analyses were undertaken at UT's Texas Advanced Computer Center (TACC).

Finally, to evaluate the effects of missing data due to allelic dropout, low sequence coverage, or the random effect of next generation sequencing, we created three different datasets with varying levels of locus coverage across samples. That is, we created four progressively smaller data matrices comprising all the loci that were present in at least  $\approx 30\%$ ,  $60\%$ ,  $75\%$ , and  $90\%$  of individuals.

### 2.6. Phylogenetic Analysis

We used two different methods to infer the phylogenetic relationships among the set of *Sapajus* and outgroup samples. First, we conducted a Maximum Likelihood (ML) analysis using the concatenated RAD sequence data from all loci in the final genotype matrix [68]. Second, we used a coalescent-based approach using quartet-based phylogenetic inference under a Multispecies Coalescent (MSC) theory framework [69–72].

Model selection was performed for the whole concatenated dataset using ModelFinder [73] implemented in the IQ-TREE v. 1.6.12 software [74,75], using the corrected Akaike's information criterion (AICc). All nucleotide substitution models supported by IQ-TREE were tested. We then used the best fitted substitution model for phylogenetic tree reconstruction under the ML framework in the IQ-TREE v. 1.6.12 software [74,75]. For the ML analysis, 500 initial independent searches were done on the original alignment. Each search started from 100 parsimony trees; then, the 20 best scoring parsimony trees were selected to optimize the search, and only the 5 top scoring trees were retained in the candidate set to improve ML tree search efficiency. Searches were run using default tree search parameters in IQ-TREE. Node support was assessed with 1000 standard nonparametric bootstrap replicates [76]. For the MSC analysis, we used the SVDquartets approach [72] to estimate the species tree. SVDquartets, which is a single-site quartet framework, exhaustively samples subsets of four individuals from the data matrix to produce the best quartet tree and then constructs a species tree from all sampled quartets. We implemented the

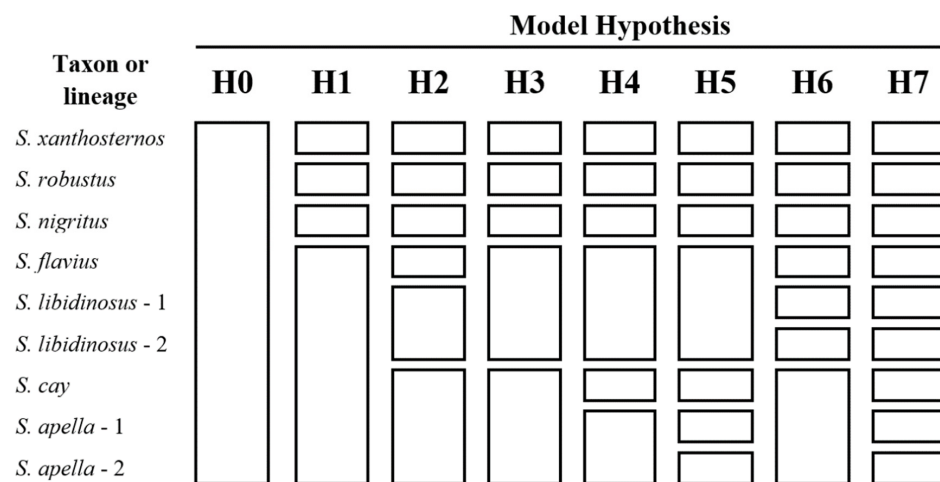
SVDquartets species tree inference using the program Tetrads v0.9.13, within the ipyrad analysis toolkit [62]. Statistics of support for each node were estimated in Tetrads through 1000 bootstrap replicates by resampling the number of loci with replacement to the same size as the original dataset. Phylogenetic analyses were undertaken using the Lonestar high-performance mainframe computing infrastructure at UT's TACC.

### 2.7. Bayesian Analysis and Divergence Dating

In order to more robustly infer species boundaries and evaluate the validity of the various monophyletic groups recovered in our ML and MSC analyses, we used a validation method to test alternative hypotheses of species delimitation based on the assignment of samples to candidate species, which is a robust approach, as it explicitly models the process of lineage diversification among a set of presumed candidate lineages [77]. Statistically delimiting species boundaries using multi-locus or genome-scale DNA sequences is increasing an objective of certain taxonomic analyses—especially for identification of cryptic species—and is a prominent research field [78–80]. In addition, model-based genome-wide approaches using the MSC framework are advantageous, as they account for coalescent processes when estimating phylogenetic relationships while, at the same time, considering variation in demographic parameters, variation in molecular sequences, and incomplete lineage sorting [79,81].

Overall, we tested eight alternative species delimitation models (H0 to H7) (Figure 3). Model H0 was our null hypothesis, which considers all *Sapajus* samples as belonging to a single species. All the other models (H1 to H7) considered, as species, *S. xanthosternos*, the yellow-breasted capuchin, *S. robustus*, the crested capuchin, and *S. nigritus*, the black-horned capuchin, with varying assumptions regarding other putative forms of *Sapajus*, taking into account the fact that the most recent phylogenetic reconstructions of the *Sapajus* radiation [5,13], as well as our own ML and MSC analyses, consistently divide the Atlantic Forest capuchin forms into three different monophyletic clades. Model H1 posits a widespread *S. apella* species, which includes all but those three forms mentioned above, in a single species, as suggested by recent mtDNA- and UCE-based analyses [5,13]. Model H2 posits *S. flavius* and *S. libidinosus*—with all samples from Cerrado and Caatinga—as separate species, plus a widespread Amazonian clade that includes *S. cay*, *S. macrocephalus*, and *S. apella*, which was suggested by SNPs recovered from a UCEs analysis [13]. Model H3 posits a situation similar to H2 but considers *S. flavius* and *S. libidinosus* as the same species. Models H4 and H5 were derived from the results of our IQ-TREE analysis; H4 is similar to H3 but considers *S. cay* as a separate species from the other Amazonian forms (*S. macrocephalus* and *S. apella*), while H5 considers as separate species all monophyletic clades found in our IQ-TREE analysis. This hypothesis places *S. flavius* with samples from the Caatinga and *S. libidinosus* (*S. libidinosus* 1 = samples from Serra da Capivara + samples from Maranhão; *S. libidinosus* 2 = samples from the state of Goiás—southern Cerrado) as one species. Finally, models H6 and H7 were derived from the results of our MSC analysis. These hypotheses, similar to models H4 and H5, also posit *S. flavius* (samples from the Atlantic Forest and a couple of localities from Caatinga) and two *S. libidinosus* species (*S. libidinosus* 1 = samples from Caatinga; *S. libidinosus* 2 = samples from Cerrado) as one species. H6 and H7 differ, however, in their consideration of Amazonian *Sapajus*, with H6 positing that all the Amazonian forms belong to one unique widespread taxon, while H7 considers *S. cay*, *S. apella* 1 (the northeastern Amazonian clade from ML and MSC phylogenetic reconstructions), and *S. apella* 2 (the southwestern Amazonian clade from ML and MSC phylogenetic reconstructions), as three separate taxa.





**Figure 3.** Species delimitation model hypotheses (H0 to H7). Each species delimitation model has a specific combination of lineages (rows).

The Bayes Factor (BF) has been widely used as a model selection tool when comparing alternative models or phylogenetic hypotheses [78,80,82,83]. A Bayes Factor is calculated as the ratio of the marginal likelihood of one model to the marginal likelihood of a competing model, where the marginal likelihood measures the average fit of a model to the data [84,85]. For this study, we used a method that simultaneously estimates the species tree while evaluating alternative species delimitation models by implementing the Bayes Factor Delimitation of Species (BFD) algorithm [80] in StarBeast2 v0.15.5 [86] within the software BEAST2 [87]. The marginal likelihood was estimated for each model through Path Sampling (PS) using BEAST2's [87] PathSampleAnalyser package application. Path sampling has been shown to generate highly accurate results for model selection of species delimitation [80,88]. PS was run for a chain length of 25 million generations for 20 path steps (totaling 500 million generations). All BFD StarBeast2 analyses were performed assuming a strict clock model, using the HKY site model with four gamma categories, and a Birth–Death speciation model. The convergence of the runs was assessed using ESS parameters with Tracer v1.7 [89]. We used marginal likelihood values to rank model hypotheses H1 to H7 and Bayes Factors to estimate the support for each model relative to the model with the highest marginal likelihood. The strength of support from Bayes Factor (BF) estimates for competing model hypotheses was evaluated according to Kass and Raftery's (1995) [84] framework. Therefore, the BF scale was used as follows: (a)  $0 < BF < 2$  means “not worth more than a bare mention”,  $2 < BF < 6$  means positive evidence,  $6 < BF < 10$  means strong support, and  $BF > 10$  means decisive support to distinguish between competing species delimitation model hypotheses. BFD analyses were run twice to confirm the consistency between runs. A final tree for the higher ranked model hypothesis was then obtained by combining posterior replicates with LogCombiner [87] and summarizing with TreeAnnotator v2.6 [87] under the maximum clade credibility trees and excluding 20% burn-in. Bayesian analyses were done using high-performance computing facilities at CIPRES [90] and TACC.

Finally, evolutionary timescale and confidence intervals for divergence dates for the capuchin radiation were inferred using a Bayesian MCMC method using the software StarBeast2 v0.15.5 [86] implemented in BEAST2 [87], assuming a strict molecular clock, using the constant population size model, using an HKY site model with four gamma categories, under the Birth–Death model prior for lineage branching, and using default hyperpriors. We used one calibration point on the root node to obtain the posterior distribution of the estimated divergence times: the divergence of *Saimiri* and capuchin monkeys (*Cebus* and *Sapajus*) based on the two fossil records of *Neosaimiri fieldsi* and *Panamacebus* [91,92]. Therefore, considering the minimum and maximum date estimates for these fossil records, we ran three divergence time analyses according to the calibration

models proposed by [93], as follows: (a) model calibration 1: an exponential distribution for the prior with a predefined offset of 12.1 Ma; (b) model calibration 2: a lognormal distribution with an offset of 12.5 Ma, mean of 1.8 million years, and standard deviation of 0.4 million years; and (c) model calibration 3: a lognormal distribution with an offset of 20.0 Ma, mean of 2.0 million years, and standard deviation of 0.5 million years.

### 3. Results

#### 3.1. Molecular Data

Overall, we generated  $\sim 2 \times 10^9$  raw reads for all the samples spread across 18 pooled libraries each containing between 19 and 24 samples. All libraries showed good quality, with mean Phred scores of 39 or above for R1 and R2 reads. As expected for the next generation sequencing approach, individual base quality scores decreased at the end of the reads, with the last 5 bp having mean Phred scores of 36 and 30 for the R1 and R2 reads, respectively. Approximately 95% of the raw reads were kept after all BBduk adapter trimming, PhiX removal, and quality filtering and trimming steps (Figure S1), and approximately 99% of these reads were successfully assigned to an individual sample in the demultiplexing step.

An average of 58,340 loci with depth  $>6x$  (min. = 9870, max. = 201,459) were assembled within each individual sample after all quality filtering and trimming steps (Table S3). The number of total loci recovered in the final genotyping matrix (across all samples) varied according to the criterion chosen for the number of samples a locus had to be present in for its inclusion in the final genotype matrix (Table 1). Overall, the number of total loci in the final ipyrad genotype matrices varied from 16,880 to 64,081 for the matrices with  $\approx 20\%$  (minimum of 150 samples used to select a locus) to  $\approx 40\%$  (minimum of 50 samples used to select a locus) missing data, respectively (Table 1).

**Table 1.** Total number of loci and the total number of SNPs included in the final genotype matrices, based on the minimum number of samples per locus for the output matrix.

Minimum % to Call a Locus <sup>1</sup>	Minimum # to Call a Locus <sup>2</sup>	Number of Loci <sup>3</sup>	SNPs Matrix Size	% Missing SNPs	Total # of Variable Sites	Parsimony-Informative Sites <sup>4</sup>	% Missing Sequence Matrix
$\approx 90\%$	150	16,880	353,995	15.38%	327,622	178,012	20.4%
$\approx 75\%$	130	31,216	614,050	19.34%	572,259	299,054	24.2%
$\approx 60\%$	100	43,736	830,442	24.41%	777,155	394,844	29.2%
$\approx 30\%$	50	64,081	1,125,828	34.32%	1,054,796	518,460	40.1%

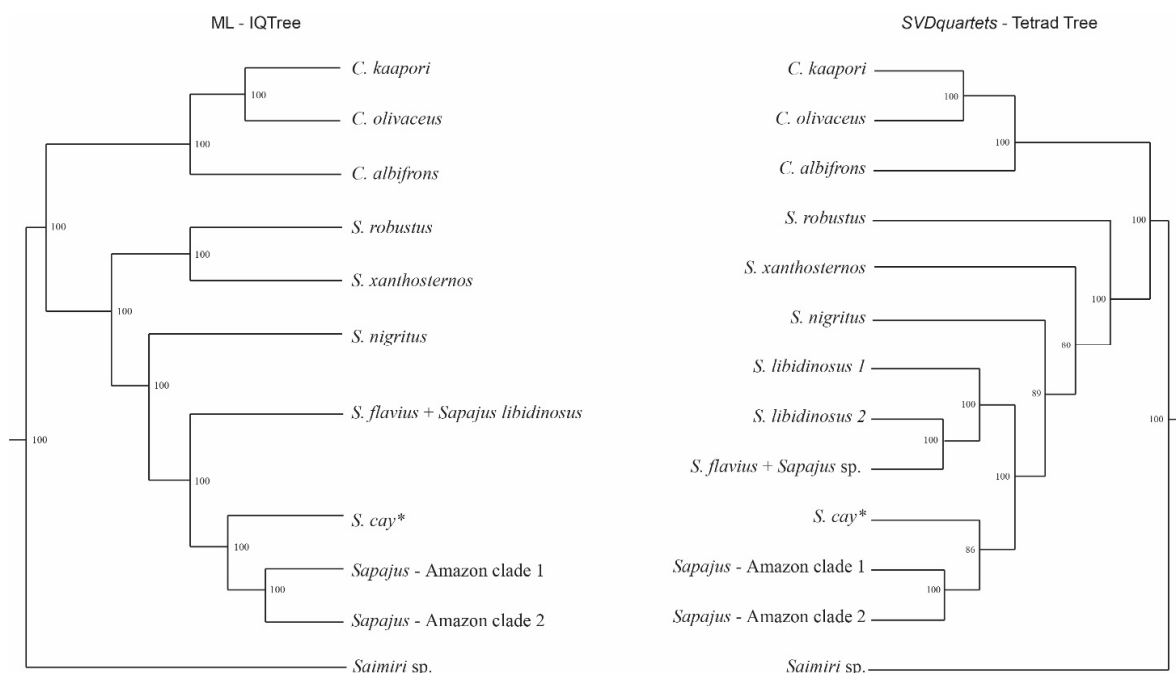
<sup>1</sup> Minimum number of samples a locus must be present in to be included in the final genotype matrix, represented as a percentage of the total number of samples. <sup>2</sup> Number of samples used as parameter for ipyrad. <sup>3</sup> Number of loci in the final genotype matrix. <sup>4</sup> Total number of parsimony-informative sites estimated by IQ-TREE.

#### 3.2. *Sapajus* Phylogenetic Reconstructions

Both ML and MSC analyses recovered the same tree topology when varying the degree of missing data. Results shown for ML and MSC are for the analysis with up to 34% missing data (i.e., genotype data available for a minimum of 50 samples for a locus to be included in the final genotyping matrix).

As expected from previous studies, we recovered strong support (100% of bootstrap replicates) for reciprocal monophyly of the *Sapajus* and *Cebus* genera from both the ML and MSC analyses (Figure 4, Figures S2, S3 and S7–S9). Within *Sapajus*, our ML analysis recovered *S. robustus* and *S. xanthosternos* as sister clades, with these two groups closer to *S. nigritus*, to the exclusion of all other clades, with full bootstrap support. Our ML phylogenetic reconstruction also recovered a monophyletic clade comprising *S. flavius* and *S. libidinosus*, to the exclusion of the capuchin monkeys occurring in the Pantanal and Amazon biomes, which were recovered as three monophyletic taxa, including *Sapajus cay* and two other Amazonian clades, all with strong support (100% bootstrap), see Figure 4. However, one of the samples from *S. cay* was recovered within one of the Amazonian

clusters (Figures S2, S3 and S10–S12). Complementarily, our MSC analysis also recovered *S. robustus*, *S. xanthosternos*, and *S. nigritus* as monophyletic groups, with *S. robustus* as the sister group to all *Sapajus* with strong support (100%), while the support values for the split of the monophyletic clades comprising either *S. xanthosternos* or *S. nigritus* were more moderate (80% and 89%, respectively). The MSC phylogenetic reconstruction (Figure 4) also recovered a monophyletic clade comprising *S. flavius* and *S. libidinosus* (100% support) and a grouping of three additional monophyletic clades, *S. cay* (with moderate support of 86%) and two Amazonian forms with strong support (100%). Again, as in the ML reconstruction, one of the samples putatively assigned to *S. cay* was recovered within one of the Amazonian clusters. Therefore, both ML and MSC analyses recovered *S. robustus* and *S. xanthosternos* as early splits within the *Sapajus* radiation and *S. nigritus* as the sister group to all other *Sapajus* clades. Both analyses also recovered Amazonian *Sapajus* as divided into a northeastern clade and a southeastern group, with maximum support in both analyses.



**Figure 4.** Maximum likelihood (ML-IQ-TREE, left) and SVDquartets species tree (SVDquartets—Tetrad tree, right) phylogenies for *Sapajus* radiation. Numbers at each node represent bootstrap support values (based on 1000 replicates). \* one of the samples putatively assigned to *S. cay* was recovered within one of the Amazonian clusters. See Figures S2–S5 for the complete trees.

Overall, the species trees recovered from both the ML and MSC analyses were congruent in recovering mostly the same set of monophyletic clades. However, aside from the support values, the main difference between our ML and MSC analyses was regarding the placement of the samples identified as *S. flavius* and *S. libidinosus* species. While both analyses recovered strong structure within the clade comprising samples from both *S. flavius* and *S. libidinosus* (with the latter being paraphyletic), the assignment of individual samples to the reciprocally monophyletic clades recovered was not congruent between analyses. The ML phylogenetic analysis suggests that samples of *S. libidinosus* from the southern Cerrado (Goiás state) belong to a different clade than samples of *S. libidinosus* from the northern Cerrado (Maranhão and Piauí states), while it placed all *S. libidinosus* from the Caatinga biome as closer to *S. flavius*, within a structured monophyletic group of *S. flavius* + *S. libidinosus* from the Caatinga (Figure 4, Figures S2 and S4). That is, the ML-based inference of phylogenetic relationships within the *S. flavius* + *S. libidinosus* group does not agree with that expected based on the geographic distribution of the samples, nor with morphological characteristics previously described for the specimens of each species [3,94].

By contrast, the MSC species tree reconstruction (Figure 4, Figures S3 and S5) recovered three reciprocally monophyletic clades within the *S. flavius* + *S. libidinosus* group, with the first clade comprising samples of *S. libidinosus* from the Caatinga biome, the second comprising samples of *S. libidinosus* from the Cerrado, and the third comprising all putative *S. flavius* samples. Interestingly, all samples from the eastern Caatinga that were previously identified (based on morphological characters) as coming from the blond capuchin monkey were indeed recovered within the *S. flavius* clade in both the ML and MSC analyses.

### 3.3. Species Delimitation and Divergence Dating

We tested eight species delimitation model hypotheses using the BFD method (Figure 3). The rankings of the alternative models based on their Marginal Likelihood Estimates (MLE) and Bayes Factor Delimitation (BFD) are shown in Table 2. Model H7, which agrees with our MSC phylogenetic reconstruction, received “decisive” support over all other model hypotheses ( $2\ln\text{BF} = 192.38\text{--}6867.94$ ). Among the remaining models, model H5 was the next closest in rank, based on both MLE and BFD, while models H0 and H1 were the least favored species delimitation model hypotheses. Model H7 posits the existence of nine distinct lineages within the *Sapajus* radiation; these include *S. xanthosternos*, *S. robustus*, *S. nigritus*, *S. flavius*, and *S. cay*, plus two lineages within what has, heretofore, been called *S. libidinosus* (one lineage comprising samples from the Cerrado and the other comprising samples from the Caatinga) and two lineages within what has been called *S. apella* from the Amazon. Note, however, that in the Bayesian analysis, the three Amazonian lineages (*S. cay* and two lineages with *S. apella*) received only moderate bootstrap support (Figure 5), and one of the *S. cay* samples was recovered within one of the Amazonian clades.

**Table 2.** Summary of results from the Bayes Factor species delimitation analysis. Species delimitation models (H0–H7) ordered by rank, their marginal likelihood estimates (MLE), and Bayes Factor testing results ( $2\ln\text{Bf}$ ) from the analyses with the Path Sampling (PS) methods.

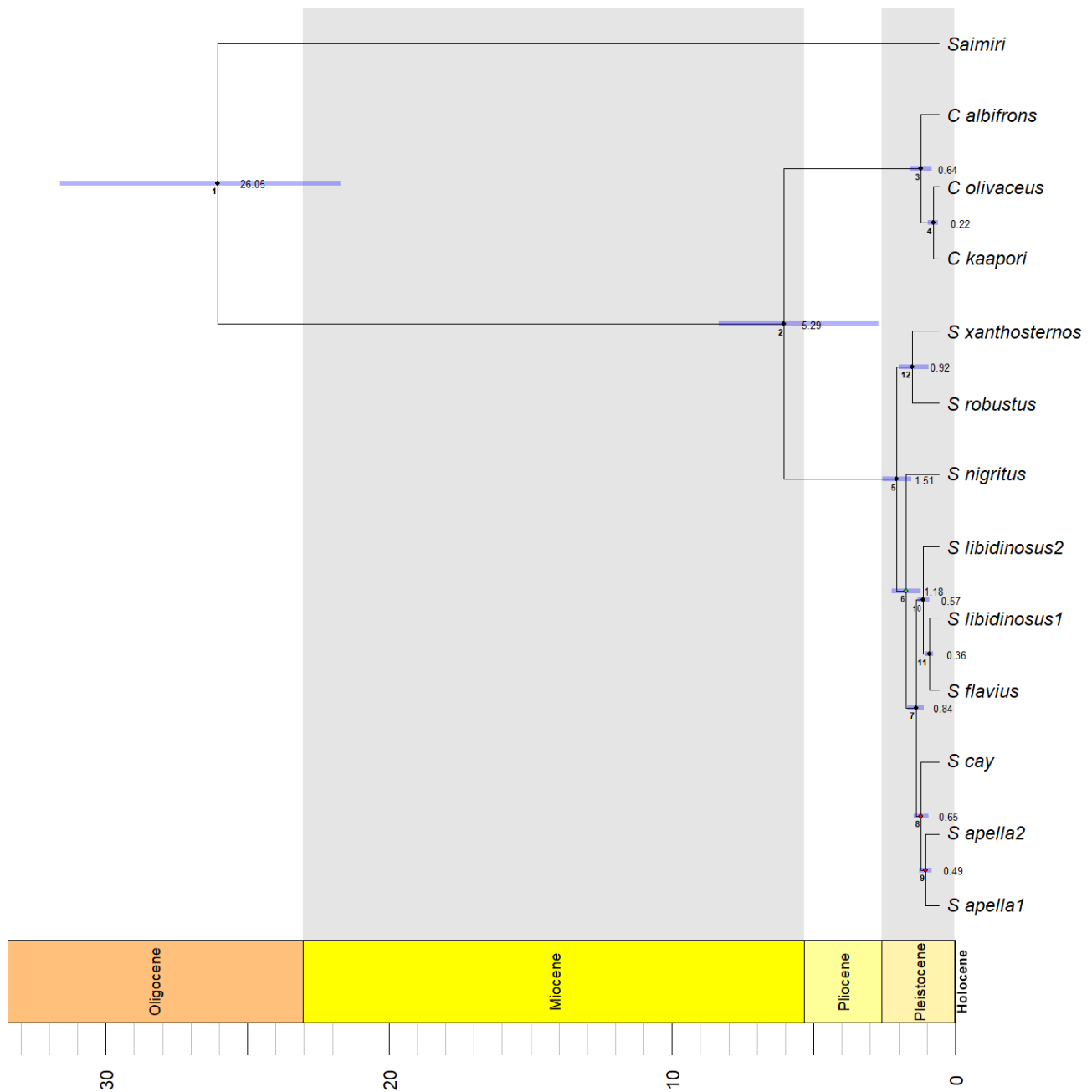
Model Hypothesis <sup>1</sup> (Ranked by MLE)	Marginal Likelihood Estimate (MLE)	Bayes Factor ( $2\ln\text{BF}$ ) <sup>2</sup> to the 1st Ranked Model H7	Bayes Factor ( $2\ln\text{BF}$ ) <sup>2</sup> to the 2nd Ranked Model H5
H7	−185,731.7644	-	−192.38
H5	−185,827.9526	192.38	-
H6	−186,006.9226	550.32	357.94
H1	−186,153.4149	843.30	650.92
H4	−186,337.5626	1211.60	1019.22
H3	−186,424.9737	1386.42	1194.04
H1	−187,530.4374	3597.35	3404.97
H0	−189,165.7362	6867.94	6675.57

<sup>1</sup> Model hypothesis as shown in Figure 3. <sup>2</sup> Bayes Factor, calculated by multiplying twice the ratio of the MLE of one model by the MLE of a competing model [ $\text{BF} = 2 \times (\text{MLE}_x - \text{MLE}_y)$ ], according to [79].

**Table 3.** Summary of the posterior distribution of divergence times (in Ma) estimated using the software Starbeast2.

Node in Time Tree <sup>1</sup>	Model Calibration 1		Model Calibration 2		Model Calibration 3 <sup>2</sup>	
	Median	95% HPD	Median	95% HPD	Median	95% HPD
1	13.04	12.1–16.39	14.03	13.08–15.46	26.05	21.71–31.62
2	2.85	0.9–3.89	3.08	1.26–3.83	5.29	1.95–7.58
3	0.34	0.13–0.53	0.37	0.16–0.53	0.64	0.26–1.03
4	0.11	0.02–0.2	0.12	0.02–0.21	0.22	0.04–0.41
5	0.78	0.52–1.07	0.83	0.57–1.04	1.51	0.99–2.02
6	0.61	0.34–0.88	0.65	0.4–0.88	1.18	0.66–1.69
7	0.43	0.29–0.59	0.46	0.33–0.58	0.84	0.56–1.14
8	0.33	0.21–0.47	0.35	0.23–0.48	0.65	0.39–0.9
9	0.25	0.13–0.37	0.26	0.16–0.38	0.49	0.28–0.72
10	0.29	0.18–0.4	0.32	0.21–0.39	0.57	0.35–0.75
11	0.19	0.1–0.27	0.2	0.13–0.26	0.36	0.21–0.51
12	0.5	0.17–0.74	0.53	0.2–0.72	0.92	0.36–1.42

<sup>1</sup> See Figure 5 for node labels. <sup>2</sup> Posterior distribution of divergence times shown in Figure 5.



**Figure 5.** StarBeast2 Bayesian divergence time analysis with node heights scaled to median divergence time estimates. Small numbers closer to the nodes correspond to the node labels in Table 3; numbers at the right of each node indicate the posterior means of the ages for each node. Blue bars represent 95% Highest Posterior Density (HPD) intervals. Colors of the nodes indicate the support values: black = high support/posterior  $\geq 0.99$ ; green = moderate support/posterior  $\geq 0.95 < 0.99$ ; red = low support/posterior  $< 0.95$  (see Figure S6 for detailed support values for each node).

Table 3 summarizes the results of the three different model calibrations used as priors in our StarBeast2 Bayesian divergence time analysis. The different root calibrations yielded variation in the posterior distribution of divergence times, with Model calibration 3 yielding the oldest estimates for all nodes, e.g., the split between the *Cebus* and *Sapajus* lineages [Node 2] as Median = 5.29 Ma; 95% Highest Posterior Density (HPD) interval = 1.95–7.58. In Figure 5, we present the divergence time estimates from this Model calibration 3 analysis.

#### 4. Discussion

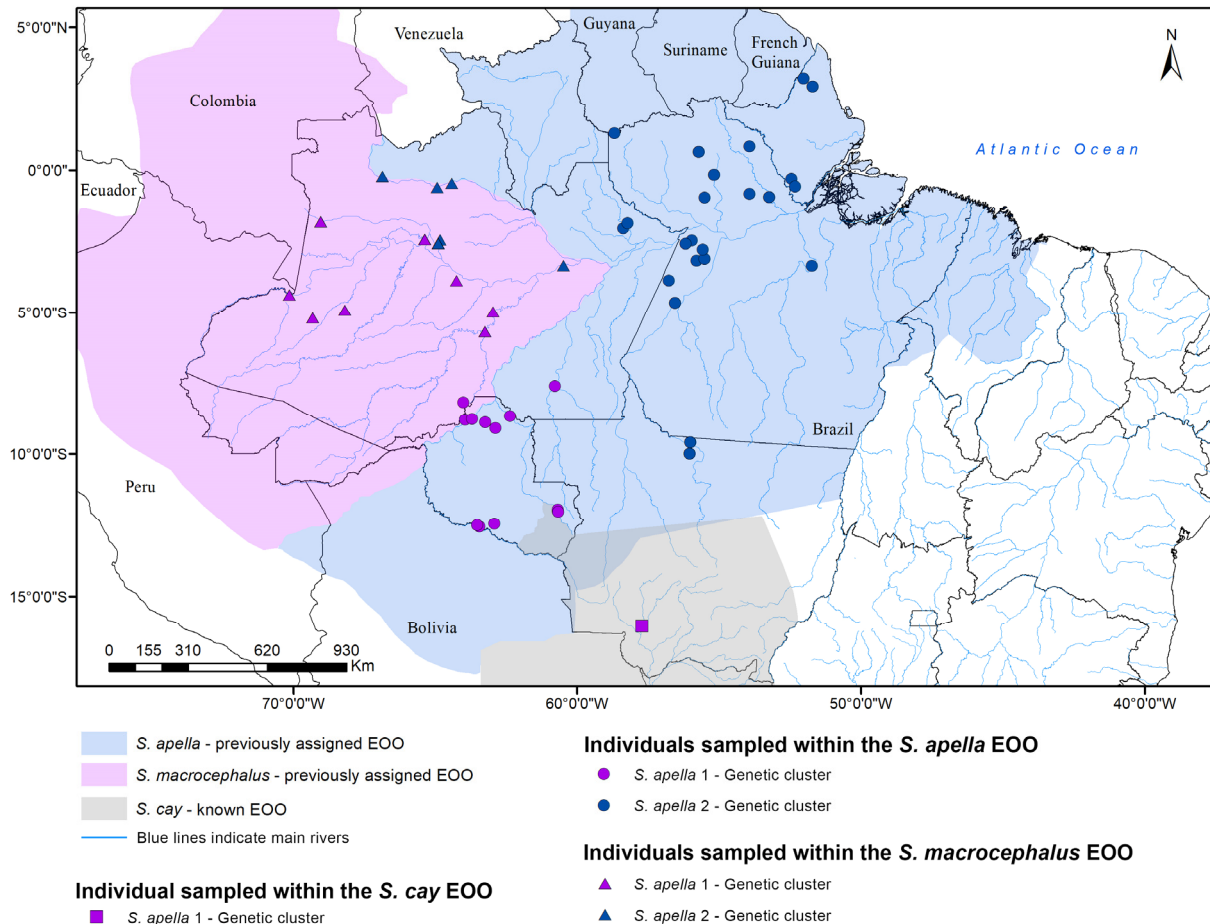
In this study, we used a phylogenomic approach to investigate the evolutionary relationships within the robust capuchin monkey radiation. Only a handful of prior investigations have used genetic evidence to evaluate phylogenetic relationships specifically within the genus *Sapajus* [5,13,95], and while these studies have provided some insight into the existing species diversity within the robust capuchin radiation, thus far, the picture has remained incomplete. Here, we used a larger sample size than that employed in any prior genetic study of the radiation, and we used a large set of phylogenomic markers to reconstruct the *Sapajus* phylogeny using both Maximum Likelihood (ML) and Multispecies Coalescent (MSC) methods. Importantly, we also applied a species delimitation approach within our MSC analysis, taking advantage of recent progress in genome-wide marker discovery and next generation sequence techniques, such as ddRADseq. Therefore, the results from this study provide new information to better understand species diversity within the robust capuchin monkeys and the evolutionary history among *Sapajus* lineages.

Both of our ML and MSC phylogenetic reconstructions provided support for the species status of *S. robustus*, the crested capuchin; *S. xanthosternos*, the yellow-breasted capuchin; and *S. nigritus*, the black-horned capuchin, corroborating previous findings [5,96] that also recovered these lineages as different taxa, though neither previous studies nor this study were able to confidently resolve the precise relationships among these species. Our findings support the placement of these three species from the Atlantic Forest of Brazil as the first splits within the robust capuchin radiation, with *S. nigritus* recovered as sister to all other *Sapajus* lineages from the Pantanal, Amazon, Cerrado, Caatinga, and northeastern Atlantic Forest regions. This finding disagrees with previous morphology-based taxonomies that have placed *S. robustus* as a subspecies of *S. nigritus* [11,12], but it corroborates other [3,7] taxonomic accounts for capuchin monkeys, including an earlier phylogeographic study that suggested an Atlantic Forest origin for the robust capuchins [97].

All of our analyses suggest that capuchin monkeys from the Pantanal and Amazon are divided into three reciprocally monophyletic clades. However, interestingly, one of the samples identified as *Sapajus cay* clustered with the *S. apella* 1 genetic cluster and the division of the samples within the two Amazonian clades do not agree with morphotypes or with the geographic division previously described for *S. apella* and *S. macrocephalus* species [3,7], as shown in Figure 6. Lima et al. (2018) [13] found no support for the molecular distinctiveness of the Pantanal and Amazon robust capuchin forms when using SNPs derived from Ultraconserved Element genomic markers (UCEs). These authors found some weak evidence for a northeastern and a southeastern clade within the Amazonian forms but also with the genetic lineages identified not agreeing with the current species hypotheses for the Amazonian *Sapajus*.

Silva-Júnior (2001) [3] analyzed morphological characters of more than 200 individuals of several localities throughout the Amazon for *S. apella* and 40 individuals from 12 localities for *S. macrocephalus* and described the distribution of both taxa based on these morphological characters while also considering the major rivers as possible barriers or limiting boundaries for these 2 Amazonian robust capuchin species. However, despite some evidence of rivers playing a role as a barrier for primate dispersal [98,99], other studies have demonstrated that, for some primate species, rivers might not, in fact, hinder animal movement and thus gene flow [100,101]. Therefore, even though our results do not support the geographic distributions previously inferred for two potential lineages of capuchins in the Amazon based on morphological analyses, they nonetheless demonstrate that two clades are indeed present: a “northeastern” clade potentially from the north of the Japurá and the Negro rivers at its most western portion in Brazil, widely overlapping with the previously described range for *S. apella*, and another “southwestern” clade to the south of these same rivers but extending through the Rondônia state of Brazil and somewhat overlapping with the previously described *S. macrocephalus* species range (Figure 6). Giving that these two lineages were recovered with strong support in both our ML and MSC analyses, and also considering the fact that these lineages were supported by the species

delimitation approach used, we suggest that new studies are needed to better diagnose the potential morphological characteristics distinguishing these two lineages, the limits of their distributions (not only in Brazil but also in Colombia where potentially both clades might occur), and whether these lineages should be recognized as distinct subspecies.



**Figure 6.** Map showing the genetic clusters found for Amazonian robust capuchin monkeys in both ML and MSC phylogenetic reconstructions, the previously assigned Extension of Occurrence (EOO) for *S. macrocephalus*, *S. apella*, and *S. cay*, and morphology-based identification of individuals samples, according to [3].

In our study, too, Azara's capuchin monkey, *S. cay*, was recovered as a sister clade to the rest of the widespread Amazon species cluster. However, only a small sample size for *S. cay* was used in this study ( $N = 3$  individuals), and one of these samples clustered within the *S. apella* 1 clade, thus it is not possible yet to assess the taxonomic validity of *S. cay*. Still, it is interesting to note that both the northern range limits of *S. cay* and the southern limits of *S. apella* species are not completely defined, raising questions on the correct taxonomic identification of the one *S. cay* sample that clustered within the Amazonian clade.

Overall, our phylogenetic reconstructions suggest that putative *Sapajus libidinosus*, sampled from across the dry biomes of the Caatinga and Cerrado, are paraphyletic and fall within a widespread clade with a geographic range that spans the Cerrado, Caatinga, and northeastern Atlantic Forest and includes all samples putatively from both *S. flavius* and *Sapajus libidinosus*. MSC phylogenetic reconstruction recovered two reciprocally monophyletic clades within putative *S. libidinosus*, corresponding to the geographic distribution of the samples, with one genetic cluster composed of the samples from the Cerrado, while the second corresponded to the samples from the Caatinga. In our ML analysis, the putative blond capuchin monkey samples were recovered as clustered within the set of samples of *S. libidinosus* from the eastern Caatinga. Contrarily, both of our MSC phylogenetic re-

constructions recovered *S. flavius* as a monophyletic clade, with the SVDquartets analysis recovering the blond capuchin as sister to the *S. libidinosus* clade from the Cerrado, and our Bayesian phylogenetic reconstruction instead suggesting that *S. flavius* is sister to the *S. libidinosus* clade from the Caatinga, both with strong support. Species delimitation analysis gave decisive support to model hypothesis H7, which considered three species for the capuchin monkeys within the Cerrado, Caatinga, and northeastern Atlantic Forest, corroborating the MSC phylogenetic reconstruction.

The fact that both our MSC and species delimitation analyses found, with strong support, that the widespread robust capuchins occupying the Cerrado and Caatinga belong to two reciprocally monophyletic clades challenges the current taxonomy for both *S. libidinosus* and *S. flavius*. Phylogenetic reconstruction incongruences, such as those that we found among the capuchin monkeys from the Cerrado, Caatinga, and northern Atlantic Forest, could indicate either the finding of a new cryptic lineage in the Caatinga biome or, rather, true inconsistencies due to incomplete lineage sorting or hybridization, for example. Incomplete lineage sorting (ILS), a process by which ancestral polymorphisms can persist through species divergences, and gene flow across species boundaries caused by introgressive hybridization, might generate gene tree discordances, hampering species tree estimation [102–104]. While MSC phylogenetic approaches have improved model complexity, making it possible to specifically account for lineage sorting and intraspecific variation within individuals [105], such models cannot account for high levels of gene flow, which has been shown to affect species tree inferences by decreasing posterior clade probabilities, underestimating divergence time estimates, and altering the species tree topology [102].

Phylogenetic relationships among taxa within the *Sapajus* radiation have continued to be contentious, with some studies suggesting *Sapajus* is a recent evolving lineage characterized by a high degree of past gene flow among certain lineages as well as ongoing admixture [5,13,106]. In addition, more recent phylogenetic and phylogenomic studies have supported recognizing either four or six species within the *Sapajus* radiation [5,13]. In this study, we applied a species delimitation method to explicitly evaluate different model hypotheses for the number of species in the robust capuchin genus and to identify potential evolutionarily independent lineages (e.g., distinct species). Our Bayes Factor species delimitation analysis suggests that the *Sapajus* radiation is composed of nine distinct lineages: *S. xanthosternos*, *S. robustus*, *S. nigritus*, *S. libidinosus* cluster 2 (samples from the Cerrado biome), *S. libidinosus* cluster 1 (samples from the Caatinga), *S. flavius*, *S. cay*, *S. apella* cluster 1 (corresponding to *S. macrocephalus*), and *S. apella* cluster 2. Overall, this result agrees with the morphology-based taxonomies of Silva-Júnior (2001) [3] and Rylands and colleagues (2013) [7], except that it highlights the paraphyly of *S. libidinosus* and suggests the geographic distributions of both Amazonian capuchin monkeys need to be reconsidered.

Importantly, this study has filled a longstanding gap regarding sample collection from both the eastern *S. libidinosus* and westernmost *S. flavius* ranges, corroborating previous morphology-based studies that suggested the presence of blond capuchin monkeys in some areas of the dry Caatinga [16,94], as all the samples putatively identified, based on morphology, as *S. flavius* from the Caatinga was indeed clustered with blond capuchin samples from the Atlantic Forest in all our analyses. However, in light of this new genetic evidence for the presence of the blond capuchin monkey in the Caatinga, as well as genetic evidence of hybridization among *Sapajus* lineages [13], additional analyses are needed to further investigate whether the paraphyletic arrangement found in the *S. libidinosus* samples represents a new genetic lineage within the Caatinga or, rather, reflects gene flow and introgression between the bearded capuchin monkey and the blond capuchin in the transition areas between the Atlantic Forest and Caatinga biomes.

Identifying species limits has never been a straightforward task [107]. Therefore, to robustly infer species boundaries, the use of integrative taxonomy has become increasingly common. Although sampling markers from across the genome and using models that account for different coalescent histories and discordance among loci can provide more objective measures for assessing species relationships and delimiting taxa [19,69,70,78,79],



species delimitation approaches, such as the one used in this study, should be seen as one source of evidence, which should be analyzed along with other lines of evidence as much as possible [107,108]. Therefore, new studies are necessary to better understand capuchin monkey species distributions and the occurrence of hybridization. Additionally, it will be important to further increase sampling of species and areas still poorly represented in this and other previous studies.

Finally, the time tree generated from our Starbeast2 analysis (Figure 5, Table 3), calibrated with the upper bound time estimate for the divergence of *Saimiri* and the capuchin monkeys, *sensu lato* (Model calibration 3), placed the estimated divergence time for *Cebus* and *Sapajus* genera at 5.29 Ma. While this is the oldest of the divergence times estimated for this particular node in our study, it is nonetheless more recent than the mean divergence time estimated between gracile and robust capuchin monkeys in other recent studies (5.8 Ma [5]; 6.6 Ma [13]; 6 Ma [109]; and 6.6 Ma [110]), although it is close to the divergence time estimated by [101], at 5.39 Ma based on whole mtDNA genomes. This divergence time estimate corresponds to a late Miocene divergence time for *Cebus* and *Sapajus* genera, which agrees with the hypothesis of the savanna-like environments in the Cerrado favoring a vicariance event separating primate populations of different genera, including populations of a capuchin ancestor occurring throughout the Amazon and Atlantic Forest [111]. The divergence time estimates for all other nodes within the *Sapajus* genus were more recent than those estimates found by [13], indicating the reinvasion of *Sapajus* into the Pantanal, Amazon, Cerrado, Caatinga, and northeastern Atlantic Forest was a recent event, occurring only 0.6 to 1.7 Ma, based on the divergence time estimated for the split between the *S. nigritus* and all other robust capuchin lineages outside the Atlantic Forest to the south of the São Francisco River.

## 5. Conclusions

In this study, we used the largest sample size to date to study the evolutionary history of the *Sapajus* genus and successfully generated data matrices with thousands of genomic markers for all putative species of robust capuchin monkeys. All of our analyses (ML and MSC phylogenetic reconstructions as well as species delimitation model testing under Bayesian inference) were congruent regarding the evolutionary history of the species from the Atlantic Forest south of the São Francisco River and of the species occurring in the Pantanal and Amazon in Brazil. *Sapajus robustus*, *S. xanthosternos*, and *Sapajus nigritus* were recovered as three monophyletic clades and as the first splits in the robust capuchin radiation, with *S. nigritus* recovered as closer to all other robust capuchin lineages. In addition, the Pantanal and Amazonian *Sapajus* were recovered as being structured into three monophyletic clades, although *Sapajus cay*, Azara's capuchin monkey, only received strong support from one (ML) out of three phylogenetic reconstructions, while the division of the Amazonian capuchin monkeys into two reciprocally monophyletic clades was strongly supported by two (ML and MSC) out of three phylogenetic reconstructions. We suggest these two clades should be considered at least as valid subspecies, with *Sapajus apella apella* as the lineage occurring to the north of the Jupará and Negro rivers and extending as a northeastern clade and with *Sapajus apella macrocephalus* as the lineage occurring in the southwestern Amazon to the south of the Negro River. However, new morphological assessments are necessary, as these Amazonian clades do not agree with previous morphology-based taxonomic distributions for the Amazonian capuchin monkeys.

Our phylogenetic reconstructions for the capuchin monkeys occurring in the Cerrado, Caatinga, and northeastern Atlantic Forest were less congruent, as we recovered the bearded capuchin monkey as a paraphyletic clade, with samples from the Caatinga biome belonging to either a monophyletic clade (MSC) or grouped with samples of the blond capuchin monkey (ML). Despite the strong support from both the MSC and the species delimitation approaches, further analyses are necessary to indicate whether this incongruence regarding the placement of the *Sapajus libidinosus* samples from the Caatinga is due to the occurrence of gene flow. Finally, our species delimitation approach supported

the division of the robust capuchin monkey into nine different species. However, this result should be seen as a taxonomic hypothesis and, as such, subject to further testing.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14050970/s1>, Figure S1: Results for adapter trimming, PhiX removal, and quality filtering/trimming; Figure S2: Maximum likelihood (IQ-TREE) phylogeny for *Sapajus* radiation; Figure S3: *SVDquartets* species tree (Tetrad) phylogeny for *Sapajus* radiation; Figure S4: Maximum likelihood (IQ-TREE) phylogeny for *Sapajus* radiation with full bootstrap numbers at each node; Figure S5: *SVDquartets* species tree (Tetrad) phylogeny for *Sapajus* radiation with full bootstrap numbers at each node; Figure S6: StarBeast2 Bayesian divergence analysis with full bootstrap numbers at each node; Figures S7–S9: Maximum likelihood (IQ-TREE) phylogeny for *Sapajus* radiation using data matrices comprising all the loci present in at least  $\approx 60\%$ ,  $75\%$ , and  $90\%$  of individuals, respectively, with full bootstrap numbers at each node; Figures S10–S12: *SVDquartets* species tree (Tetrad) phylogeny for *Sapajus* radiation using data matrices comprising all the loci present in at least  $\approx 60\%$ ,  $75\%$ , and  $90\%$  of individuals, respectively, with full bootstrap numbers at each node; Table S1: Morphology-based and phylogenetic taxonomies of robust capuchin monkeys; Table S2: Samples used in the study; Table S3: Total number of loci recovered for each sample that are present in the final data matrix under varying levels of missing data.

**Author Contributions:** Conceptualization, A.B.M., M.M.V.-M. and A.D.F.; methodology and formal analysis, A.B.M.; resources, samples, and data curation A.B.M., M.M.V.-M., J.W.L., M.G.M.L., W.K.S., J.d.S.e.S.-J., F.R., J.P.B. and A.D.F.; writing—original draft preparation, A.B.M.; writing—review and editing, A.B.M., A.D.F., J.W.L. and M.M.V.-M.; supervision, A.D.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** Financial support for this study was generously provided by the National Science Foundation (DDRI BSC—1650844), Primate Conservation, Inc., CAPES/CNPq—Science Without Borders Program, Rhonda L. Andrews Memorial Fellowship Award, Summer Writing Fellowship, The University of Texas at Austin, and a UT Graduate Continuing Fellowship.

**Institutional Review Board Statement:** The animal capture procedures, measurements, and sample collection used were approved by the Brazilian SISBIO/ICMBio (Research Permit Number 19927), and all protocols used were approved by the Institutional Care and Use Committee (IACUC) of The University of Texas at Austin (Protocol ID: AUP-2016-00077). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We are very grateful to Lina María Valencia Rodríguez and Edgardo Ortiz for generous support in the development of the ddRAD-seq protocol and to Jessica Podnar and Scott Hunicke-Smith from the Genome Sequencing and Analysis Facility (GSAF) at the University of Texas for their help and support through all stages of NGS library preparation and sequencing. We especially thank Marcos de Souza Fialho, Leandro Jerusalinsky, Plautino de Oliveira Laroque, Renata Bocorny de Azevedo, Gerson Buss, Bárbara Lins Caldas de Moraes, Bruna Barboza Bezerra, Rodrigo Ferraz Jardim Marques, Maria do Socorro da Silva, Wallace Pinto Batista, Usina Monte Alegre, Chesf, CPRH-PE, SUDEMA, MONA do São Francisco/ICMBio, and 15° Grupamento de Bombeiros Militar in Paulo Afonso for permits, logistic support, and assistance during key steps of data collection in the field.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hill, W.C.O. *Primates: Comparative Anatomy and Taxonomy. IV. Cebidae, Part A*; Interscience Publishers: Geneva, Switzerland, 1960.
2. 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. *Neotrop. Primates* **2000**, *8*, 61–93.
3. Silva Júnior, J.S. *Especiação Nos Macacos-Prego e Caiararas, Gênero Cebus Erxleben, 1777 (Primates, Cebidae)*; Universidade Federal do Rio de Janeiro: Rio de Janeiro, Brazil, 2001.
4. Alfaro, J.W.L.; Silva, J.D.S.E.; Rylands, A.B. How Different Are Robust and Gracile Capuchin Monkeys? An Argument for the Use of *Sapajus* and *Cebus*: *Sapajus* and *Cebus*. *Am. J. Primatol.* **2012**, *74*, 273–286. [[CrossRef](#)] [[PubMed](#)]

5. Lima, M.G.M.; Buckner, J.C.; Silva-Júnior, J.d.S.e.; Aleixo, A.; Martins, A.B.; Boubli, J.P.; Link, A.; Farias, I.P.; da Silva, M.N.; Röhe, F.; et al. Capuchin Monkey Biogeography: Understanding *Sapajus* Pleistocene Range Expansion and the Current Sympatry between *Cebus* and *Sapajus*. *J. Biogeogr.* **2017**, *44*, 810–820. [[CrossRef](#)]
6. Fragaszy, D.M.; Fedigan, L.M.; Visalberghi, E. *The Complete Capuchin: The Biology of the Genus Cebus*; Cambridge University Press: Cambridge, UK, 2004; ISBN 978-0-521-66768-5.
7. Rylands, A.B.; Bezerra, B.M.; Paim, F.P.; Queiroz, H. Species Accounts of Cebidae. In *Handbook of the Mammals of the World, 3 (Primates)*; Lynx Edicions: Barcelona, Spain, 2013; pp. 390–413.
8. Elliot, D.G. *A Review of Primates*; American Museum of Natural History: New York, NY, USA, 1913.
9. Hershkovitz, P. Mammals of Northern Colombia, Preliminary Report No. 4: Monkeys (Primates), with Taxonomic Revisions of Some Forms. *Proc. United States Natl. Mus.* **1949**, *98*, 323–427, Plates 15–17, Figures 52–59. [[CrossRef](#)]
10. Cabrera, A. Catálogo de Los Mamíferos de América Del Sur I (Metatheria, Unguiculata, Carnívora). In *Revista del Museo Argentino de Ciencias Naturales “Bernardino Rivadavia” e Instituto Nacional de Investigación de las Ciencias Naturales*; Ciencias zoológicas; Casa Editora Coni: Buenos Aires, Argentina, 1957; Volume 4, pp. 1–307.
11. Groves, C.P. Primate Taxonomy. In *Smithsonian Series in Comparative Evolutionary Biology*; Smithsonian Institution Press: Washington, DC, USA, 2001; ISBN 1-56098-872-X.
12. Groves, C.P. Order Primates. In *Mammal Species of the World: A Taxonomic and Geographic Reference*; Johns Hopkins University Press: Baltimore, MD, USA, 2005; Volume 1, pp. 111–184.
13. Lima, M.G.M.; Silva-Júnior, J.d.S.e.; Černý, D.; Buckner, J.C.; Aleixo, A.; Chang, J.; Zheng, J.; Alfaro, M.E.; Martins, A.; Di Fiore, A.; et al. A Phylogenomic Perspective on the Robust Capuchin Monkey (*Sapajus*) Radiation: First Evidence for Extensive Population Admixture across South America. *Mol. Phylogenetics Evol.* **2018**, *124*, 137–150. [[CrossRef](#)]
14. Rylands, A.B.; Mittermeier, R.A.; Silva, J.S. Neotropical Primates: Taxonomy and Recently Described Species and Subspecies: Neotropical Primate Taxonomy. *Int. Zoo Yearb.* **2012**, *46*, 11–24. [[CrossRef](#)]
15. Oliveira, M.M.; Langguth, A. Rediscovery of Marcgrave’s Capuchin Monkey and Designation of a Neotype for *Simia Flavia* Schreber, 1774 (Primates, Cebidae). *Bol. Mus. Nac.* **2006**, *523*, 1–16.
16. Ferreira, R.G.; Jerusalinsky, L.; Silva, T.C.F.; de Souza Fialho, M.; de Araújo Roque, A.; Fernandes, A.; Arruda, F. On the Occurrence of *Cebus flavius* (Schreber 1774) in the Caatinga, and the Use of Semi-Arid Environments by *Cebus* Species in the Brazilian State of Rio Grande Do Norte. *Primates* **2009**, *50*, 357–362. [[CrossRef](#)]
17. Fialho, M.d.S.; Valença-Montenegro, M.M.; da Silva, T.C.F.; Ferreira, J.G.; Laroque, P.d.O. Ocorrência de *Sapajus Flavius* e *Alouatta Belzebul* No Centro de Endemismo Pernambuco. *Neotrop. Primates* **2014**, *21*, 214–218. [[CrossRef](#)]
18. Giarla, T.C.; Esselstyn, J.A. The Challenges of Resolving a Rapid, Recent Radiation: Empirical and Simulated Phylogenomics of Philippine Shrews. *Syst. Biol.* **2015**, *64*, 727–740. [[CrossRef](#)]
19. Harris, R.B.; Alström, P.; Ödeen, A.; Leaché, A.D. Discordance between Genomic Divergence and Phenotypic Variation in a Rapidly Evolving Avian Genus (*Motacilla*). *Mol. Phylogenetics Evol.* **2018**, *120*, 183–195. [[CrossRef](#)] [[PubMed](#)]
20. Pollock, D.D.; Zwickl, D.J.; McGuire, J.A.; Hillis, D.M. Increased Taxon Sampling Is Advantageous for Phylogenetic Inference. *Syst. Biol.* **2002**, *51*, 664–671. [[CrossRef](#)] [[PubMed](#)]
21. Zwickl, D.J.; Hillis, D.M. Increased Taxon Sampling Greatly Reduces Phylogenetic Error. *Syst. Biol.* **2002**, *51*, 588–598. [[CrossRef](#)] [[PubMed](#)]
22. Heath, T.A.; Hedtke, S.M.; Hillis, D.M. Taxon Sampling and the Accuracy of Phylogenetic Analyses. *J. Syst. Evol.* **2008**, *46*, 239–257.
23. Kiesling, J.; Yi, S.V.; Xu, K.; Gianluca Sperone, F.; Wildman, D.E. The Tempo and Mode of New World Monkey Evolution and Biogeography in the Context of Phylogenomic Analysis. *Mol. Phylogenetics Evol.* **2015**, *82*, 386–399. [[CrossRef](#)]
24. Moyle, R.G.; Filardi, C.E.; Smith, C.E.; Diamond, J. Explosive Pleistocene Diversification and Hemispheric Expansion of a “Great Speciator”. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1863–1868. [[CrossRef](#)] [[PubMed](#)]
25. Nee, S.; Mooers, A.O.; Harvey, P.H. Tempo and Mode of Evolution Revealed from Molecular Phylogenies. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8322–8326. [[CrossRef](#)] [[PubMed](#)]
26. Rundell, R.J.; Price, T.D. Adaptive Radiation, Nonadaptive Radiation, Ecological Speciation and Nonecological Speciation. *Trends Ecol. Evol.* **2009**, *24*, 394–399. [[CrossRef](#)]
27. Braun, E.L.; Kimball, R.T. Polytomies, the Power of Phylogenetic Inference, and the Stochastic Nature of Molecular Evolution: A Comment on Walsh et al. (1999). *Evolution* **2001**, *55*, 1261. [[CrossRef](#)]
28. Degnan, J.H.; Rosenberg, N.A. Discordance of Species Trees with Their Most Likely Gene Trees. *PLoS Genet.* **2006**, *2*, e68. [[CrossRef](#)]
29. Maddison, W.P.; Knowles, L.L. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Syst. Biol.* **2006**, *55*, 21–30. [[CrossRef](#)] [[PubMed](#)]
30. Knowles, L.L.; Carstens, B.C. Delimiting Species without Monophyletic Gene Trees. *Syst. Biol.* **2007**, *56*, 887–895. [[CrossRef](#)] [[PubMed](#)]
31. Harrison, R.G.; Larson, E.L. Hybridization, Introgression, and the Nature of Species Boundaries. *J. Hered.* **2014**, *105*, 795–809. [[CrossRef](#)] [[PubMed](#)]

32. Eckert, A.; Carstens, B. Does Gene Flow Destroy Phylogenetic Signal? The Performance of Three Methods for Estimating Species Phylogenies in the Presence of Gene Flow. *Mol. Phylogenetics Evol.* **2008**, *49*, 832–842. [[CrossRef](#)] [[PubMed](#)]
33. Fontenot, B.E.; Makowsky, R.; Chippindale, P.T. Nuclear–Mitochondrial Discordance and Gene Flow in a Recent Radiation of Toads. *Mol. Phylogenetics Evol.* **2011**, *59*, 66–80. [[CrossRef](#)]
34. Kutschera, V.E.; Bidon, T.; Hailer, F.; Rodi, J.L.; Fain, S.R.; Janke, A. Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow. *Mol. Biol. Evol.* **2014**, *31*, 2004–2017. [[CrossRef](#)] [[PubMed](#)]
35. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-Wide Genetic Marker Discovery and Genotyping Using next-Generation Sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [[CrossRef](#)]
36. McCormack, J.E.; Hird, S.M.; Zellmer, A.J.; Carstens, B.C.; Brumfield, R.T. Applications of Next-Generation Sequencing to Phylogeography and Phylogenetics. *Mol. Phylogenetics Evol.* **2013**, *66*, 526–538. [[CrossRef](#)]
37. Ekblom, R.; Galindo, J. Applications of next Generation Sequencing in Molecular Ecology of Non-Model Organisms. *Heredity* **2011**, *107*, 1–15. [[CrossRef](#)]
38. Van Orsouw, N.J.; Hogers, R.C.J.; Janssen, A.; Yalcin, F.; Snoeijers, S.; Verstege, E.; Schneiders, H.; van der Poel, H.; van Oeveren, J.; Verstegen, H.; et al. Complexity Reduction of Polymorphic Sequences (CRoPS™): A Novel Approach for Large-Scale Polymorphism Discovery in Complex Genomes. *PLoS ONE* **2007**, *2*, e1172. [[CrossRef](#)]
39. Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)] [[PubMed](#)]
40. 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* **2011**, *6*, e19379. [[CrossRef](#)]
41. Van Tassel, C.P.; Smith, T.P.L.; Matukumalli, L.K.; Taylor, J.F.; Schnabel, R.D.; Lawley, C.T.; Haudenschild, C.D.; Moore, S.S.; Warren, W.C.; Sonstegard, T.S. SNP Discovery and Allele Frequency Estimation by Deep Sequencing of Reduced Representation Libraries. *Nat. Methods* **2008**, *5*, 247–252. [[CrossRef](#)]
42. Davey, J.W.; Blaxter, M.L. RADSeq: Next-Generation Population Genetics. *Brief. Funct. Genom.* **2010**, *9*, 416–423. [[CrossRef](#)] [[PubMed](#)]
43. Bergey, C.M.; Pozzi, L.; Disotell, T.R.; Burrell, A.S. A New Method for Genome-Wide Marker Development and Genotyping Holds Great Promise for Molecular Primatology. *Int. J. Primatol.* **2013**, *34*, 303–314. [[CrossRef](#)]
44. Reitzel, A.M.; Herrera, S.; Layden, M.J.; Martindale, M.Q.; Shank, T.M. Going Where Traditional Markers Have Not Gone before: Utility of and Promise for RAD Sequencing in Marine Invertebrate Phylogeography and Population Genomics. *Mol. Ecol.* **2013**, *22*, 2953–2970. [[CrossRef](#)]
45. Hohenlohe, P.A.; Day, M.D.; Amish, S.J.; Miller, M.R.; Kamps-Hughes, N.; Boyer, M.C.; Muhlfeld, C.C.; Allendorf, F.W.; Johnson, E.A.; Luikart, G. Genomic Patterns of Introgression in Rainbow and Westslope Cutthroat Trout Illuminated by Overlapping Paired-End RAD Sequencing. *Mol. Ecol.* **2013**, *22*, 3002–3013. [[CrossRef](#)]
46. Jones, J.C.; Fan, S.; Franchini, P.; Schartl, M.; Meyer, A. The Evolutionary History of Xiphophorus Fish and Their Sexually Selected Sword: A Genome-Wide Approach Using Restriction Site-Associated DNA Sequencing. *Mol. Ecol.* **2013**, *22*, 2986–3001. [[CrossRef](#)]
47. Takahashi, T.; Nagano, A.J.; Kawaguchi, L.; Onikura, N.; Nakajima, J.; Miyake, T.; Suzuki, N.; Kanoh, Y.; Tsuruta, T.; Tanimoto, T.; et al. A DdRAD-Based Population Genetics and Phylogenetics of an Endangered Freshwater Fish from Japan. *Conserv. Genet.* **2020**, *21*, 641–652. [[CrossRef](#)]
48. Nadeau, N.J.; Ruiz, M.; Salazar, P.; Counterman, B.; Medina, J.A.; Ortiz-Zuazaga, H.; Morrison, A.; McMillan, W.O.; Jiggins, C.D.; Papa, R. Population Genomics of Parallel Hybrid Zones in the Mimetic Butterflies, *H. Melpomene* and *H. Erato*. *Genome Res.* **2014**, *24*, 1316–1333. [[CrossRef](#)]
49. Kozlov, M.V.; Mutanen, M.; Lee, K.M.; Huemer, P. Cryptic Diversity in the Long-Horn Moth *Nemophora degeerella* (Lepidoptera: Adelidae) Revealed by Morphology, DNA Barcodes and Genome-Wide DdRAD-Seq Data: Cryptic Diversity in *Nemophora degeerella*. *Syst. Entomol.* **2017**, *42*, 329–346. [[CrossRef](#)]
50. Lavretsky, P.; DaCosta, J.M.; Sorenson, M.D.; McCracken, K.G.; Peters, J.L. DdRAD-seq Data Reveal Significant Genome-wide Population Structure and Divergent Genomic Regions That Distinguish the Mallard and Close Relatives in North America. *Mol. Ecol.* **2019**, *28*, 2594–2609. [[CrossRef](#)] [[PubMed](#)]
51. Lah, L.; Trense, D.; Benke, H.; Berggren, P.; Gunnlaugsson, P.; Lockyer, C.; Öztürk, A.; Öztürk, B.; Pawliczka, I.; Roos, A.; et al. Spatially Explicit Analysis of Genome-Wide SNPs Detects Subtle Population Structure in a Mobile Marine Mammal, the Harbor Porpoise. *PLoS ONE* **2016**, *11*, e0162792. [[CrossRef](#)]
52. Mynhardt, S.; Bennett, N.C.; Bloomer, P. New Insights from RADseq Data on Differentiation in the Hottentot Golden Mole Species Complex from South Africa. *Mol. Phylogenetics Evol.* **2020**, *143*, 106667. [[CrossRef](#)] [[PubMed](#)]
53. Scally, A.; Yngvadottir, B.; Xue, Y.; Ayub, Q.; Durbin, R.; Tyler-Smith, C. A Genome-Wide Survey of Genetic Variation in Gorillas Using Reduced Representation Sequencing. *PLoS ONE* **2013**, *8*, e65066. [[CrossRef](#)]
54. Ennes Silva, F.; Valsecchi do Amaral, J.; Roos, C.; Bowler, M.; Röhe, F.; Sampaio, R.; Cora Janiak, M.; Bertuol, F.; Ismar Santana, M.; de Souza Silva Júnior, J.; et al. Molecular Phylogeny and Systematics of Bald Uakaris, Genus *Cacajao* (Primates: Pitheciidae), with the Description of a New Species. *Mol. Phylogenetics Evol.* **2022**, *173*, 107509. [[CrossRef](#)]
55. Costa-Araújo, R.; de Melo, F.R.; Canale, G.R.; Hernández-Rangel, S.M.; Messias, M.R.; Rossi, R.V.; Silva, F.E.; da Silva, M.N.F.; Nash, S.D.; Boubli, J.P.; et al. The Munduruku Marmoset: A New Monkey Species from Southern Amazonia. *PeerJ* **2019**, *7*, e7019. [[CrossRef](#)]

56. 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* **2012**, *7*, e37135. [CrossRef]
57. Buckner, J.C.; Jack, K.M.; Melin, A.D.; Schoof, V.A.M.; Gutiérrez-Espeleta, G.A.; Lima, M.G.M.; Lynch, J.W. Major Histocompatibility Complex Class II DR and DQ Evolution and Variation in Wild Capuchin Monkey Species (Cebinae). *PLoS ONE* **2021**, *16*, e0254604. [CrossRef]
58. 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* **2018**, *13*, e0201254. [CrossRef]
59. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 10 April 2023).
60. Bushnell, B. BMAP: A Fast, Accurate, Splice-Aware Aligner. In Proceedings of the 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, USA, 17–20 March 2014.
61. Renaud, G.; Stenzel, U.; Maricic, T.; Wiebe, V.; Kelso, J. DeML: Robust Demultiplexing of Illumina Sequences Using a Likelihood-Based Approach. *Bioinformatics* **2015**, *31*, 770–772. [CrossRef] [PubMed]
62. Eaton, D.A.R.; Overcast, I. Ipyrad: Interactive Assembly and Analysis of RADseq Datasets. *Bioinformatics* **2020**, *36*, 2592–2594. [CrossRef] [PubMed]
63. Byrne, H.; Webster, T.H.; Brosnan, S.F.; Izar, P.; Lynch, J.W. Signatures of Adaptive Evolution in Platyrrhine Primate Genomes. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116681119. [CrossRef]
64. Fumagalli, M. Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. *PLoS ONE* **2013**, *8*, e79667. [CrossRef] [PubMed]
65. Harvey, M.G.; Judy, C.D.; Seeholzer, G.F.; Maley, J.M.; Graves, G.R.; Brumfield, R.T. Similarity Thresholds Used in DNA Sequence Assembly from Short Reads Can Reduce the Comparability of Population Histories across Species. *PeerJ* **2015**, *3*, e895. [CrossRef]
66. Rubin, B.E.R.; Ree, R.H.; Moreau, C.S. Inferring Phylogenies from RAD Sequence Data. *PLoS ONE* **2012**, *7*, e33394. [CrossRef]
67. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
68. De Queiroz, A.; Gatesy, J. The Supermatrix Approach to Systematics. *Trends Ecol. Evol.* **2007**, *22*, 34–41. [CrossRef] [PubMed]
69. Liu, L.; Yu, L.; Pearl, D.K.; Edwards, S.V. Estimating Species Phylogenies Using Coalescence Times among Sequences. *Syst. Biol.* **2009**, *58*, 468–477. [CrossRef]
70. Liu, L.; Wu, S.; Yu, L. Coalescent Methods for Estimating Species Trees from Phylogenomic Data: Estimating Species Trees from Phylogenomic Data. *J. Syst. Evol.* **2015**, *53*, 380–390. [CrossRef]
71. Bryant, D.; Bouckaert, R.; Felsenstein, J.; Rosenberg, N.A.; RoyChoudhury, A. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Mol. Biol. Evol.* **2012**, *29*, 1917–1932. [CrossRef] [PubMed]
72. Chifman, J.; Kubatko, L. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics* **2014**, *30*, 3317–3324. [CrossRef] [PubMed]
73. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
74. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
75. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
76. Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **1985**, *39*, 783–791. [CrossRef] [PubMed]
77. Carstens, B.C.; Pelletier, T.A.; Reid, N.M.; Satler, J.D. How to Fail at Species Delimitation. *Mol. Ecol.* **2013**, *22*, 4369–4383. [CrossRef] [PubMed]
78. Leache, A.D.; Fujita, M.K.; Minin, V.N.; Bouckaert, R.R. Species Delimitation Using Genome-Wide SNP Data. *Syst. Biol.* **2014**, *63*, 534–542. [CrossRef]
79. Fujita, M.K.; Leaché, A.D.; Burbrink, F.T.; McGuire, J.A.; Moritz, C. Coalescent-Based Species Delimitation in an Integrative Taxonomy. *Trends Ecol. Evol.* **2012**, *27*, 480–488. [CrossRef] [PubMed]
80. Grummer, J.A.; Bryson, R.W.; Reeder, T.W. Species Delimitation Using Bayes Factors: Simulations and Application to the *Sceloporus scalaris* Species Group (Squamata: Phrynosomatidae). *Syst. Biol.* **2014**, *63*, 119–133. [CrossRef]
81. Edwards, S.V.; Xi, Z.; Janke, A.; Faircloth, B.C.; McCormack, J.E.; Glenn, T.C.; Zhong, B.; Wu, S.; Lemmon, E.M.; Lemmon, A.R.; et al. Implementing and Testing the Multispecies Coalescent Model: A Valuable Paradigm for Phylogenomics. *Mol. Phylogenetics Evol.* **2016**, *94*, 447–462. [CrossRef]
82. Afonso Silva, A.C.; Santos, N.; Ogilvie, H.A.; Moritz, C. Validation and Description of Two New North-Western Australian Rainbow Skinks with Multispecies Coalescent Methods and Morphology. *PeerJ* **2017**, *5*, e3724. [CrossRef] [PubMed]
83. Solano-Zavaleta, I.; Nieto-Montes de Oca, A. Species Limits in the Morelet's Alligator Lizard (Anguillidae: Gerrhonotinae). *Mol. Phylogenetics Evol.* **2018**, *120*, 16–27. [CrossRef]
84. Kass, R.E.; Raftery, A.E. Bayes Factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [CrossRef]

85. Oaks, J.R.; Cobb, K.A.; Minin, V.N.; Leaché, A.D. Marginal Likelihoods in Phylogenetics: A Review of Methods and Applications. *Syst. Biol.* **2019**, *68*, 681–697. [[CrossRef](#)] [[PubMed](#)]
86. Ogilvie, H.A.; Bouckaert, R.R.; Drummond, A.J. StarBEAST2 Brings Faster Species Tree Inference and Accurate Estimates of Substitution Rates. *Mol. Biol. Evol.* **2017**, *34*, 2101–2114. [[CrossRef](#)]
87. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kühnert, D.; De Maio, N.; et al. BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006650. [[CrossRef](#)] [[PubMed](#)]
88. Aydin, Z.; Marcussen, T.; Ertekin, A.S.; Oxelman, B. Marginal Likelihood Estimate Comparisons to Obtain Optimal Species Delimitations in *Silene* Sect. *Cryptoneurae* (Caryophyllaceae). *PLoS ONE* **2014**, *9*, e106990. [[CrossRef](#)]
89. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [[CrossRef](#)]
90. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8.
91. Kay, R.F. Biogeography in Deep Time—What Do Phylogenetics, Geology, and Paleoclimate Tell Us about Early Platyrrhine Evolution? *Mol. Phylogenetics Evol.* **2015**, *82*, 358–374. [[CrossRef](#)]
92. Bloch, J.I.; Woodruff, E.D.; Wood, A.R.; Rincon, A.F.; Harrington, A.R.; Morgan, G.S.; Foster, D.A.; Montes, C.; Jaramillo, C.A.; Jud, N.A.; et al. First North American Fossil Monkey and Early Miocene Tropical Biotic Interchange. *Nature* **2016**, *533*, 243–246. [[CrossRef](#)]
93. Di Fiore, A.; Chaves, P.B.; Cornejo, F.M.; Schmitt, C.A.; Shanee, S.; Cortés-Ortiz, L.; Fagundes, V.; Roos, C.; Pacheco, V. The Rise and Fall of a Genus: Complete MtDNA Genomes Shed Light on the Phylogenetic Position of Yellow-Tailed Woolly Monkeys, *Lagothrix flavicauda*, and on the Evolutionary History of the Family *Atelidae* (Primates: Platyrrhini). *Mol. Phylogenetics Evol.* **2015**, *82*, 495–510. [[CrossRef](#)]
94. Silva, T.C.F. Estudo Da Variação Na Pelagem e Da Distribuição Geográfica Em *Cebus flavius* (Schreber, 1774) e *Cebus libidinosus* (Spix, 1823) Do Nordeste Do Brasil. Master Thesis, Universidade Federal da Paraíba, João Pessoa, Brazil, 2010.
95. Ruiz-García, M.; Castillo, M.I.; Luengas-Villamil, K. It Is Misleading to Use *Sapajus* (Robust Capuchins) as a Genus? A Review of the Evolution of the Capuchins and Suggestions on Their Systematics. In *Phylogeny, Molecular Population Genetics, Evolutionary Biology and Conservation of the Neotropical Primates*; Nova Science Publisher Inc.: New York, NY, USA, 2016; pp. 209–268.
96. Ruiz-García, M.; Castillo, M.I.; Lichilín-Ortiz, N.; Pinedo-Castro, M. Molecular Relationships and Classification of Several Tufted Capuchin Lineages (*Cebus apella*, *Cebus xanthosternos* and *Cebus nigrurus*, Cebidae), by Means of Mitochondrial Cytochrome Oxidase II Gene Sequences. *IJFP* **2012**, *83*, 100–125. [[CrossRef](#)] [[PubMed](#)]
97. Lynch Alfaro, J.W.; Boubli, J.P.; Olson, L.E.; Di Fiore, A.; Wilson, B.; Gutiérrez-Espeleta, G.A.; Chiou, K.L.; Schulte, M.; Neitzel, S.; Ross, V.; et al. Explosive Pleistocene Range Expansion Leads to Widespread Amazonian Sympatry between Robust and Gracile Capuchin Monkeys: Biogeography of Neotropical Capuchin Monkeys. *J. Biogeogr.* **2012**, *39*, 272–288. [[CrossRef](#)]
98. Boubli, J.P.; Ribas, C.; Lynch Alfaro, J.W.; Alfaro, M.E.; da Silva, M.N.F.; Pinho, G.M.; Farias, I.P. Spatial and Temporal Patterns of Diversification on the Amazon: A Test of the Riverine Hypothesis for All Diurnal Primates of Rio Negro and Rio Branco in Brazil. *Mol. Phylogenetics Evol.* **2015**, *82*, 400–412. [[CrossRef](#)] [[PubMed](#)]
99. Lynch Alfaro, J.W.; Boubli, J.P.; Paim, F.P.; Ribas, C.C.; da Silva, M.N.F.; Messias, M.R.; Röhe, F.; Mercês, M.P.; Silva Júnior, J.S.; Silva, C.R.; et al. Biogeography of Squirrel Monkeys (*Genus saimiri*): South-Central Amazon Origin and Rapid Pan-Amazonian Diversification of a Lowland Primate. *Mol. Phylogenetics Evol.* **2015**, *82*, 436–454. [[CrossRef](#)] [[PubMed](#)]
100. Piel, A.K.; Stewart, F.A.; Pintea, L.; Li, Y.; Ramirez, M.A.; Loy, D.E.; Crystal, P.A.; Learn, G.H.; Knapp, L.A.; Sharp, P.M.; et al. The Malagarasi River Does Not Form an Absolute Barrier to Chimpanzee Movement in Western Tanzania. *PLoS ONE* **2013**, *8*, e58965. [[CrossRef](#)] [[PubMed](#)]
101. Link, A.; Valencia, L.M.; Céspedes, L.N.; Duque, L.D.; Cadena, C.D.; Di Fiore, A. Phylogeography of the Critically Endangered Brown Spider Monkey (*Ateles hybridus*): Testing the Riverine Barrier Hypothesis. *Int. J. Primatol.* **2015**, *36*, 530–547. [[CrossRef](#)]
102. Leaché, A.D.; Harris, R.B.; Rannala, B.; Yang, Z. The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Syst. Biol.* **2014**, *63*, 17–30. [[CrossRef](#)]
103. Tajima, F. Evolutionary Relationship of DNA Sequences in Finite Populations. *Genetics* **1983**, *105*, 437–460. [[CrossRef](#)]
104. Pamilo, P.; Nei, M. Relationships between Gene Trees and Species Trees. *Mol. Biol. Evol.* **1988**, *5*, 568–583. [[CrossRef](#)]
105. Heled, J.; Drummond, A.J. Bayesian Inference of Species Trees from Multilocus Data. *Mol. Biol. Evol.* **2010**, *27*, 570–580. [[CrossRef](#)] [[PubMed](#)]
106. Martins-Junior, A.M.G.; Amorim, N.; Carneiro, J.C.; de Mello Affonso, P.R.A.; Sampaio, I.; Schneider, H. *Alu* Elements and the Phylogeny of Capuchin (*Cebus* and *Sapajus*) Monkeys: *Alu* Elements and Capuchin Monkeys. *Am. J. Primatol.* **2015**, *77*, 368–375. [[CrossRef](#)] [[PubMed](#)]
107. De Queiroz, K. Species Concepts and Species Delimitation. *Syst. Biol.* **2007**, *56*, 879–886. [[CrossRef](#)] [[PubMed](#)]
108. Edwards, D.L.; Knowles, L.L. Species Detection and Individual Assignment in Species Delimitation: Can Integrative Data Increase Efficacy? *Proc. R. Soc. B* **2014**, *281*, 20132765. [[CrossRef](#)]
109. Perelman, P.; Johnson, W.E.; Roos, C.; Seuánez, H.N.; Horvath, J.E.; Moreira, M.A.M.; Kessing, B.; Pontius, J.; Roelke, M.; Rumppler, Y.; et al. A Molecular Phylogeny of Living Primates. *PLoS Genet.* **2011**, *7*, e1001342. [[CrossRef](#)]

110. Springer, M.S.; Meredith, R.W.; Gatesy, J.; Emerling, C.A.; Park, J.; Rabosky, D.L.; Stadler, T.; Steiner, C.; Ryder, O.A.; Janečka, J.E.; et al. Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix. *PLoS ONE* **2012**, *7*, e49521. [[CrossRef](#)] [[PubMed](#)]
111. Lynch Alfaro, J.W.; Cortés-Ortiz, L.; Di Fiore, A.; Boubli, J.P. Special Issue: Comparative Biogeography of Neotropical Primates. *Mol. Phylogenetics Evol.* **2015**, *82*, 518–529. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.