A phylogenomic perspective on the robust capuchin monkey (*Sapajus*) radiation

Marcela G. M. Lima[1,2], Jessica W. Lynch Alfaro[1,3], Janet C. Buckner[4], Alexandre Aleixo[2], David Cerny[4], Jimmy Zheng[4], Michael E. Alfaro[4], Amely Martins[5,6], Jean P. Boubli[7], José de Sousa e Silva-Júnior[2]

1. Institute for Society and Genetics, University of California, Los Angeles, CA, USA
2. Curso de Pós-Graduação em Zoologia, Universidade Federal do Pará/Museu Paraense Emílio Goeldi, Belém, PA, Brazil
3. Department of Anthropology, UCLA, Los Angeles, CA, USA
4. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA
5. Department of Anthropology, University of Texas at Austin, Austin, TX, USA
6. Centro Nacional de Pesquisa e Conservação de Primatas Brasileiros, ICMBio, MMA, Brazil
7. School of Environment and Life Sciences, University of Salford, UK

Corresponding Author: Marcela G. M. Lima

Universidade Federal do Pará, Rua Augusto Corrêa, 01 – Guamá, Belém, PA, Brasil, CEP 66075-110

E-mail: marcela_gml@yahoo.com.br

Graphical Abstract

Highlights

- Phylogenomic analyses support *Sapajus* and *Cebus* clades within capuchin monkeys
- Molecular data support *Sapajus nigritus, S. robustus* and *S. xanthosternos* as species
- UCE phylogeny lumps *Sapajus* Amazonian and grassland morphospecies
- SNP data separate *S. flavius* and *S. libidinosus* as sister species
- We recommend collapsing *S. apella*, *S. macrocephalus* and *S. cay* as one species

Abstract

Phylogenetic relationships among robust capuchin monkeys (*Sapajus*) are poorly understood. Taxonomies for this group based on morphology have considered from one to twelve different species. Current IUCN classification lists eight robust capuchins: *S. xanthosternos*, *S. nigritus*, *S. robustus*, *S. flavius*, *S. libidinosus*, *S. cay*, *S. apella* and *S. macrocephalus*. Here we assembled the first phylogenomic data set for robust capuchin monkeys using ultra-conserved elements (UCEs) to construct a robust capuchin phylogeny using RAxML. We extracted SNPs from the UCE data set, and created SNP phylogenies using Bayesian and Maximum Likelihood methods. We estimated a species tree using SVDquartets analyses. All phylogenomic analyses strongly supported *Sapa-*

jus and *Cebus* clades within capuchin monkeys, and *Sapajus nigritus, S. robustus* and *S. xanthosternos* as species. However, the UCE phylogeny lumped morphospecies *S. cay, flavius, libidinosus, apella, macrocephalus,* and *flavius* together as a single widespread evolutionary lineage. The Bayesian SNP phylogeny was better resolved, and recovered *S. flavius* and *S. libidinosus* as sister species, together as sister to an *S. apella* + *macrocephalus* + *cay* clade; *S. apella, S. cay,* and *S. apella* individuals were interspersed together in the topology with no evidence for monophyly for any of these three morphological species. The species tree topology differed from the UCE and SNP topologies in that it reconstructed two major clades for robust capuchin monkeys: one Atlantic Forest clade (*S. robustus*, *S. xanthosternos*, and *S. nigritus*) and one widely distributed clade (*S. flavius*, *S. libidinosus*, plus north and south Amazonian robust capuchins). As morphological and molecular subdivisions of the Amazonian group + southern grasslands group (currently recognized as *S. cay*, *S. apella* and *S. macrocephalus*) are discordant, we recommend lumping all Amazonian plus southern grassland robust capuchin taxa as *S. apella* without subspecies.

Keywords

Neotropical primates, phylogeny, single nucleotide polymorphisms (SNPs), species tree, Ultraconserved elements (UCEs)

1. Introduction

Robust capuchin monkeys (*Sapajus*) comprise a widespread Neotropical primate genus found across cis-Andean Latin America, from the Colombian Llanos to the Guianas and throughout the Amazon basin as well as in the Atlantic Forest, Cerrado, Caatinga and Central Grasslands of South America, as far south as northern Argentina (Rylands et al., 2013). These primates as a group are true habitat generalists, with an incredible diet breadth compared to other Neotropical primates. While fruit and insects form the bulk of their diets, their robust jaw morphology coupled with behavioral adaptations for tool use and manipulative and extractive foraging together allow for the exploitation of encased and hidden foods unavailable to most other non-human animals (Fragaszy et al., 2004; Lynch Alfaro et al., 2012b).

Taxonomists have disagreed about the proximity of the relationship of robust capuchins to gracile capuchins. Elliot (1913) created a taxonomic key that divided the genus *Cebus* into tufted and non-tufted groups on the basis of the presence or absence of hair tufts on the frontal region of the head. However, only after Hershkovitz (1949) was there a general consensus about this division, with just one species (*Cebus apella* Linnaeus, 1758) recognized among the tufted group. Hill (1960) also considered all robust capuchins as one cosmopolitan species, *Cebus apella*, placed within the gracile capuchin genus, *Cebus*. Groves (2001, 2005) considered capuchins to form two species groups: (1) *C. capucinus* group with *C. capucinus*, *C. albifrons*, *C. olivaceus*, and *C. kaapori*; and (2) *C. apella* group with *C. apella*, *C. libidinosus*, *C. nigritus*, and *C. xanthosternos* (Table 1). Silva-Júnior (2001) separated robust capuchins as a different subgenus (*Sapajus*) from gracile capuchins (*Cebus*) on the basis of distinct cranial, postcranial and pelage morphology. Subsequently, genetic research validated the separation of robust and gracile capuchins as two distinct and equally diverse clades using mitochondrial (Lynch Alfaro et al., 2012a; Lima et al., 2017) and a combination of mtDNA

and nuclear (Perelman et al., 2011) markers. Two Alu elements provide strong evidence for the monophyly of robust *versus* gracile capuchins: Alu element S49P is present in *Sapajus* but not *Cebus* (Viana et al., 2015) and the AluSc8 insertion is found in *Cebus* but not *Sapajus* (Martins Jr. et al., 2015). A recent review justified the splitting of robust and gracile capuchins into two genera (*Cebus* for gracile capuchins and *Sapajus* for robust capuchins) based on the distinct morphology, biogeographic history, behavior, and ecology of each type (Lynch Alfaro et al., 2012b).

Taxonomists have also disagreed about the number of species encompassed by extant robust capuchins based on morphology (Table 1). Elliot (1913) recognized twelve species of robust capuchins, but Cabrera (1957) and Hill (1960) placed all robust forms into one species, *Cebus apella*, while retaining 11 and 16 subspecies, respectively. For the four decades between 1960 and 2000, most researchers lumped all robust capuchins as one species irrespective of place of origin, usually without regard for subspecies designations (e.g. Cole, 1992; Daegling, 1992; Ford and Hobbs, 1996; Masterson, 1997; Wright, 2005a; 2005b, 2007), leading to obfuscation of species or population differences within the robust capuchin literature (see Lynch Alfaro et al., 2014 for discussion). However, Torres de Assumpção (1983) pointed to distinct geographical variation in morphology among robust capuchin populations within Brazil, and especially within the Atlantic Forest. More recent morphological analyses have provided evidence for multiple *Sapajus* species (Groves, 2001, 2005; Silva-Júnior, 2001, 2002, 2005; Rylands et al., 2005, 2012, 2013; Rylands and Mittermeier, 2009). The robust capuchin group is now considered by most taxonomists to be comprised of four to eight species (Silva-Júnior., 2001; Groves, 2001; Rylands and Mittermeier, 2009; Rylands et al., 2005, 2012, 2013). The IUCN (2015) currently recognizes eight species: *Sapajus flavius*, the blonde capuchin; *S. xanthosternos*, the yellow-breasted capuchin; *S. robustus*,

the robust tufted capuchin; *S. nigritus*, the black-horned capuchin; *S. apella*, the brown capuchin; *S. macrocephalus*, the large-headed capuchin; *S. cay*, Azara's capuchin; and *S. libidinosus*, the bearded capuchin.

Recent biogeographic analyses based on mitochondrial DNA suggest that the time depth of the radiation of extant robust capuchins is about 2.5 My of diversification, with diversity accumulating first in the Atlantic Coastal Forest of Brazil, and a recent expansion of robust capuchins throughout the Amazon Basin and Cerrado, Caatinga and Central Grasslands in the last 500,000 years (Lynch Alfaro et al., 2012a; Lima et al., 2017). These analyses suggest that while the Atlantic Forest populations are relatively old and distinct, and can be separated as up to four different species, the Amazon/Grasslands radiation is better considered a highly polymorphic single species or species complex (Lima et al., 2017). If our current nuclear data set is congruent with the mtDNA data, we would expect to see evidence for four to five species: *S. nigritus*, *S. robustus*, and *S. xanthosternos* each as reciprocally monophyletic clades, with *S. flavius* either nested within or as the sister group to a single clade that extends across the Amazon and grasslands habitats in South America (and encompasses *S. apella*, *S. libidinosus*, *S. macrocephalus* and *S. cay* morphospecies) (Lima et al., 2017).

Here we use phylogenomic markers, ultraconserved elements (UCEs), to infer the phylogeny for robust capuchin monkeys, and to assess the evidence for congruence with species assignment by morphology and by mitochondrial and Alu markers. The UCE-based approach enriches DNA libraries for hundreds or thousands of UCEs and their flanking regions; then employs massively parallel sequencing for these libraries, and informatic tools to assemble, align and analyze the data (Faircloth et al., 2013). The UCE approach has been used successfully to resolve historically contentious taxonomical questions (McCormack et al., 2012; Crawford et al., 2012) including Pleistocene

144 radiations (McCormack et al., 2015). Previous studies using nuclear markers for capu-
145 chin phylogeny have utilized a limited number of taxa and used captive individuals
146 from unknown provenance as species exemplars (i.e. Perelman et al., 2011, Springer et
147 al., 2012). The present study marks the first test of robust capuchin phylogeny using
148 phylogenomic markers to analyze genetic relationships across species-representative
149 individuals from known provenance and assigned morphologically to each of the eight
150 currently recognized *Sapajus* species. Based on the most comprehensive mtDNA analy-
151 sis for the capuchin monkey radiation (Lima et al., 2017) we expect that much of the
152 diversification within the *Sapajus* genus has occurred relatively recently, within the
153 Pleistocene. We use SNP (Single Nucleotide Polymorphisms) data recovered within the
154 UCE results in order to refine our understanding of robust capuchin diversification, as
155 this technique was successful recently in elucidating the scrub-jay phylogeny across a
156 similar geologic time frame (McCormack et al., 2015).

157

158 **2. Material and methods**

159 *2.1. Samples, DNA extraction and sequencing*

160 We sampled 67 individuals from 8 species of the genus *Sapajus* and 4 species of
161 the genus *Cebus* from 62 localities distributed throughout the Atlantic Forest, Amazon,
162 Central Grasslands habitats and Central America (Figure 1 and Table 2). The total ge-
163 nomic DNA was extracted from muscle and blood samples using the Qiagen DNeasy
164 Blood & Tissue Kit, according to the manufacturer's protocol. Library preparation, se-
165 quence capture and sequencing of ultraconserved elements were performed by RAPiD
166 Genomics (Gainesville, FL, USA). Samples were quantified, normalized and sheared to
167 an average fragment length of 350 base pairs (bp) for library preparation. Samples were

dual-indexed with unique i5 and i7 8bp indexes. Libraries were then pooled with equimolar concentrations and the target sequence was captured using a custom set of 4715 probes targeting approximately 2300 UCE loci and 46 exons. Capture libraries were then pooled with equimolar concentrations for multiplexed dual-end (2x100bp) sequencing on an Illumina HiSeq 2500 v4 machine.

2.2. Sequence read quality control, assembly and UCE identification

We performed quality control using the trimming tool Trimmomatic 0.32.1 (Bolger et al., 2014) which trimmed sequences for adapter contamination, barcodes and low-quality regions using the parallel wrapper script in Illumiprocessor 2.0.6 (Faircloth, 2013) (https://github.com/faircloth-lab/illumiprocessor). We assembled the contigs for each sample using Trinity software package (vers. 2-25-2013) with default parameters using Phyluce 1.5.0 (Faircloth, 2016). We matched our assembled contigs to 4715 UCE loci custom-designed probe set using phyluce_assembly_match_contigs_to_probes integrating LASTZ 1.02.00 (Harris, 2007) from the Phyluce 1.5.0 (Faircloth, 2016) to remove any contigs that did not match probes or that matched multiple probes designed from different UCE loci. We performed in Phyluce 1.5.0 (Faircloth, 2016) the alignment of the contigs using the program phyluce_align_seqcap_align with MAFFT 7.271 (Katoh and Standley, 2013).

2.3. Phylogenetic analyses

For the phylogenetic analyses, we used a concatenated data set in a single alignment constructed in Phyluce 1.5.0 (Faircloth et al., 2012; Faircloth, 2016). We

191 used two data sets of UCE alignments that included greater than 95% of taxa present for
192 each UCE locus (5% missing) and greater than 75% of taxa present for each UCE locus
193 (25% missing), totaling 1838 UCEs with five exons (RAPGEF1, NAT15, GRIA21,
194 CLOCK e BDNF) and 1388 UCEs with two exons (NAT15, GRIA21) respectively. We
195 performed phylogenetic tree reconstruction under maximum likelihood (ML) in
196 RAxML 8.0.19 (Stamatakis, 2014), using a GTRCAT model of nucleotide substitution,
197 1000 replicate searches to identify the optimal tree and we generated non-parametric
198 bootstrap replicates using the autoMRE option of RAxML. To find the best partitioning
199 scheme, we used PartitionFinder (Lanfear et al., 2012). We considered each UCE as a
200 data block and enabled hcluster (Lanfear et al., 2014) with equal weights. To evaluate
201 the fit of each model we used the Bayesian information criterion (BIC).

202

203 *2.4. SNPs Analyses*

204 Upon identifying the target UCE loci, we computed the coverage at each base of
205 each contig using a python wrapper included in Phyluce
206 (phyluce_assembly_get_trinity_coverage_for_uce_loci). We then employed a de novo
207 SNPs calling approach by aligning all raw reads against our sample of *S. robustus*, the
208 reference sample with the highest coverage across all UCE loci enriched. This method
209 integrated BWA (v 0.7.7-1) and PICARD (v 1.106-0) to output de novo aligned align-
210 ments in BAM format, repair any formatting violations, add read group header infor-
211 mation, and mark duplicates in each BAM. We then merged all resulting BAMs into
212 one file, realigning the data and calling SNPs and indels using GATK (v 3.5-0-
213 g36282e4). To ensure high-quality SNPs in downstream analyses, we hierarchically
214 filtered the data according to stringent quality and validation parameters, excluding

SNPs with QUAL under 25, low variant confidence, and poor validation. Finally, the resulting VCF was passed through VCFTOOLS (v 0.1.14) to remove all loci that missed SNP calls for over 25% of all 67 samples.

On a parallel track, we passed our SNP data through a recently developed automatic pipeline called SNPhylo (Lee et al. 2014), designed to efficiently reconstruct trees based on genome wide SNPs. We modified our filtered VCF file by manually filling in autosomal chromosome positions for each SNP call, a necessary condition in order to run the program. We then set the Minor Allele Frequency threshold to 0.04 and negated the LD threshold to enable a more inclusive dataset for phylogenetic inference. We also bypassed the default low-quality data removal step, because the dataset had already undergone quality filtration with GATK. As a final step, the SNPhylo pipeline employs DNAML to generate a maximum likelihood hypothesis and passes the tree through PHANGORN, which generates 1000 bootstrap replicates for the final result.

Additionally, in ExaBayes 1.4.1 (Aberer et al., 2014), we performed two independent runs, each with four chains (three heated and one cold), from random starting topologies for 10 million generations with a sampling frequency of 500 generations. Posterior distributions of trees were summarized with the consensus script and combined with the postProcParam script. Convergence and stationarity of parameter estimates were verified using Tracer 1.6.0 (Rambaut et al., 2013).

We estimated a species tree using SVDquartets analyses (Singular Value Decomposition Scores for Species Quartets; Chifman and Kubatko, 2014) implemented in PAUP* v4.0a147 (Swofford, 2002). This method infers quartets based on summaries of SNPs in a concatenated sequence matrix species using a coalescent model. We randomly sampled 10 million quartets from the data matrix to infer a species tree and we meas-

ured uncertainty in relationships using nonparametric bootstrapping with 1000 replicates. For this analysis, we did not include the samples from the widely distributed clade that did not form a part of the Northern Amazon or Southern Amazon subclades in the Bayesian (Exabayes) and maximum likelihood (SNPhylo) trees.

2.5. Divergence dating analyses

For the purposes of divergence time estimation, the 75% complete dataset was re-analyzed in PartitionFinder 2 (Lanfear et al., 2017) using the *k*-means algorithm described by Frandsen et al. (2015) and the BIC as the model selection method. We identified the fastest-evolving partition based on the rate multipliers reported in auxiliary files generated using the "--save-phylofiles" flag. This partition, totaling 10,316 sites, was then used to conduct a time tree analysis in BEAST 1.8.2 (Drummond et al., 2012).

We used the birth-death branching process (Gernhard 2008) with default hyperpriors placed on the growth rate and relative death rate hyperparameters to generate the joint prior distribution on tree topology and node heights. The uncorrelated lognormal relaxed clock was used to model the distribution of branch rates across the tree. In order to constrain the branch rate distribution to biologically realistic values, we placed a lognormal hyperprior with a mean of 0.005 (in real space) and a standard deviation of 1 on the ucld.mean hyperparameter (initial value of 0.005), and assigned a truncated exponential distribution with support from 0 to 1 and a mean of 0.3 to the ucld.stdev hyperparameter (initial value of 0.1). GTR+Γ was specified as the nucleotide substitution model; all of its free parameters were assigned default priors, the base frequencies were estimated rather than fixed, and the gamma rate heterogeneity distribution was discretized into 4 categories.

We ran the analysis under the fixed topology operator mix as specified in BEAUTi v1.8.4 (Drummond et al., 2012), with the tuning of the ucld.mean and ucld.stdev operators set to 0.9 and their weight increased to 6.0. All remaining operators were kept at their default values. The topological constraint we employed (Supplementary Figure 1) was based on the species tree inferred with SVDquarters (see below), with one callitrichid and seven catarrhine outgroups manually added to the tree based on the generally accepted phylogeny of the Simiiformes (Perelman et al., 2011; Springer et al., 2012). The data for outgroup species were generated from Faircloth et al. 2012. Since most of the calibration points we used were concentrated within the catarrhine part of the tree, we pruned the capuchin taxon sample down to 4 species, with 2 representatives of the genus *Cebus* (*C. capucinus* and *C. olivaceus*) and 2 representatives of *Sapajus* (*S. apella* and *S. xanthosternos*) in order to increase the ratio of calibrated to uncalibrated nodes, as well as to achieve a more uniform placement of fossil data throughout the tree.

To calibrate the tree, we used all of the fossil dates previously employed by Springer et al. (2012) that were applicable to our restricted taxon sample (Table 3). To assess the sensitivity of the posterior node age distribution to the root age prior, we also ran an additional analysis using an older root calibration derived from the age of *Perupithecus* (Bond et al., 2015). Each calibration point was assigned an offset exponential density such that the upper bound specified by Springer et al. (2012) corresponded to the 95th percentile of the distribution. In contrast to the uniform densities utilized by Springer et al. (2012), exponential distributions have the advantage of concentrating most probability mass close to the lower bound. As single-parameter distributions, exponentials are also less arbitrary than lognormal priors commonly used in BEAST time

tree analyses, which can render the posterior overly sensitive to the choice of calibration density hyperparameters (Warnock et al., 2012).

The Markov chain Monte Carlo analysis was run for 400 million generations, sampling every 1000 generations and removing the initial 10% of samples as burnin. We assessed convergence of the chain using the effective sample sizes (ESS) reported for each parameter in Tracer 1.6.0 (Rambaut et al., 2013) by ensuring that all the ESS values exceeded 200. The posterior distribution of time trees was summarized into a maximum clade credibility tree using TreeAnnotator 1.8.3 (Rambaut and Drummond, 2015a).

3. Results

3.1. Quality control

We sequenced a total of 178 million read pairs (mean = 2,661,695.4) for all samples. An average of 3309 contigs per sample (min = 1162, max = 6170) was assembled from 67 individuals (Table 2). After alignment and trimming as described above, we got an average of 1882 unique contigs matching UCE loci from each sample. We produced a 75% complete data matrix containing 1843 alignments of UCE loci, which produced a concatenated matrix of 550,515 bp (average length: 298.70 bp per alignment) and a 95% complete data matrix containing 1390 alignments of UCE loci, which produced a concatenated matrix of 439,190 bp (average length: 315.96 bp per alignment).

3.2. Phylogenomic analyses

We recovered strong support in the tree topology from our RAxML (75% and 95%) analyses for reciprocal monophyly between the *Sapajus* and *Cebus* clades (Figure 2 and Supplementary Figure 2). Our analyses show strong molecular support for three of the morphological species within the genus *Sapajus*: *S. robustus, S. xanthosternos* and *S. nigritus,* all within the Atlantic Forest of Brazil. All other morphologically defined species within the genus (*S. flavius, S. libidinosus*, *S. apella*, *S. cay*, and *S. macrocephalus*) group together with high support in a widely distributed clade (from the Atlantic Forest to the Amazon), but there is no support for any subclades within this group in either the 75% or 95% taxa sets. Thus, the RAxML tree suggests four species of *Sapajus*: *S. robustus, S. xanthosternos* and *S. nigritus* from the Atlantic Forest of Brazil, and a widespread species that encompasses morphotypes *S. flavius, S. libidinosus, S. apella*, *S. cay,* and *S. macrocephalus*.

3.3. SNPs Analyses

After filtering out low quality SNPs, we retained a total of 19,583 SNPs across all samples. We then filtered for missing data and included only the SNPs that were parsimony-informative sites, generating a 75% complete matrix with a total of 11,462 informative high quality SNPs.

Similar to the RAxML analyses, our Maximum Likelihood and Bayesian trees using SNPs from the UCE data recover *S. xanthosternos* and *S. nigritus* as monophyletic clades, with the single *S. robustus* sample as the sister group to *S. xanthosternos* (Figure 3). However, within the widely distributed clade in the SNP trees, there are two distinct subclades. One subclade recovers monophyly of the species *Sapajus flavius* and also contains all *S. libidinosus* samples in a clade with *S. apella* specimens from Tucu-

ruí. The other subclade contains *S. cay*, *S. apella*, and *S. macrocephalus*; clusters within this subclade are geographically coherent but do not correspond to the current morphological taxonomy of the genus *Sapajus*. There is a clear division between Amazonian *Sapajus* north and south of the Amazon River, with some exceptions. Thus, our phylogenomic SNP data provides some support for six distinct species within *Sapajus*: *S. nigritus*, *S. robustus*, *S. xanthosternos*, *S. flavius*, *S. libidinosus* and a widespread Amazonian and southern grasslands species.

While the ExaBayes and SNPhylo had similar topologies, the two trees differed in the strength of their support for particular clades. For example, the SNPhylo tree resolved *S. nigritus* as the sister group to the widespread *Sapajus* clade (98), and *S. robustus* as sister to *S. xanthosternos* (96). SNPhylo also resolved *S. flavius* + (*S. libidinosus* + Tucuruí *S. apella*) clade as the sister group to *S. apella* + *S. macrocephalus* + *S. cay* (100). On the other hand, the ExaBayes tree provided higher support for the *S. flavius* + (*S. libidinosus* + Tucuruí *S. apella*) clade (0.99) and for the *S. cay* + Rondônia *S. apella* clade (0.99). Within the widespread Amazonian *S. apella* + *S. macrocephalus* + *S. cay* clade, ExaBayes recovered a northwestern *S. macrocephalus* subclade (0.99) and a northeastern *S. apella* subclade (0.99) that were strongly supported as sister to each other (0.97). ExaBayes also supported the sister relationship (0.95) between the *S. cay* + Rondônia *S. apella* subclade and a south-central Amazonian *S. macrocephalus* clade (Atalaia, Purus, Jirau, Canutama, Cujubim, Mamiraua, Japura, Jamari; 0.91). In contrast, the internal topology for the subclades of the *S. apella* + *S. macrocephalus* + *S. cay* clade was less well-supported in SNPhylo.

In the species tree recovered using SVDquartets analyses (Figure 4), we found strong support (100) in the tree topology for reciprocal monophyly between *Sapajus* and *Cebus.* The internal topology differed in some regards for *Sapajus* when compared to

359 our RAxML, ML and Bayesian trees using SNPs from the UCE data. As in other anal-
360 yses, *Sapajus xanthosternos* and *S. robustus* were strongly supported as sister taxa
361 (100), but here *S. nigritus* was weakly supported (77) as sister to *S. xanthosternos* + *S.*
362 *robustus.* While in the other trees, *S. apella*, *S. macrocephalus*, *S. cay*, *S. flavius*, and *S.*
363 *libidinosus* formed a subclade nested within the Atlantic forest robust capuchin clade
364 and sister to *S. nigritus*, here this widespread group forms a second and well-supported
365 (100) clade distinct from the Atlantic forest clade, with *S. flavius* supported (90) as sis-
366 ter to *S. libidinosus*, and Northern Amazonian and Southern Amazonian robust capu-
367 chins together forming a clade (100).

368

369 *3.4. Divergence time analyses*

370 While the BEAST run with the younger root calibration (based on *Aegypto-*
371 *pithecus* at 28.3 Ma) reached convergence after the specified number of generations
372 (ESS values ≥ 250 for all parameters), the analysis employing the *Perupithecus*-derived
373 36 Ma minimum on the age of the root failed to converge, as indicated by an effective
374 sample size of <200 for the age of the hominoid-cercopithecid divergence (node 6). An
375 additional run of 100 million generations was performed and combined with the first
376 chain using LogCombiner 1.8.3 (Rambaut and Drummond, 2015b); however, the result-
377 ing ESS values were lower than those obtained from the first run alone, suggesting that
378 the two chains had sampled from different distributions. To overcome this problem, a
379 third chain of 500 million generations was run in BEAST under the same settings. The
380 ESS values for both the third run alone and the total combined run of 900,000 samples
381 exceeded 200 for all parameters.

Regardless of the choice of root prior, the 95% highest posterior density (HPD) intervals of all calibrated nodes were well within the bounds used to construct the respective calibration densities (compare Tables 3 and 4). Use of the *Perupithecus* calibration shifted the marginal posterior distribution of the root age from the Late to Middle Eocene but exercised comparatively little influence on the estimated ages of shallower divergences (Table 4). The intrageneric divergences within both *Cebus* and *Sapajus* (Table 4; nodes 7 and 8) were consistently older and less precise (marked by wider 95% HPD intervals) when estimated under the *Perupithecus*-derived root age prior. The mean estimated split between robust and gracile capuchins (Table 4; node 9) shifted from 5.4 to 6.8 Ma when *Perupithecus* was used to calibrate the platyrrhine-catarrhine divergence, while the width of the corresponding 95% HPD interval remained unchanged.

4. Discussion

Together our analyses provide genetic support for six distinct species within *Sapajus*: five morphological species (strong support for *S. robustus, S. xanthosternos, S. nigritus,* and more equivocal support for *S. libidinosus* and *S. flavius*) and one morphologically diverse Amazonian + Central Grasslands species that contains two major clades separated by distributions in Northern *versus* Southern Amazonia. Recent mitochondrial studies provide some additional support for the species status of *S. robustus, S. xanthosternos* and *S. nigritus* though the exact relationships among species varies (Lima et al., 2017; Ruiz-Garcia et al., 2012). *S. flavius* is recovered as a monophyletic group with mitochondrial data, but is embedded within the widespread clade, or positioned as sister to the widespread clade (Lima et al., 2017), whereas the nuclear results

here place *S. flavius* and *S. libidinosus* as sister taxa. Both the mtDNA and the nuclear DNA topologies are discordant with Groves' (2001) taxonomic hypothesis that *S. robustus* is a subspecies of *S. nigritus*, because *S. nigritus* and *S. robustus* do not group together as sister taxa within *Sapajus.* In the previous studies employing large numbers of concatenated loci to elucidate primate relationships (Perelman et al., 2011; Springer et al., 2012), *S. robustus* and *S. xanthosternos* are recovered as sister taxa to the exclusion of *S. apella*. In Springer et al. (2012) *S. apella* is recovered as sister to *S. libidinosus*, consistent with our present phylogeny.

While all *Sapajus libidinosus* samples with light yellow pelage phenotype found across *S. libidinosus* distribution in the relatively dry biomes of Caatinga and Cerrado cluster together in one clade, that clade also includes samples that present standard *S. apella* pelage at the border of the two species distributions, near Tucuruí, Pará, on the eastern side of the lake that was formed by the damming of the Tocantins River for a Hydroelectric Plant (Figure 5b). These same individuals with *S. apella* morphotypes from Tucuruí cluster genetically with all sampled individuals with *S. libidinosus* pelage from within *S. libidinosus* distribution when using mitochondrial markers (Lima et al., 2017). Tucuruí capuchins have darker pelage and live in tropical forest habitat, while nearby *S. libidinosus* are adapted to open Cerrado and Caatinga habitats, and have lighter pelage. *S. libidinosus* has also been shown to have cranial and post-cranial adaptations to increased ground use and encased fruit extraction (Wright et al., 2015). Morphometric data are not available for the Tucuruí specimens, to determine if their cranial and post-cranial characteristics cluster with *S. libidinosus* or *S. apella*. Their external coloration should also be studied in detail to compare with other *Sapajus* specimens. The unexpected topology leaves us with various possibilities; it may be that the *S. libidinosus* lineage has expanded from the Cerrado biome to make inroads into the Amazon,

and that *S. libidinosus* populations living in forested areas evolve darker pelage, so that they converge in appearance with *S. apella*. This could be a result of genetic adaptation, or it could be that capuchins have a developmental response with coat color adjusting to habitat conditions. Either way, this suggests ecological forces may be driving coat color and morphological characteristics. A second possibility is that *S. apella* east of the Tocantins River became isolated from other robust Amazonian capuchins, and over time gave rise to the Caatinga and Cerrado populations of *S. libidinosus*. A third possibility is that *S. apella* and *S. libidinosus* have come into secondary contact at the borders of their distribution, and that despite significant gene flow, the two populations maintain their pelage characteristics. More morphological, genetic and ecological data will need to be collected in the Cerrado-Amazon transition zone in order to better understand relationships among capuchin populations here.

Note that *Sapajus libidinosus* + Tucuruí samples formed a clade with *S. flavius*. For this study, we sampled across western Caatinga and Cerrado for *S. libidinosus*, but we do not have samples here for eastern Caatinga where *S. libidinosus* is found close to *S. flavius* in northeastern Brazil (Figure 5b). More data from the Cerrado-Amazon transition zone and the Caatinga-Atlantic Forest transition zone could resolve if *S. flavius* and *S. libidinosus* are geographical variants of the same species, two distinct species, or are best lumped within the widespread *S. apella* group described below.

The molecular distinctiveness of the other morphological species currently assigned to *Sapajus* is not supported. Within the widespread *Sapajus* clade recovered in the SNP tree, there were strong indications for shared evolutionary history among morphotypes *S. cay*, *S. apella* and *S. macrocephalus*. There was no reciprocal monophyly between any of these morphologically defined species; instead, we observed geographic coherence for recovered lineages that did not correspond to current species hypotheses

for Amazonian and grassland *Sapajus*. The pattern is more concordant with an isolation-by-distance model across the entire 'widespread *Sapajus'* clade, and morphological variation driven by habitat type. The samples designated as *S. cay* formed a clade with geographically proximate *S. apella* samples, indicating either a high index of gene flow between the two, or that the two types actually are within the same species and have evolved phenotypic variation related to habitat type. Another possibility is that there is more than one taxon encompassed within the current taxonomic classification of *S. cay*. Some studies have already indicated that *S. cay* from the Brazilian Pantanal and from Paraguay may not be a monophyletic group (Casado et al., 2010; Lima et al., 2017), but in this study, we do not have samples from both areas. *S. macrocephalus* as defined by Rylands et al. (2013) is also paraphyletic in our study, with two distinct lineages, one found north of the Solimões and Japurá rivers and south of the Rio Negro (recovered as sister to *S. apella* north of the Amazon River: Figure 5c) and the other in south-central Amazon south of the Amazon and Solimões rivers (recovered as the sister group to south Amazonian *S. apella* and *S. cay*: Figure 5d). Note that our study extends the *S. macrocephalus* morphotype east of the Madeira River, into the Brazilian state of Rondônia. *S. apella* appears in multiple places across the topology of both the RAxML and SNP trees, divided among various lineages which do not form a monophyletic group, but instead are interspersed with clades of *S. libidinosus*, *S. macrocephalus*, and *S. cay*.

It is important to note that the geographic boundaries and taxonomic affinities for *S. apella, S. cay*, *S. libidinosus* and *S. macrocephalus* are disputed by the two predominant morphological authorities (Groves 2001, 2005; Silva-Júnior, 2001, 2002). For example, Groves (2001) considers *S. cay* as two distinct subspecies of *S. libidinosus* (called *Cebus libidinosus paraguayanus* and *Cebus libidinosus pallidus*), and *S. macrocephalus* as a subspecies of *S. apella* (*Cebus apella macrocephalus*). Neither mitochon-

481 drial (Lynch Alfaro et al., 2012a; Lima et al., 2017) nor nuclear data from the present
482 study recovered reciprocal monophyly for *S. cay*, *S. apella*, or *S. macrocephalus*. Com-
483 bining genetic and morphological data, we interpret that these morphotypes are not
484 clearly defined and discrete species, but instead form one morphologically diverse, re-
485 cently evolved pan-Amazonian plus grassland clade of robust capuchins. If we collapse
486 these three taxa into one species, the taxonomic name would be *Sapajus apella*, which
487 has priority over the other names because it was given first by Linnaeus in 1758. We do
488 not recommend the use of subspecies within this cosmopolitan species, because molecu-
489 lar and morphological subdivisions are discordant with one another suggesting a high
490 index of morphological plasticity and convergence within the species.

491 We also note that while the two major *Sapajus* clades within the Amazon are di-
492 vided roughly by the Amazon River (see Figures 5c and d), that some samples within
493 the Northern clade were from individuals south of the Amazon, and vice versa. In most
494 cases these were individuals that were very close geographically to the Amazon River
495 itself, and may be the result of human-mediated transport across the rivers in recent or
496 modern times. It is also possible that capuchins cross the Amazon at low frequency in
497 areas where there are many seasonal islands. Squirrel monkeys show a similar pattern in
498 the eastern Amazon basin, where the Amazon River forms the border for the distribu-
499 tions of *Saimiri sciureus* and *S. collinsi*, with some cases of limited dispersal to the op-
500 posite bank of the Amazon River for each species in the Juruti and Faro regions of Pará
501 State, Brazil (Mercês et al., 2015).

502 The time trees based generated from our BEAST analysis placed the mean esti-
503 mated divergence time for gracile and robust capuchins at 5.4 Ma using the *Aegypto-*
504 *pithecus* tree root prior, or 6.8 Ma, using the *Perupithecus* tree root prior. These com-
505 pare to previous mean estimates for divergence between *Cebus* and *Sapajus* at 5.8 Ma,

using mitochondrial data (Lima et al., 2017), at 6 Ma using a BEAST analysis for 54 nuclear genes (Perelman et al. 2011), and 6.6 Ma for the MCMC tree in PAML utilizing autocorrelated rates and soft-bounded constraints for a supermatrix of both nuclear and mitochondrial genes (Springer et al., 2012). In other words, all analyses converge on a late Miocene divergence time for robust and gracile capuchin monkeys. This timing is consistent with the formation of the savanna-like Cerrado leading to vicariance of a widespread capuchin ancestor previously spanning the Amazon to the Atlantic Forest (Lynch Alfaro et al., 2015; Lima et al., 2017).

5. Conclusions

Our phylogenomic data provided strong support for *Cebus* and *Sapajus* as two reciprocally monophyletic clades. This is concordant with morphological evaluations of distinctiveness between robust and gracile capuchins (Elliott, 1913; Hershkovitz, 1949; Groves, 2001, 2005; Silva-Júnior, 2001, 2002; Lynch Alfaro et al., 2012b), and mitochondrial and Alu element data that also point to this split (Lynch Alfaro et al., 2012a; Lima et al., 2017; Martins Jr. et al., 2015**;** Viana et al., 2015). We recovered a late Miocene split for robust and gracile capuchins, concordant with previous molecular studies. The timetree mean estimate for the initial diversification of robust capuchins was at 2.1 Ma (using the *Aegyptopithecus* root calibration) or 2.6 Ma (using the *Perupithecus* root calibration); this early Pleistocene diversification is also consistent with previous studies using mitochondrial data (Lynch Alfaro et al. 2012a; Lima et al., 2017).

In general, our phylogenies based on ultraconserved elements were congruent with mitochondrial phylogenies for robust capuchins (Lynch Alfaro et al., 2012; Lima et al., 2017), although the placement of *S. robustus* as sister to *S. xanthosternos* was

unique to the nuclear phylogenomic data, as was the recovery of a sister relationship between *S. flavius* and *S. libidinosus*. Our UCE tree distinguished only four *Sapajus* species, but the ExaBayes SNP tree provided more support for six robust capuchin species, *S. xanthosternos, S. robustus, S. nigritus, S. flavius, S. libidinosus*, and *S. apella* (which subsumes *S. cay* and *S. macrocephalus*), although *S. apella* morphotypes from Tucuruí were found within the *S. libidinosus* clade. The major division for Amazonian capuchins according to molecular data is a North-South division (both in the present work and from mitochondrial data in Lima et al., 2017), whereas the morphological division of *S. macrocephalus* and *S. apella* has been described as more of an East-West division, with the Madeira and Negro rivers as the suggested dividing line (Groves, 2001, 2005; Silva-Júnior, 2001, 2002). As morphological and molecular subdivisions of the Amazonian group are discordant, we recommend lumping all Amazonian plus southern grassland robust capuchin taxa as *S. apella* without subspecies. However, this does not discount the importance of populational differences in behavior, morphology and ecology in *S. apella* across the Amazon and southern grasslands; these populational differences may serve as a model for understanding the rapid evolution of populational differences across diverse habitats in other highly polymorphic species, such as humans.

The taxonomic relationship of *S. nigritus* to other capuchins is not well supported, with the species tree placing it as the sister group to *S. xanthosternos* + *S. robustus*, but the gene trees placing it as the sister group to the widespread clade of robust capuchins (*S. flavius, S. libidinosus, S. apella* as above). In contrast, mitochondrial phylogenetic reconstructions have placed *S. nigritus* as the sister to all other *Sapajus* (Lima et al., 2017). More work needs to be done delineating the relationship and geographical boundaries between *S. nigritus nigritus* from Minas Gerais to Sao Paulo, Brazil and *S. n. cucullatus* from southern Brazil and Argentina, and their relationships to other capu-

chins. Future work is also needed to determine the relationship of Critically Endangered *S. apella margaritae* endemic to Margarita Island, Venezuela to the other Amazonian and Guianan robust capuchins.

Acknowledgments

Special thanks to Stephen D. Nash/IUCN SSC Primate Specialist Group to Illustrations copyright 2013. Support to M.G.M.L.'s PhD research was provided by a CNPq PhD fellowship (142141/2012-7) and CNPq SWE fellowship (201172/2014-3). Some of the field expeditions were funded by CNPq/FAPEAM SISBIOTA Program (563348/2010-0) to J.P.B. UCE data was generated with support from NSF-FAPESP (grant 1241066 - Dimensions US-BIOTA-São Paulo: Assembly and evolution of the Amazonian biota and its environment: an integrated approach) to A.A. Divergence time analyses were performed with support from the Whitcome Research Fellowship to D.C.

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556. doi:10.1093/molbev/msu236

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114–2120.

576 Bond, M., Tejedor, M.F., Campbell Jr., K.E., Chornogubsky, L., Novo, N., Goin, F.,
577 2015. Eocene primates of South America and the African origins of New World
578 monkeys. Nature 520, 538–541. doi:10.1038/nature14120

579 Cabrera, A., 1957. Catálogo de los mamíferos de América del Sur. I (Metatheria, Un-
580 guiculata, Carnívora). Rev. Mus. Argentino Cien. Nat., Cien. Zool. 4, 1-307.

581 Casado, F., Bonvicino, C.R., Nagle, C., Comas, B., Manzur, T.D., Lahoz, M.M., Seu-
582 ánez, H.N., 2010. Mitochondrial divergence between 2 populations of the hood-
583 ed capuchin, *Cebus* (*Sapajus*) *cay* (Platyrrhini, Primates). J. Hered. 101, 261–9.
584 doi:10.1093/jhered/esp119

585 Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent
586 Model. Bioinformatics 30, 3317–3324. doi:10.1093/bioinformatics/btu530

587 Cole, T.M., 1992. Postnatal heterochrony of the masticatory apparatus in *Cebus apella*
588 and *Cebus albifrons*. J. Hum. Evol. 23 (3), 253–282.

589 Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K., Glenn,
590 T.C., 2012. More than 1000 ultraconserved elements provide evidence that tur-
591 tles are the sister group of archosaurs. Biol. Lett. 8, 783–6.
592 doi:10.1098/rsbl.2012.0331

593 Daegling, D.J., 1992. Mandibular morphology and diet in the genus *Cebus*. Int. J. Pri-
594 matol. 13, 545–570. doi:10.1007/BF02547832

595 Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics
596 with BEAUTi and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
597 doi:10.1093/molbev/mss075

598 Elliot, D. G., 1913. A Review of Primates. Monograph Series, American Museum of
599 Natural History, New York.

600 Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T.,
601 Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic mark-
602 ers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.
603 doi:10.1093/sysbio/sys004

604 Faircloth, B.C., 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and
605 quality trimming. http://dx.doi.org/10.6079/J9ILL.

606 Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved
607 genomic loci. Bioinformatics 32, 786–788.
608 doi:10.6084/m9.figshare.1284521.Contact

609 Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspec-
610 tive on the radiation of ray-finned fishes based upon targeted sequencing of ul-
611 traconserved elements (UCEs). PLoS One 8, e65923.
612 doi:10.1371/journal.pone.0065923

613 Ford, S.M., Hobbs, D.G., 1996. Species definition and differentiation as seen in the
614 postcranial skeleton of *Cebus*. In: Norconk MA, Rosenberger AL, Garber, PA,
615 editors. Adaptive radiations of neotropical primates. New York: Plenum Press.
616 pp. 229–249.

617 Fragaszy, D., Visalberghi, E., Fedigan, L., 2004. The complete capuchin. Cambridge:
618 Cambridge University Press. 339 p.

619 Frandsen, P.B., Calcott, B., Mayer, C., Lanfear, R., 2015. Automatic selection of parti-
620 tioning schemes for phylogenetic analyses using iterative *k*-means clustering of
621 site rates. BMC Evol. Biol. 15, 13. doi:10.1186/s12862-015-0283-7

622 Gernhard, T., 2008. The conditioned reconstructed process. J. Theor. Biol. 253, 769–
623 778. doi:10.1016/j.jtbi.2008.04.005

624 Groves, C.P., 2001. Primate Taxonomy. Smithsonian Institution Press, Washington,
625 DC.

626 Groves, C.P., 2005. Order Primates. In: Wilson, D.E., Reeder, D.M. (Eds.), Mammal
627 Species of the World: A Taxonomic and Geographic Reference, third ed., vol. 1.
628 Johns Hopkins University Press, Baltimore, MD, pp. 111–184.

629 Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The
630 Pennsylvania State University.

631 Hershkovitz, P., 1949. Mammals of northern Colombia. Preliminary report N. 4: mon-
632 keys (Primates), with taxonomic revisions of some forms. Proc. United States Natl.
633 Museum. 98, 323–427.

634 Hill, W.C.O., 1960. Primates: Comparative Anatomy and Taxonomy. IV. Cebidae, Part
635 A. Interscience, New York.

636 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software ver-
637 sion 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–
638 780. doi:10.1093/molbev/mst010

639 Kay, R.F., 2015. Biogeography in deep time – What do phylogenetics, geology, and
640 paleoclimate tell us about early platyrrhine evolution? Mol. Phylogenet. Evol.
641 82, 358–374. doi:10.1016/j.ympev.2013.12.002

642 Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined
643 selection of partitioning schemes and substitution models for phylogenetic anal-
644 yses. Mol. Biol. Evol. 29, 1695–1701. doi:10.1093/molbev/mss020

645 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFind-
646 er 2: New methods for selecting partitioned models of evolution for molecular
647 and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.
648 doi:10.1093/molbev/msw260

649 Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal
650 partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82.
651 doi:10.1186/1471-2148-14-82

652 Lee, T.H., Guo, H., Wang, X., Kim, C., Paterson, A.H,. 2014. SNPhylo: a pipeline to
653 construct a phylogenetic tree from huge SNP data. BMC Genomics 15, 162.
654 doi:10.1186/1471-2164-15-162

655 Lima, M.G.M., Buckner, J.C., Silva-Júnior, J. de S. e, Aleixo, A., Martins, A.B., Bou-
656 bli, J.P., Link, A., Farias, I.P., da Silva, M.N., Röhe, F., Queiroz, H., Chiou,
657 K.L., Di Fiore, A., Alfaro, M.E., Lynch Alfaro, J.W., 2017. Capuchin monkey
658 biogeography: understanding *Sapajus* Pleistocene range expansion and the cur-
659 rent sympatry between *Cebus* and *Sapajus*. J. Biogeogr. 1–11.
660 doi:10.1111/jbi.12945

661 Lynch Alfaro, J.W., Boubli, J.P., Olson, L.E., Di Fiore, A., Wilson, B., Gutiérrez-
662 Espeleta, G.A., Chiou, K.L., Schulte, M., Neitzel, S., Ross, V., Schwochow, D.,
663 Nguyen, M.T.T., Farias, I., Janson, C.H., Alfaro, M.E., 2012a. Explosive Pleis-
664 tocene range expansion leads to widespread Amazonian sympatry between ro-
665 bust and gracile capuchin monkeys. J. Biogeogr. 39, 272–288.

666 Lynch Alfaro, J.W., Silva-Junior, J. S., Rylands, A.B., 2012b. How different are robust
667 and gracile capuchin monkeys? An argument for the use of *Sapajus* and *Cebus*.
668 Am. J. Primatol. 74, 273–86.

669 Lynch Alfaro, J.W., Izar, P., Ferreira, R.G., 2014. Capuchin monkey research priorities
670 and urgent issues. Am. J. Primatol. 76, 705–720. doi:10.1002/ajp.22269.

671 Lynch Alfaro J.W., Cortés-Ortiz L., Di Fiore A., Boubli J.P. 2015. Special issue: Com-
672 parative biogeography of Neotropical primates. Mol. Phylogenet. Evol. 82, 518-529.

673 Martins Jr., A.M.G., Amorim, N., Carneiro, J.C., de Mello Affonso, P.R.A., Sampaio,
674 I., Schneider, H., 2015. Alu elements and the phylogeny of capuchin (*Cebus* and
675 *Sapajus*) monkeys. Am. J. Primatol. 77, 368–375. doi:10.1002/ajp.22352

676 Masterson, TJ., 1997. Sexual dimorphism and interspecific cranial form in two capuchin
677 species: *Cebus albifrons* and *C. apella*. Am. J. Phys. Anthropol. 104, 487–511.
678 doi:10.1002/(SICI)1096-8644(199712)104:4<487::AID-AJPA5>3.0.CO;2-P

679 McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T.,
680 Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers
681 that resolve placental mammal phylogeny when combined with species-tree
682 analysis. Genome Res. 22, 746–54. doi:10.1101/gr.125864.111

683 McCormack, J., Tsai, W.L.E., Faircloth, B.C., 2015. Sequence capture of ultracon-
684 served elements from bird museum specimens. Mol. Ecol. Resour. 1–15.
685 doi:10.1101/020271

686 Mercês, M.P., Lynch Alfaro, J.W., Ferreira, W.A.S., Harada, M.L., Silva-Júnior, J.S.,
687 2015. Morphology and mitochondrial phylogenetics reveal that the Amazon
688 River separates two eastern squirrel monkey species: *Saimiri sciureus* and *S.*
689 *collinsi*. Mol. Phylogenet. Evol. 82 (PB), 426–435.

690 Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A.M.,
691 Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P.C., Silva, A.,
692 O'Brien, S.J., Pecon-Slattery, J., 2011. A molecular phylogeny of living primates.
693 PLoS Genet. 7, e1001342. doi:10.1371/journal.pgen.1001342

694 Rambaut, A., Drummond, A.J., Suchard, M., 2013. Tracer v1.6. Available from
695 http://tree.bio.ed.ac.uk/software/tracer/

696 Rambaut, A., Drummond, A.J., 2015a. TreeAnnotator v1.8.3. Available from
697 http://beast.bio.ed.ac.uk/downloads/

698 Rambaut, A., Drummond, A.J., 2015b. LogCombiner v1.8.3. Available from
699 http://beast.bio.ed.ac.uk/downloads/

700 Ruiz-García, M., Castillo, M.I., Lichilín-Ortiz, N., Pinedo-Castro, M., 2012. Molecular
701 Relationships and Classification of Several Tufted Capuchin Lineages (*Cebus*
702 *apella*, *Cebus xanthosternos* and *Cebus nigritus*, Cebidae), by Means of Mito-
703 chondrial Cytochrome Oxidase II Gene Sequence. Folia Primatol. 83, 100–125.
704 doi:10.1159/000342832

Rylands, A.B., Mittermeier, R.A., 2009. The diversity of the New World primates (Platyrrhini): An annotated taxonomy. In: Garber, P.A., Estrada, A., Bicca-Marques, J.C., Heymann, E.W., Strier, K.B. South American Primates: Comparative Perspectives in the Study of Behavior, Ecology and Conservation. Springer Science+Business Media, LLC. pp. 23–54.

Rylands, A., Kierulff, M., Mittermeier, R., 2005. Notes on the taxonomy and distributions of the tufted capuchin monkeys (*Cebus*, Cebidae) of South America. Lundiana 6, 97–110.

Rylands, A.B., Mittermeier, R.A., Silva-Júnior, J.S., 2012. Neotropical primates: taxonomy and recently described species and subspecies. Int. Zoo Yearb. 46, 11–24.

Rylands, A.B., Mittermeier, R.A., Bezerra, B.M., Paim, F.P., Queiroz, H.L., 2013. Species accounts of Cebidae. In: Mittermeier, R.A., Rylands, A.B., Wilson, D.E. (Eds.), Handbook of the Mammals of the World, vol. 3. Primates. Lynx Edicions, Barcelona, pp. 390–413

Silva Jr, J.S., 2001. Especiação nos Macacos-prego e Caiararas, Gênero Cebus Erxleben, 1777 (Primates, Cebidae). PhD Thesis: Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Silva-Júnior, J.S., 2002. Taxonomy of capuchin monkeys, *Cebus* Erxleben, 1777. Neotrop. Primates 10, 29.

Silva Jr, J.S., 2005. Especiação nos Macacos-prego e Caiararas, gênero *Cebus* Erxleben, 1777 (Primates, Cebidae). Bol. Soc. Bras. Mastoz. 42, 11-12.

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., Fisher, C.A., Murphy, W.J.,

2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS ONE 7, e49521.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Torres de Assumpção, C., 1983. An ecological study of the primates of southeastern Brazil, with a reappraisal of *Cebus apella* races. PhD Thesis, Edinburgh: University of Edinburgh.

Viana, M.C., Menezes, A.N., Moreira, M.A.M., Pissinatti, A., Seuánez, H.N., 2015. MECP2, a gene associated with Rett syndrome in humans, shows conserved coding regions, independent Alu insertions, and a novel transcript across primate evolution. BMC Genet. 16, 77. doi:10.1186/s12863-015-0240-x

Warnock, R.C.M., Yang, Z., Donoghue, P.C.J., 2012. Exploring uncertainty in the calibration of the molecular clock. Biol. Lett. 8, 156–159. doi:10.1098/rsbl.2011.0710

Wright, B.W., 2005a. Craniodental biomechanics and dietary toughness in the genus *Cebus*. J. Hum. Evol. 48, 473–492. doi:10.1016/j.jhevol.2005.01.006.

Wright, K.A., 2005b. Interspecific and ontogenetic variation in locomotor behavior, habitat use, and postcranial morphology in *Cebus apella* and *Cebus olivaceus* [dissertation]. Evanston (IL): Northwestern University. 433p.

749 Wright, K.A., 2007. The relationship between locomotor behavior and limb morpholo-
750 gy in brown (*Cebus apella*) and weeper (*Cebus olivaceus*) capuchins. Am. J.
751 Primatol. 69, 736–756. doi:10.1002/ajp.20391

752 Wright, K.A., Wright, B.W., Ford, S.M., Fragaszy, D., Izar, P., Norconk, M., Master-
753 son, T., Hobbs, D.G., Alfaro, M.E., Lynch Alfaro, J.W., 2015. The effects of
754 ecology and evolutionary history on robust capuchin morphological diversity.
755 Mol. Phylogenet. Evol. 82, 455–466.

Tables

Table 1: Taxonomies of robust capuchins.

Elliot (1913)	Hershkovitz (1949)	Cabrera (1957)	Hill (1960)	Groves (2001, 2005)	
Cebus apella	*Cebus apella*	*Cebus apella*	*Cebus apella*	*Cebus apella*	
Cebus fatuellus		*C. a. apella*	*C. a. apella*	*C. a. apella*	
C. f. fatuellus		*C. a. margaritae*	*C. a. margaritae*	*C. a. fatuellus*	
C. f. peruanus		*C. a. macrocephalus*	*C. a. fatuellus*	*C. a. macrocephalus*	
Cebus macrocephalus		*C. a. libidinosus*	*C. a. peruanus*	*C. a. peruanus*	
Cebus libidinosus		*C. a. paraguayanus*	*C. a. tocantinus*	*C. a. tocantinus*	
Cebus azarae		*C. a. pallidus*	*C. a. macrocephalus*	*C. a. margaritae*	
C. a. azarae		*C. a. xanthosternos*	*C .a. libidinosus*	*Cebus libidinosus*	
C. a. pallidus		*C. a. versutus*	*C. a. cay*	*C. l. libidinosus*	
Cebus frontatus		*C. a. nigritus*	*C. a. pallidus*	*C. l. pallidus*	
Cebus variegatus		*C. a. vellerosus*	*C. a. frontatus*	*C. l. paraguayanus*	
Cebus versuta		*C. a. robustus*	*C. a. xanthosternos*	*C. l. juruanus*	
Cebus cirrifer			*C. a. nigritus*	*Cebus nigritus*	
Cebus crassiceps			*C. a. robustus*	*C. n. nigritus*	
Cebus caliginosus			*C. a. magnus*	*C. n. robustus*	
Cebus vellerosus			*C. a. juruanus*	*C. n. cucullatus*	
			C. a. maranonis	*Cebus xanthosternos*	

763

764

765

766

767

768

769 **Table 2:** List of samples, locality data and resulting for UCE data.

Code	Species	Latitude	Longitude	Trimmed reads	Contigs Assembled
1	*S. xanthosternos*	-15.17	-39.07	2681597	3274
2	*S. xanthosternos*	-15.41	-39.5	2843593	3661
3A	*S. xanthosternos*	-14.79	-39.05	3196673	3802
3B	*S. xanthosternos*	-14.79	-39.05	3521726	4275
4	*S. robustus*	-19.95	-43.85	4538948	5198
5	*S. nigritus*	-23.86	-46.14	2762021	3471
6	*S. nigritus*	-23	-49.32	946881	1937
7	*S. flavius*	-6.56	-35.13	2713906	3096
8	*S. flavius*	-7.01	-34.96	4787966	5150
9	*S. flavius*	-7.02	-35.09	2877922	3601
10	*S. libidinosus*	-2.77	-41.81	2764451	3430
11	*S. libidinosus*	-2.8	-41.87	4348317	5094
12	*S. libidinosus*	-5.09	-42.43	2612178	3208
13	*S. libidinosus*	-7.93	-44.2	3068523	3551
14	*S. libidinosus*	-5.28	-48.3	3303530	3885
15	*S. libidinosus*	-14.14	-48.17	3381894	3603
16	*S. libidinosus*	-16.6	-49.26	3301692	3884
17A	*S. apella*	-3.83	-49.64	3541159	3793
17B	*S. apella*	-3.83	-49.64	2980533	3534

18	*S. apella*	-6.15	-49.56	1908769	2828
19	*S. apella*	-3.36	-51.74	3391742	3723
20	*S. apella*	-2.61	-51.54	5485708	6170
21	*S. apella*	-0.58	-52.33	1311929	2137
22	*S. apella*	3.22	-52.03	1757726	2338
23	*S. apella*	0.83	-53.93	2781762	2805
24	*S. apella*	1.29	-58.7	2130450	2604
25	*S. apella*	-1.49	-56.8	1572934	2413
26	*S. apella*	-2.47	-58.4	3571090	3780
27	*S. apella*	-2.6	-56.18	2394355	3227
28	*S. apella*	-3.18	-55.8	1890413	2709
29	*S. apella*	-3.88	-56.78	1276241	2039
30	*S. apella*	-4.71	-56.44	1746336	2515
31	*S. apella*	-10	-56.04	1791793	2450
32	*S. apella*	-9.2	-59.06	2103015	2895
33	*S. apella*	-12.03	-60.67	2339872	3027
34	*S. apella*	-12.56	-63.44	3883141	4558
35	*S. cay*	-16.06	-57.72	1624662	2588
36	*S. cay*	-13.52	-60.43	2361492	2991
37	*S. macrocephalus*	-12.45	-62.92	2986344	3335
38	*S. macrocephalus*	-8.67	-62.37	2962283	3477
39	*S. macrocephalus*	-9.1	-62.88	2222218	2882
40	*S. macrocephalus*	-8.89	-63.24	3054313	3411
41	*S. macrocephalus*	-8.8	-63.95	1459387	2148
42	*S. macrocephalus*	-8.19	-64.02	2196025	2741
43	*S. macrocephalus*	-5.69	-63.24	3840307	4395
44A	*S. macrocephalus*	-4.99	-62.96	3199632	3780
44B	*S. macrocephalus*	-4.99	-62.96	1163783	2218
45	*S. macrocephalus*	-4.75	-61.28	2351064	3072

46	*S. macrocephalus*	-4.44	-60.32	2219015	2938
47	*S. macrocephalus*	-3.37	-60.48	1876035	2707
48	*S. macrocephalus*	-1.05	-62.89	2044899	2699
49	*S. macrocephalus*	-0.48	-64.41	2723327	3234
50	*S. macrocephalus*	-0.61	-64.92	3169376	3983
51	*S. macrocephalus*	-0.23	-66.85	2105443	2681
52	*S. macrocephalus*	-2.47	-64.83	3117247	3756
53	*S. macrocephalus*	-2.59	-64.89	2484843	2946
54	*S. macrocephalus*	-2.45	-65.36	1918138	2692
55	*S. macrocephalus*	-1.84	-69.03	2085573	2716
56	*S. macrocephalus*	-4.4	-70.14	3522837	4000
57	*S. macrocephalus*	-4.94	-68.17	4107017	4659
-	*C. unicolor*	-9.22	-66.74	2057387	3279
-	*C. o. castaneus*	-0.58	-52.33	2107696	3145
-	*C. o. castaneus*	1.84	-52.74	1401630	2151
-	*C. kaapori*	-2.33	-46.08	2885841	3593
-	*C. capucinus*	10.95	-84.55	3954729	4702
-	*C. capucinus*	10.88	-85.78	508807	1162
-	*C. albifrons*	-2.59	-64.89	3111458	3951

Table 3: Fossil calibrations used for divergence time estimation (see Supplementary Figure 1 for node labels).

Calibrated node	Divergence	Fossil	Reference
1	Hominina / *Pan*	*Ardipithecus kadabba*	Springer et al., 2012
4	Hominini / *Gorilla*	*Sivapithecus* sp.	Springer et al., 2012
5	*Papio* / *Macaca*	*Macaca libyca*	Springer et al., 2012
6	Hominoidea / Cercopithecidae	*Afropithecus turkanensis*	Springer et al., 2012

10	Callitrichidae / Cebidae (*sensu* Rylands et al., 2012)	*Patasola magdalenae*; *Lagonimico conclucatus*	Kay, 2015 (minimum); Springer et al., 2012 (maxi
11	Catarrhini / Platyrrhini	*Aegyptopithecus zeuxis* / *Perupithecus ucayaliensis*	Springer et al., 2012 / Bond et al., 2015

773

774

775

776

777

778

779

780 **Table 4:** Summary of the posterior distribution of divergence times (in Ma) estimated

781 using BEAST (see Supplementary Figure 1 for node labels).

Node	Springer et al. root calibration			
	Median	Mean	95% HPD	Median
1	5.6	5.8	5.1–7.3	5.7
2	7.9	8.2	5.7–11.1	8.6
3	14.2	14.5	10.1–19.5	11.4
4	17.4	17.7	12.1–23.2	15.5
5	6.5	6.9	5.5–9.6	6.3
6	23.6	24.4	20.6–30.8	25.8
7	1.7	1.8	0.4–3.4	2.2
8	1.9	2.1	0.6–4.0	2.4
9	5.2	5.4	3.2–8.0	6.6
10	17.0	17.4	13.4–22.1	14.9

11	34.1	35.4	28.3–46.7	41.4

Figure Captions

Graphical Abstract. (a) Maximum likelihood and **(b)** Bayesian inference for robust capuchin phylogeny based on SNP data.

Figure 1. Map showing the sampled localities for *Sapajus*

Figure 2. Maximum likelihood (RAxML) 75% phylogeny for UCE data.

Figure 3. (a) Maximum likelihood and **(b)** Bayesian inference for robust capuchin phylogeny based on SNP data.

Figure 4. Species tree for robust capuchins using SNP quartets.

Figure 5. (a) Map with minimum convex polygons to show geographic distribution of major subclades within the widespread *Sapajus* clade, **(b)** Minimum convex polygon for range distribution for *S. flavius* and *S. libidinosus* clades within the ExaBayes phylogeny, **(c)** Minimum convex polygon for range distribution for the Northern Amazonian *Sapajus* clade within the ExaBayes phylogeny and **(d).**Minimum convex polygon for range distribution for the Southern Amazonian *Sapajus* clade within the ExaBayes phylogeny. Larger map depicts subclades of south central Amazonian *S. macrocephalus* and southern Amazonian + grasslands *S. apella + cay.*

Supplementary Figure 1. Topological constraint used for divergence time estimation in BEAST.

Supplementary Figure 2. Maximum likelihood (RaxML) 95% phylogeny for UCE data.

Graphical Abstract.

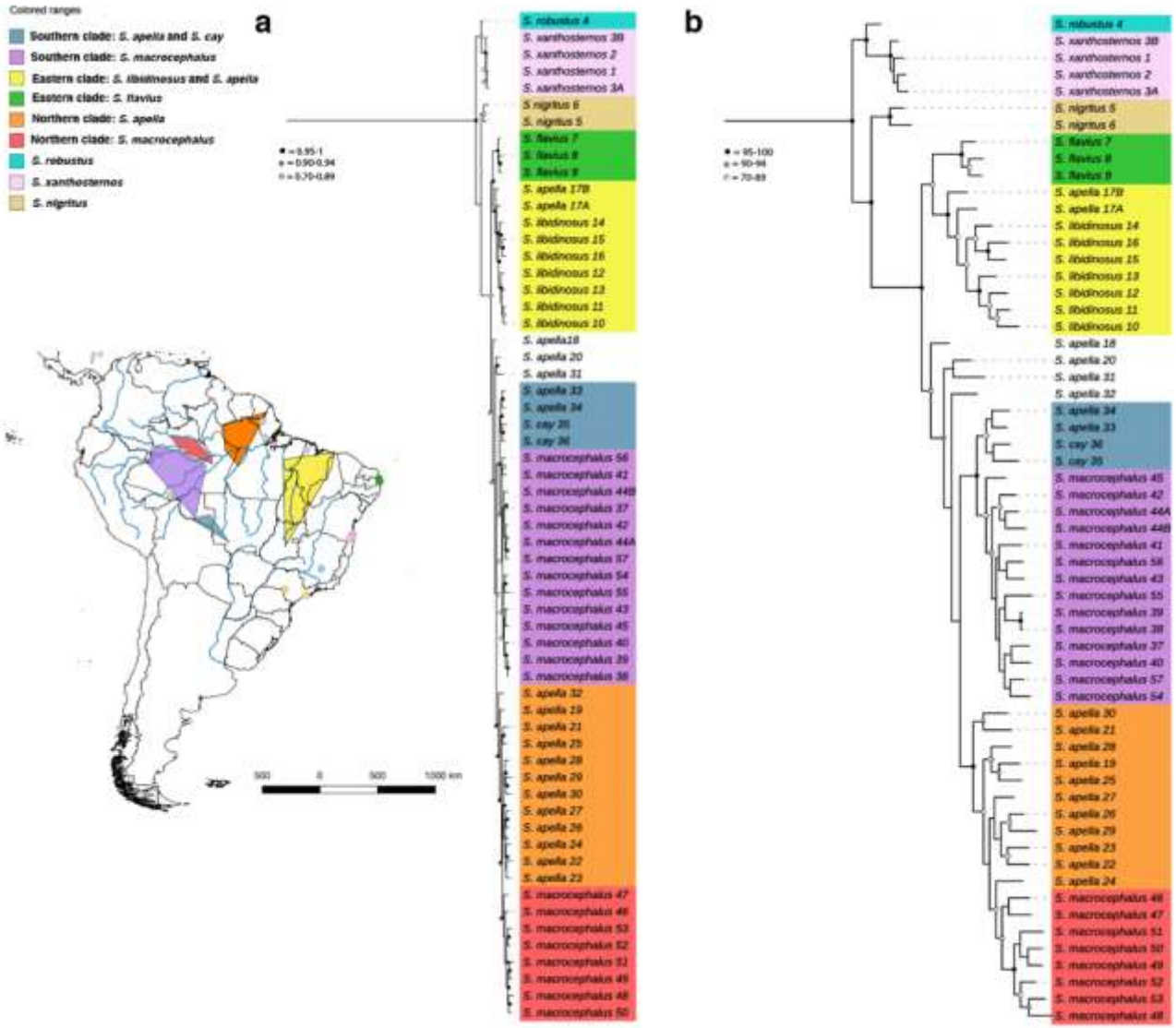


Figure 1.

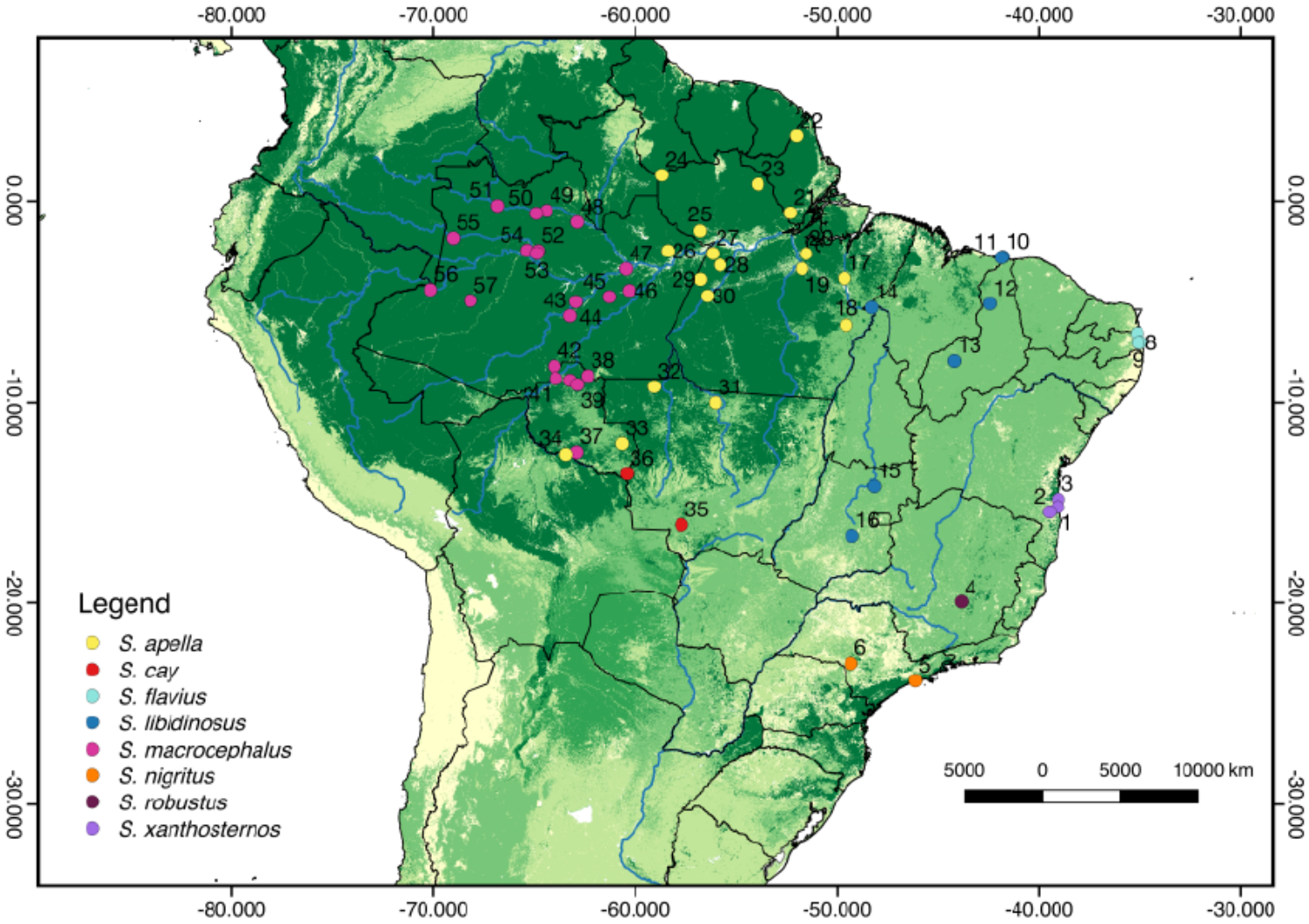

-80.000
-70.000
-60.000
-50.000
-40.000
-30.000
0.000
-10.000
-20.000
-30.000
Legend
S. apella
S. cay
S. flavius
S. libidinosus
S. macrocephalus
S. nigritus
S. robustus
S. xanthosternos
5000
0
5000
10000 km

Figure 2.

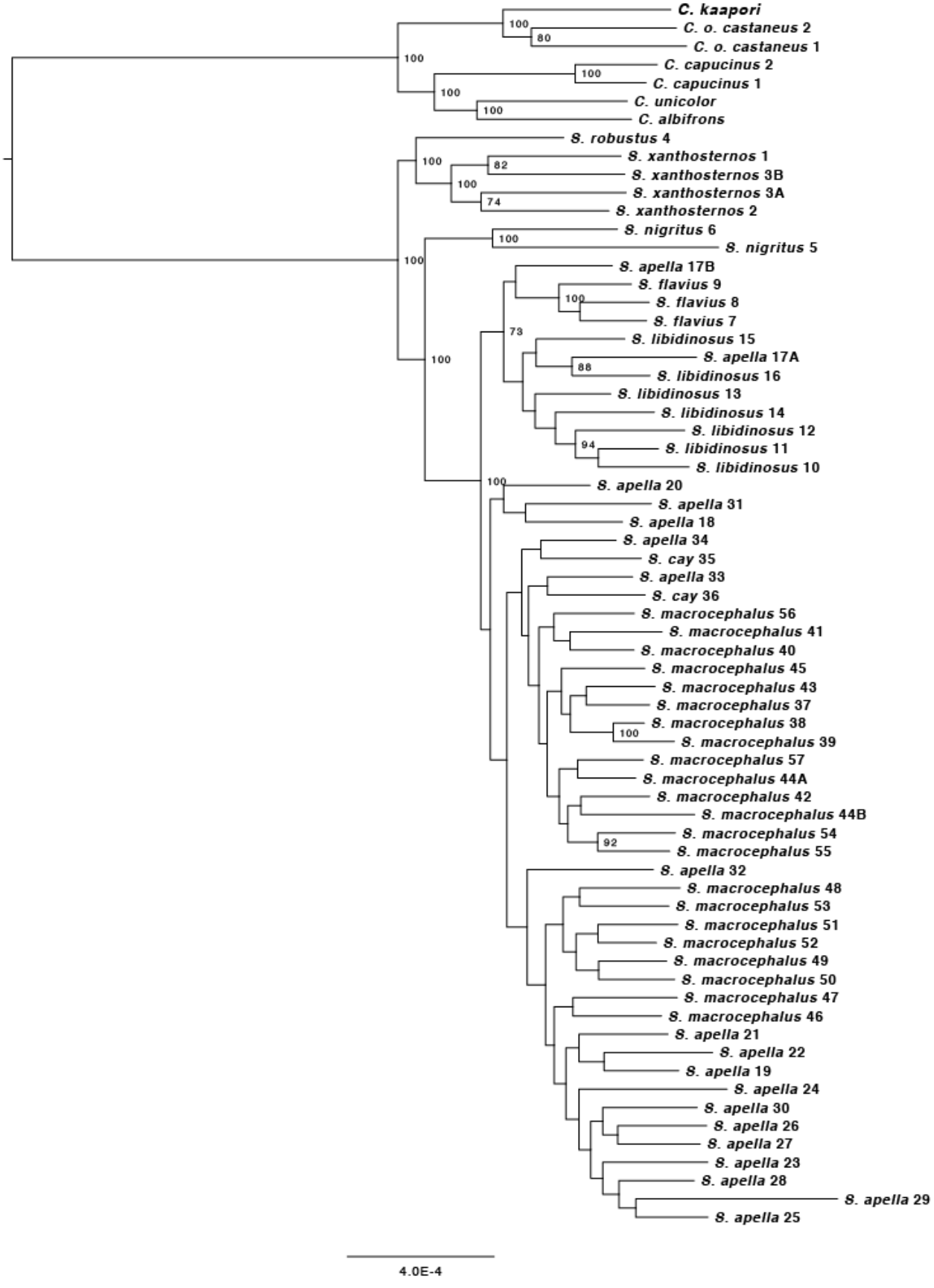

C. kaapori
C. o. castaneus 2
C. o. castaneus 1
C. capucinus 2
C. capucinus 1
C. unicolor
C. albifrons
S. robustus 4
S. xanthosternos 1
S. xanthosternos 3B
S. xanthosternos 3A
S. xanthosternos 2
S. nigritus 6
S. nigritus 5
S. apella 17B
S. flavius 9
S. flavius 8
S. flavius 7
S. libidinosus 15
S. apella 17A
S. libidinosus 16
S. libidinosus 13
S. libidinosus 14
S. libidinosus 12
S. libidinosus 11
S. libidinosus 10
S. apella 20
S. apella 31
S. apella 18
S. apella 34
S. cay 35
S. apella 33
S. cay 36
S. macrocephalus 56
S. macrocephalus 41
S. macrocephalus 40
S. macrocephalus 45
S. macrocephalus 43
S. macrocephalus 37
S. macrocephalus 38
S. macrocephalus 39
S. macrocephalus 57
S. macrocephalus 44A
S. macrocephalus 42
S. macrocephalus 44B
S. macrocephalus 54
S. macrocephalus 55
S. apella 32
S. macrocephalus 48
S. macrocephalus 53
S. macrocephalus 51
S. macrocephalus 52
S. macrocephalus 49
S. macrocephalus 50
S. macrocephalus 47
S. macrocephalus 46
S. apella 21
S. apella 22
S. apella 19
S. apella 24
S. apella 30
S. apella 26
S. apella 27
S. apella 23
S. apella 28
S. apella 29
S. apella 25
100
80
100
100
100
100
82
100
74
100
100
100
73
100
86
94
100
92
4.0E-4

Figure 3.

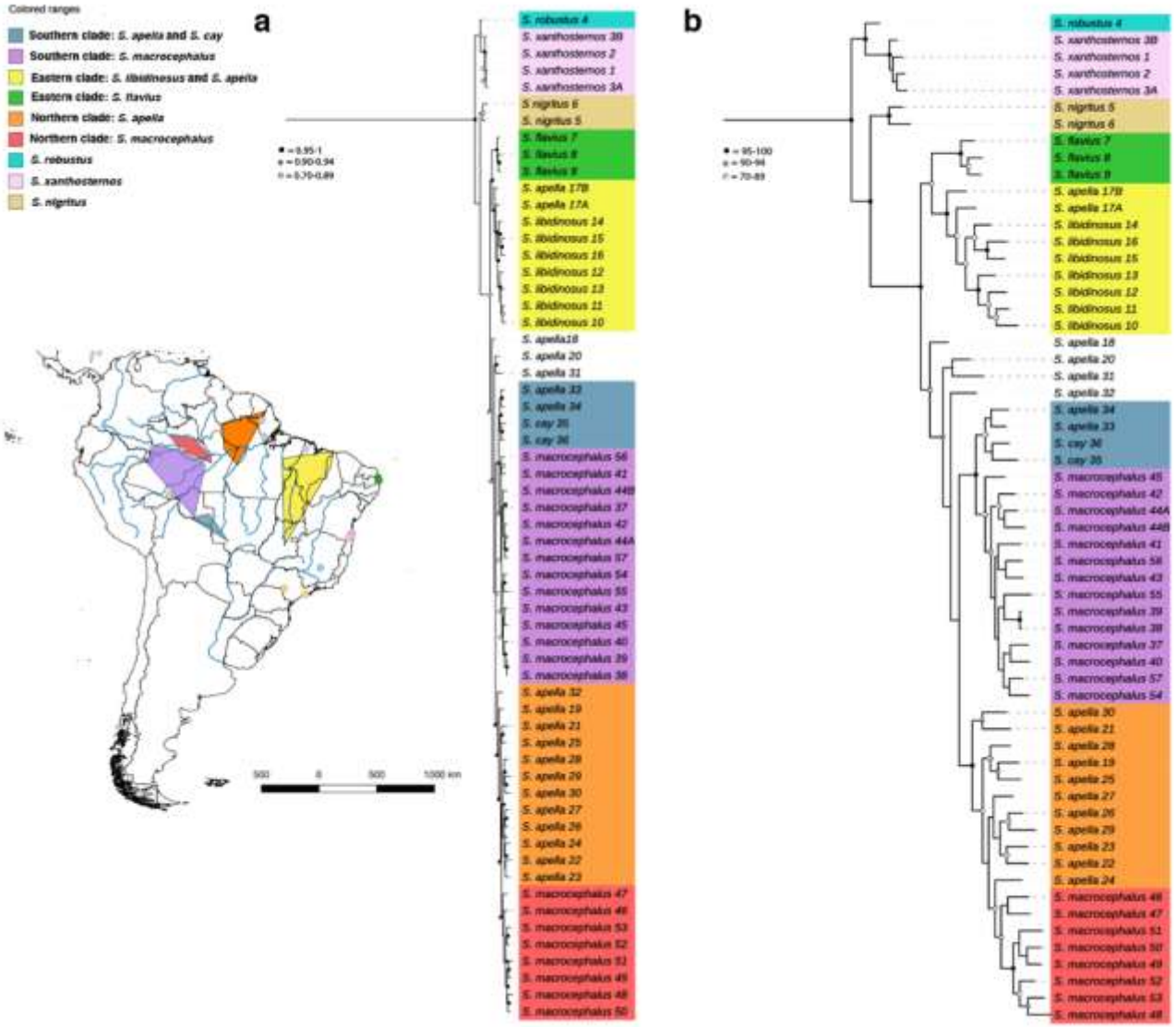

Colored ranges
Southern clade: S. apelta and S. cay
Southern clade: S. macrocephalus
Eastern clade: S. libidinosus and S. apelta
Eastern clade: S. flavius
Northern clade: S. apelta
Northern clade: S. macrocephalus
S. robustus
S. xanthostemos
S. nigritus
a
b
S. robustus 4
S. xanthostemos 3B
S. xanthostemos 2
S. xanthostemos 1
S. xanthostemos 3A
S. nigritus 6
S. nigritus 5
S. flavius 7
S. flavius 8
S. flavius 9
S. apelta 17B
S. apelta 17A
S. libidinosus 14
S. libidinosus 15
S. libidinosus 16
S. libidinosus 12
S. libidinosus 13
S. libidinosus 11
S. libidinosus 10
S. apelta18
S. apelta 20
S. apelta 31
S. apelta 33
S. apelta 34
S. cay 35
S. cay 36
S. macrocephalus 56
S. macrocephalus 41
S. macrocephalus 44B
S. macrocephalus 37
S. macrocephalus 42
S. macrocephalus 44A
S. macrocephalus 57
S. macrocephalus 54
S. macrocephalus 55
S. macrocephalus 43
S. macrocephalus 45
S. macrocephalus 40
S. macrocephalus 39
S. macrocephalus 38
S. apelta 32
S. apelta 30
S. apelta 21
S. apelta 28
S. apelta 29
S. apelta 27
S. apelta 26
S. apelta 24
S. apelta 22
S. apelta 23
S. macrocephalus 47
S. macrocephalus 46
S. macrocephalus 53
S. macrocephalus 52
S. macrocephalus 51
S. macrocephalus 49
S. macrocephalus 48
S. macrocephalus 50
500 0 500 1000 km

Figure 4.

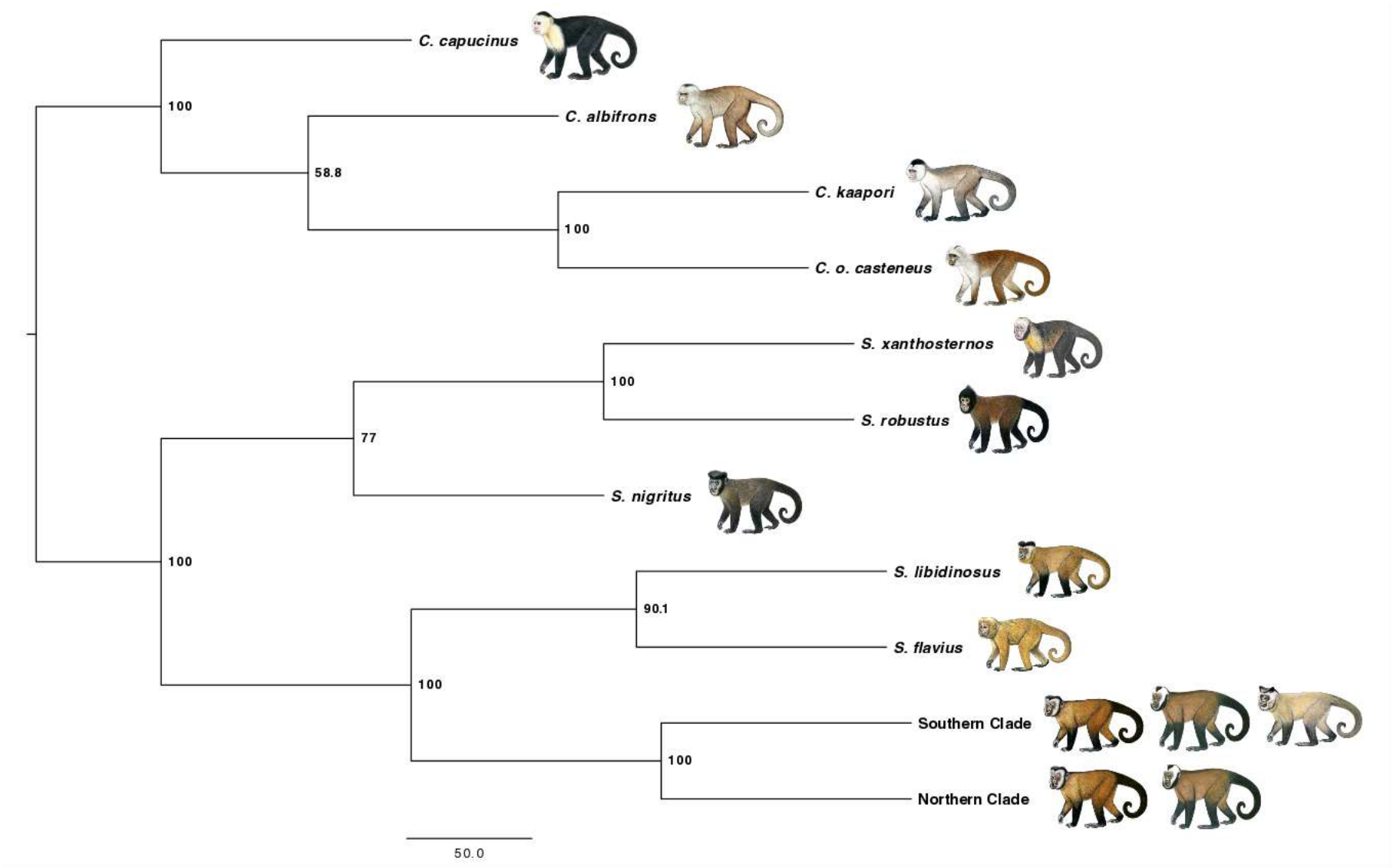
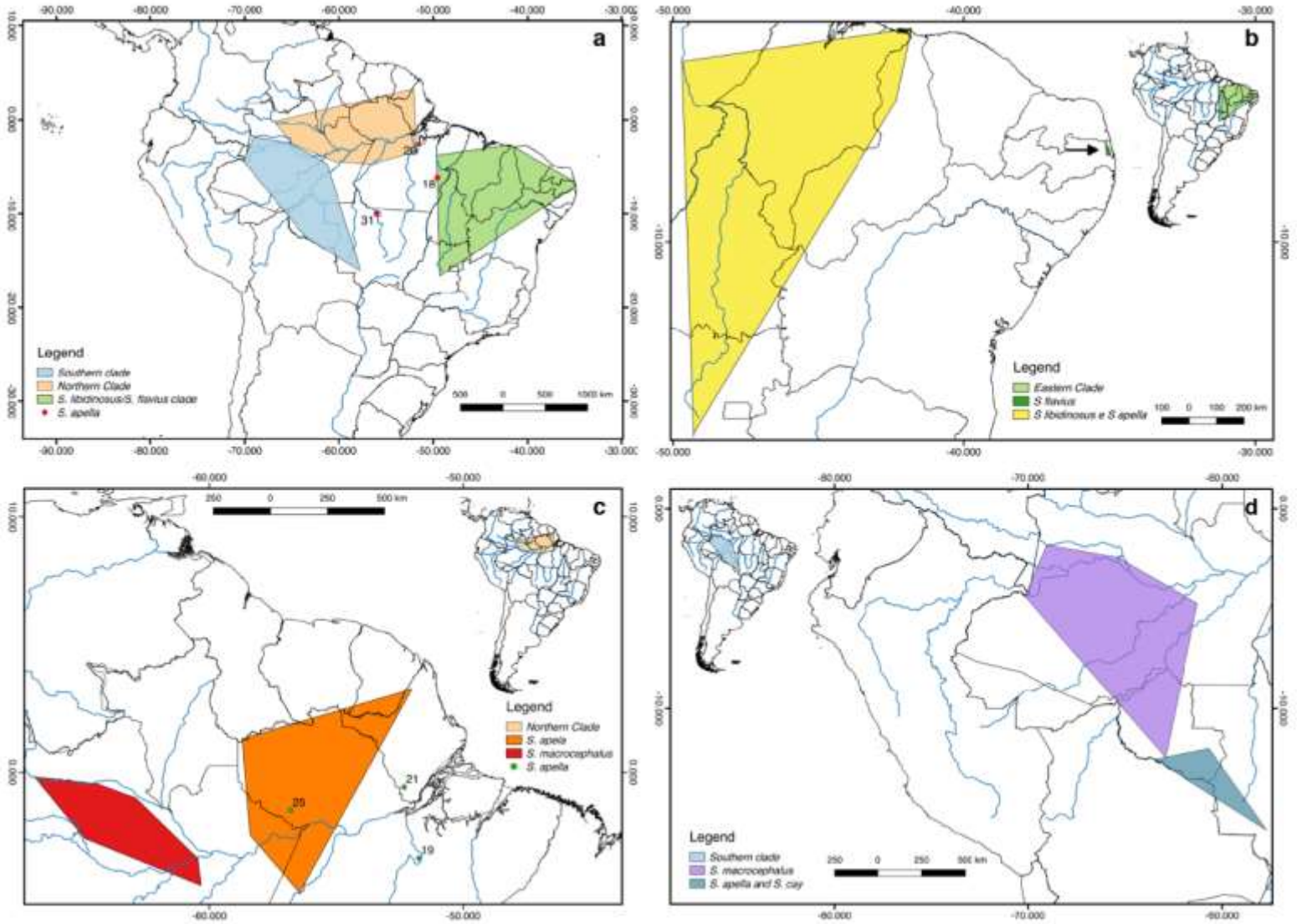
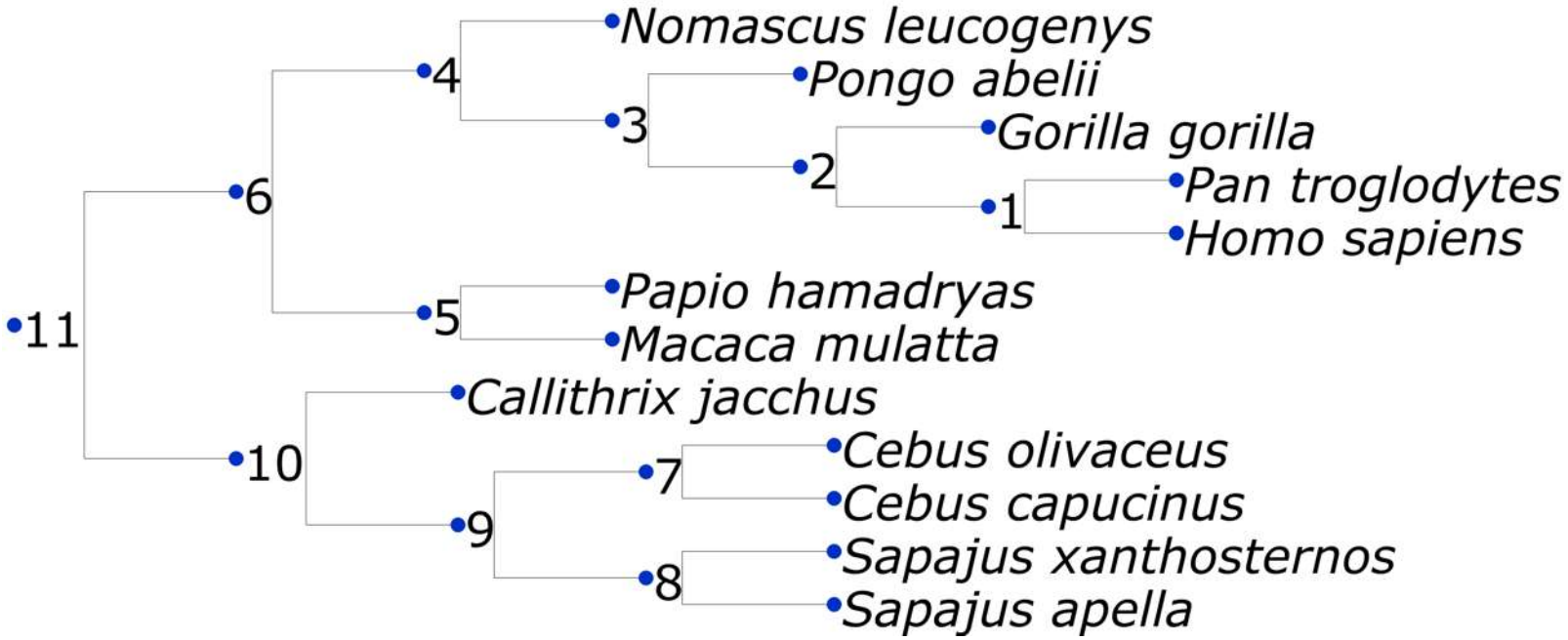

C. capucinus
C. albifrons
C. kaapori
C. o. casteneus
S. xanthosternos
S. robustus
S. nigritus
S. libidinosus
S. flavius
Southern Clade
Northern Clade
100
58.8
100
100
77
100
90.1
100
100
50.0

Figure 5.


a
Legend
Southern clade
Northern Clade
S. libidinosus/S. flavus clade
S. apella
b
Legend
Eastern Clade
S flavus
S libidinosus e S apella
c
Legend
Northern Clade
S. apela
S. macrocephalus
S. apella
d
Legend
Southern clade
S. macrocephalus
S. apella and S. cay

Supplementary Figure 1.



Supplementary Figure 2.