



Transcriptome of the Southern Muriqui *Brachyteles arachnoides* (Primates:Platyrrhini), a Critically Endangered New World Monkey: Evidence of Adaptive Evolution

Daniel A. Moreira¹, Alessandra P. Lamarca², Rafael Ferreira Soares³, Ana M. A. Coelho⁴, Carolina Furtado⁵, Nicole M. Scherer¹, Miguel A. M. Moreira⁵, Hector N. Seuánez^{5*} and Mariana Boroni^{1*}

¹ Laboratory of Bioinformatics and Computational Biology, Division of Experimental and Translational Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil, ² Laboratory of Bioinformatics and Molecular Evolution, Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, ³ Laboratory for Functional Genomics and Bioinformatics, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil, ⁴ Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, ⁵ Genetics Program, Division of Experimental and Translational Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil

OPEN ACCESS

Edited by:

TingFung Chan,
The Chinese University of Hong Kong,
China

Reviewed by:

Dapeng Wang,
University of Leeds, United Kingdom
Jeffrey Rogers,
Baylor College of Medicine,
United States

*Correspondence:

Hector N. Seuánez
hseuanez@inca.gov.br
Mariana Boroni
mariana.boroni@inca.gov.br

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 16 April 2020

Accepted: 09 July 2020

Published: 31 July 2020

Citation:

Moreira DA, Lamarca AP,
Soares RF, Coelho AMA, Furtado C,
Scherer NM, Moreira MAM,
Seuánez HN and Boroni M (2020)
Transcriptome of the Southern
Muriqui *Brachyteles arachnoides*
(Primates:Platyrrhini), a Critically
Endangered New World Monkey:
Evidence of Adaptive Evolution.
Front. Genet. 11:831.
doi: 10.3389/fgene.2020.00831

The southern muriqui (*Brachyteles arachnoides*) is the largest neotropical primate. This species is endemic to Brazil and is currently critically endangered due to its habitat destruction. The genetic basis underlying adaptive traits of New World monkeys has been a subject of interest to several investigators, with significant concern about genes related to the immune system. In the absence of a reference genome, RNA-seq and *de novo* transcriptome assembly have proved to be valuable genetic procedures for accessing gene sequences and testing evolutionary hypotheses. We present here a first report on the sequencing, assembly, annotation and adaptive selection analysis for thousands of transcripts of *B. arachnoides* from two different samples, corresponding to 13 different blood cells and fibroblasts. We assembled 284,283 transcripts with N50 of 2,940 bp, with a high rate of complete transcripts, with a median high scoring pair coverage of 88.2%, including low expressed transcripts, accounting for 72.3% of complete BUSCOs. We could predict and extract 81,400 coding sequences with 79.8% of significant BLAST hit against the Euarchontoglires SwissProt dataset. Of these 64,929 sequences, 34,084 were considered homologous to Supraprimate proteins, and of the remaining sequences (30,845), 94% were associated with a protein domain or a KEGG Orthology group, indicating potentially novel or specific protein-coding genes of *B. arachnoides*. We use the predicted protein sequences to perform a comparative analysis with 10 other primates. This analysis revealed, for the first time in an Atelid species, an expansion of *APOBEC3G*, extending this knowledge to all NWM families. Using a branch-site model, we searched for evidence of positive selection in 4,533 orthologous sets. This evolutionary analysis revealed 132 amino acid sites in 30 genes potentially evolving under positive selection, shedding light on primate genome evolution. These genes belonged to a wide variety of categories, including those encoding the innate immune system proteins (*APOBEC3G*, *OAS2*, and

CEACAM1) among others related to the immune response. This work generated a set of thousands of complete sequences that can be used in other studies on molecular evolution and may help to unveil the evolution of primate genes. Still, further functional studies are required to provide an understanding of the underlying evolutionary forces modeling the primate genome.

Keywords: RNA-seq, *de novo* transcriptome assembly, primate, positive selection, immune system

INTRODUCTION

The southern marmoset (*Brachyteles arachnoides*), also known as the woolly spider monkey, is a neotropical primate species endemic to Brazil and distributed in the states of Paraná, São Paulo, Rio de Janeiro and Minas (Koehler et al., 2005; Talebi and Soares, 2005). The genus *Brachyteles* is composed of two species, *B. arachnoides* and *Brachyteles hypoxanthus* (Chaves et al., 2019). The southern and northern marmosets are the largest neotropical primates, with males attaining 55–78 cm of head-body length, 74–80 cm of tail length, and weighing around 8–11 kg. With long, thin arms and legs, and thumbless hands, marmosets sway in trees hanging only by their prehensile tails (Talebi and Lee, 2010). They are mainly folivorous, though they may also feed on flowers, bark, bamboo, ferns, nectar, pollen, and seeds (Talebi et al., 2005). Marmosets are diurnal primates living in groups of 4–43 individuals at altitudes ranging from sea level up to 2,000 m above sea level. Both marmoset species are currently critically endangered, primarily due to their habitat destruction (Talebi et al., 2019), and the *B. arachnoides* population is estimated at only 1,100–1,200 individuals (Strier et al., 2017).

Phylogenetic reconstructions based on molecular data consistently places the genus *Brachyteles* in the Atelidae family (Schneider et al., 1993, 2001; Perelman et al., 2011; Di Fiore et al., 2015; Dumas and Mazzoleni, 2017), within which its relative position to other genera is established as *{Alouatta [Ateles (Brachyteles, Lagothrix)]}*. Divergences from the remaining neotropical families Cebidae and Pitheciidae are dated between 20–23 mya and 22–26, respectively. While relationship between the three families was previously disputed, multiple works in the last decade have clarified this conflict as caused by incomplete lineage sorting events. Karyotypic analyses with chromosome banding shows a diploid number ($2n$) of 62 chromosomes and comparative painting with whole chromosome probes revealed several interspecific chromosome homologies between *Brachyteles*, the human, and other non-human primates, mainly *Lagothrix*, a taxon sharing a very similar karyotype (de Oliveira et al., 2005). Interestingly, *Brachyteles* and *Lagothrix* show the highest diploid chromosome count among neotropical primates ($2n = 62$).

In the last two decades the genetic basis underlying adaptive traits of New World monkeys (parvorder Platyrrhini) has been a subject of interest to several investigators (Ribeiro et al., 2005a; Vargas-Pinilla et al., 2015; Fam et al., 2019). There has been a significant concern about genes related to the immune system (Soares et al., 2010; Darc et al., 2012), which have shown higher

evolution rates in Platyrrhini than in other clades (Ribeiro et al., 2005b). The coevolution between pathogens and host defensive systems may, eventually, result in zoonotic transmissions from mammals to humans, thus heightening public health concern, as seen in the recent coronavirus pandemic outbreak (Olival et al., 2017; Kaján et al., 2020; Perlman, 2020).

High throughput sequencing technologies have flourished this past decade in the field of primate genomes, scanning thousands of genes (Nielsen et al., 2005; Kosiol et al., 2008; George et al., 2011). RNA-sequencing requires less sequencing effort than whole genome sequencing for acquiring desirable coverage and depth of genes, since less than 2% of the human genome codes for proteins (International Human Genome Sequencing Consortium, 2004). This makes *de novo* assembled transcriptome a powerful approach for identifying positively selected genes and testing evolutionary hypotheses, particularly when dealing with a non-model species for which a reference genome is not available (Stapley et al., 2010; Perry et al., 2012; Yang et al., 2014; Lan et al., 2018; Bentz et al., 2019).

Here we report for the first time the assembly, annotation and expression profile of *B. arachnoides* transcriptome from two different cell samples. Ortholog sequences from 10 other primates were retrieved and compared. Subsequently, a branch-site model analysis identified genes under positive selection in *B. arachnoides*, revealing positive selection at 132 amino acid sites in 30 genes, with several sites located in genes related to the immune response. We also showed an expansion on *APOBEC3G* increasing the knowledge of retrocopy gene birth to an Atelid species.

MATERIALS AND METHODS

Sample Collection and Ethical Considerations

Peripheral blood samples and a skin biopsy were collected from a captive male *B. arachnoides* (CPRJ2506) kept in the Centro de Primatologia do Rio de Janeiro (CPRJ-INEA; see Schrago et al., 2012), where blood samples were regularly collected for checkups and control of captive animals. Since this species is critically endangered, minor invasive methods (blood sample and skin biopsy) were adopted minimizing the risk for the specimen. Sample collection was carried out according to the IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Brazil; permanent license number 111375-1) national guidelines and provisions. This license was granted by IBAMA following the approval of its Ethics Committee.

Establishment of Cell Cultures From Skin Biopsy

A skin biopsy (0.5 cm × 0.5 cm), aseptically collected from the animal, was minced into small fragments of approximately 1 mm² and digested with 10 mL aliquots of a trypsin-collagenase solution prepared with 500 mL of RPMI medium, 2.5 g of collagenase, 50 mL of 2.5% trypsin, and 5 mL of each of the following antimicrobial agents: gentamicin, fungizone, penicillin-streptomycin and mycostatin, and HEPES buffer to a final concentration of 25 mM. Fragments were digested under continuous agitation using a magnetic stir plate. At 15 min intervals, supernatants were collected, mixed with Dulbecco MEM and 10% fetal calf serum, centrifuged at 500 g. Pellets were suspended in Dulbecco MEM medium with 10% fetal calf serum and incubated in flasks at 37°C and 5% CO₂. This procedure was repeated by adding fresh solution to remaining fragments until all were digested. Confluent flasks were harvested and cell pellets stored at –20°C for nucleic acid isolation.

RNA Isolation, Preparation of Transcriptomic Libraries and Sequencing

Total RNA was isolated from blood cells and fibroblast cell culture with RNeasy[®] (Qiagen). RNA was subsequently quantified with a NanoDrop spectrophotometer (Thermo Scientific). Quality and integrity were checked by electrophoresis in 1% agarose gels. Two aliquots (4 µg) of total RNA (one from blood and another one from fibroblasts) were used for preparing 300–400 bp cDNA libraries with TruSeq RNA Sample Prep Kit v2[®] (Illumina) following the protocol recommended by the manufacturer. Library size was checked with an Agilent Bioanalyzer 2100 DNA using a high sensitivity DNA chip. Library concentrations were estimated by qPCR using a Library Quantification kit (KAPA – KK4824). Two single-end runs from blood and three paired-end runs from fibroblasts were carried out in an Illumina HiSeq 2500 platform generating a total of 350 million reads.

Data Preprocessing, Transcriptome Assembly and Annotation

Adapter sequences were removed from raw reads with Cutadapt v1.14 (Martin, 2011). Low quality reads and poly-A tails were trimmed with Trimmomatic v0.38 (Bolger et al., 2014) with the following parameters: *leading:20 trailing:20 slidingwindow:4:20 minlen:36* and Prinseq v0.20.4 (Schmieder and Edwards, 2011) with the following parameters: *trim_tail_left 7 trim_tail_right 7 min_len 36*. Following this preprocessing step, data quality was assessed with FastQC v0.11.5. In order to improve assembly efficiency, the dataset was *in silico* normalized using a perl script built-in Trinity software following best practices for *de novo* assembly using Trinity. This normalization step reduced the quantity of input reads to be assembled, while maintaining transcriptome complexity and capability for full-length transcript reconstruction (Haas et al., 2013).

All five runs were combined and used for *de novo* assembly with Trinity v2.8.4 (Grabherr et al., 2011; Haas et al., 2013) in three different runs, altering *-min_kmer_cov*, which stands

for the minimal count for *K*-mers to be assembled by Inchworm (parameters: *-no_normalize_reads -run_as_paired -min_kmer_cov 2/3/5*). The three assemblies were evaluated with Detonate software v1.11 RSEM-EVAL built-in package (Li et al., 2014). This package combined multiple factors into a single score, including the compactness of assemblies and the read support. The assembly with the highest score was selected for subsequent analyses.

The completeness of the transcriptome was evaluated by a quantitative assessment of transcriptome assembly provided by BUSCO v3 (Simão et al., 2015). The number of full-length assembled transcripts was evaluated using BLASTX (SwissProt database for *Homo sapiens*, *e*-value < 1e-20 (Altschul et al., 1990) and examining the percentage of the target aligned to the best matching transcript with a perl script built-in Trinity software.

In order to decrease the number of fragmented transcripts, the transcriptome was filtered for assembled transcripts with at least 300 nt. Subsequently, Transdecoder v5.5 was used for identifying candidate coding regions (CDS) within assembled transcript sequences. In order to functionally annotate these potential proteins, all predicted CDS were searched in SwissProt (Euarchontoglires), eggNOG 5.0, KEGG, and Pfam public databases, using BLASTP (best top hit, *e*-value < 1e-10), eggNOG-mapper (min. hit *e*-value: 0.001, min. hit bit-score: 60, min.% of query cov.: 20) (Huerta-Cepas et al., 2017, 2019) and HMMER v3.1b2.

The transcript expression level of each cell sample was estimated by RSEM v1.2.28 (Li and Dewey, 2011) after aligning reads to the transcriptome with Bowtie2 v2.3.4.3 (Langmead and Salzberg, 2012). Transcripts with a Transcripts Per Million (TPM) ≥ 1 were used to compare the expression profiles of the two RNAs, and this comparison was visualized with Venn diagrams. Predicted CDS were annotated using the top hit of comparisons with the SwissProt dataset for Supraprimates (Euarchontoglires) using BLASTP (*e*-value < 1e-10). The 500 transcripts with the highest TPM estimates of each cell sample and with alignment coverage >80% of high scoring pair (HSP) length were used for pathway enrichment analysis (Over-Representation Analysis – ORA) with the Reactome database using the WebGestalt web tool (Wang et al., 2017). In order to get a better sense of the richness of the blood sample, we run the CIBERSORTx on a dataset containing TPM values of transcripts that had a significant BLAST hit against a Euarchontoglires protein and showed identity and HSP coverage over 60%. The CIBERSORTx was performed using a validated leukocyte gene signature matrix that contains 547 genes distinguishing 22 human hematopoietic cell phenotypes (Newman et al., 2019).

Adaptive Selection Analysis

Predicted coding sequences with alignment coverage of HSP length > 80% were further used for positive selection analysis and only isoforms with the highest TPM value were kept for downstream analysis. In order to represent the evolutionary dynamics within the neotropical primates lineage, we selected the best-annotated genomes available for this group for the selection tests; namely *Callithrix jacchus* (Ensembl release 96), *Saimiri boliviensis* (Ensembl release 96), and the outgroup *Homo sapiens*

(Ensembl release 96). Orthologs were then obtained by best reciprocal hit against predicted coding sequences. Alignments between the orthologous CDS identified were subsequently converted to amino acids, aligned using MAFFT v7.4.27 (Katoh and Standley, 2013), and converted back to nucleotides. Terminal stop codons were also removed from alignments. Reference topology for all subsequent analyses was generated by retaining only the taxa of interest in the primate phylogeny reconstructed by Perelman et al. (2011).

With the purpose of reducing the number of false-positives, we tested orthologs for positive selection using three independent validation processes. Firstly, the aBSREL model (Smith et al., 2015) available in the HyPhy 12.14 software package (Pond and Muse, 2005) was used for testing the four species for branch-site positive selection. Branches were considered under selection when the model of multiple ω rate classes along the sequence alignment was the best fit (Likelihood Ratio Test, adjusted p -value < 0.05). Findings were then confirmed by testing the predicted CDS for positive selection with the BUSTED model (Murrell et al., 2015) in HyPhy, which attributed three rate classes, one with $\omega \geq 1$ and two other with $\omega \leq 1$. The model fit was tested against the null hypothesis with a higher ω fixed to 1 (LRT, adjusted p -value < 0.05). Lastly, CDS showing positive selection in both tests were thereafter analyzed with a higher number of taxonomic units (up to 11 species depending on the availability of ortholog sequence in datasets). These included the primate species *Aotus nancymaae*, *Cebus capucinus*, *Chlorocebus sabaeus*, *Macaca nemestrina*, *Otolemur garnettii*, *Papio anubis*, and *Pongo abelii*. Because alignment errors highly increase the false-positive rate of the branch-site test (Fletcher and Yang, 2010), we manually verified these sequence alignments, discarding misaligned regions and subsequently used the branch-site test of CODEML built-in the PAML 4 package (Yang, 2007).

The genes identified as potentially evolving under positive selection were associated with Gene Ontology (GO) terms, Reactome and Panther pathways¹ (accessed: 14 October 2019). We also identified pervasive selection by comparing the synonymous and non-synonymous rates using the fixed effects likelihood test (FEL) (Kosakovsky Pond and Frost, 2005). In this last analysis, we only used the alignments with 11 species that returned positive to all selection tests. Finally, we reconstructed the evolutionary trees of the final positively selected CDS using IQTREE 1.7 software (Nguyen et al., 2015) with 1000 bootstrap replicates. Evolutionary models were selected employing the Model Finder algorithm available with IQTREE. Trees were rooted on the outgroup species *Otolemur garnettii* or, when orthologous sequence for this species was not available, on the diversification between New World and Old World primates.

Comparative Genomic Analysis and Orthogroup Identification

For the comparative analysis, we downloaded the proteomes of the same 10 primate species used in adaptive selection analysis (Ensembl release 100). To increase accuracy of the orthology assignment, we used a script provided with OrthoFinder to

extract just the longest transcript variant per gene. In the same way, to avoid misassembled transcripts and reduce redundancy, we used only peptide sequences predicted from transcripts with TPM > 1 and we ran CD-HIT (Fu et al., 2012) to cluster similar sequences with a threshold of 95%. The identification of orthogroups and the comparative analysis were performed by OrthoFinder v2 (Emms and Kelly, 2019). Briefly, OrthoFinder analysis consists of an all-to-all BLASTP step and clustering step using MCL algorithm to build the orthogroups.

The sequences of *APOBEC3G* identified by the OrthoFinder were aligned using Clustal Omega (Thompson et al., 1994) and the maximum likelihood tree was constructed using PHYML (Guindon et al., 2010) with NNIs as tree topology search, BioNJ initial tree, Blosum62 + GAMMA amino acid substitution model and 100 bootstraps. Tree was rooted using human and NWM *APOBEC3A* sequences.

Molecular Modeling

In order to understand whether the positive selection on gene sequences affects their structure, we build and compare 3D models for Human and *B. arachnoides* of *APOBEC3G* and *CEACAM1* proteins using homology modeling approach. BLASTP algorithm (Altschul et al., 1990), with default settings, was used to obtain the two template structures from protein data bank. The structure of *APOBEC3G* of *Macaca mulatta* under PDB ID 6P3X (Yang H. et al., 2020) was used as a template for modeling the *APOBEC* sequences of the *H. sapiens* and *B. arachnoides*, and the structure *CEA* (expressed by *CEACAM5* gene) from *H. sapiens* under PDB ID 1E07 (Boehm and Perkins, 2000) was used to build the *CEACAM1* proteins for *H. sapiens* and *B. arachnoides*.

All models were built using the software Modeller v9.24 (Webb and Sali, 2016) and the primary structures of templates and models were aligned using the web server Clustal Omega (Madeira et al., 2019). One thousand models were generated for each model, giving a total of four thousand structures, the models with the best objective function value were chosen. For *APOBEC3G*, the best models were 359 and 823 for *H. sapiens* and *B. arachnoides*, respectively. And the best models for *CEACAM1* were 326 and 158 for *H. sapiens* and *B. arachnoides*, respectively. The electrostatic profile of the protein models were calculated by APBS and PDB2PQR (Baker et al., 2001; Dolinsky et al., 2007; Unni et al., 2011) and rendered on Pymol v2.3 (DeLano, 2002).

RESULTS

De novo Transcriptome Assembly and Quality Assessment

More than 350 million reads of blood and fibroblast transcripts of *B. arachnoides* were generated by the Illumina HiSeq 2500 platform (**Supplementary Material 1**). Following adapter removal and low-quality base trimming, 315,430,518 (89.26%) reads, ranging from 36 to 124 nucleotide in length and with a mean Phred score > 30 , were used for assembling and comparing the three assemblies. The assembly with a minimum k -mer count of 2 performed slightly better than the others, presenting

¹<http://geneontology.org/>

a higher number of transcripts (284,283), a re-mapping ratio of 93.5%, a percentage of complete BUSCOs (72.3%), and N50 length (2,940 bp). We found 9,826 proteins (SwissProt *Homo sapiens*) represented by nearly full-length transcripts, with >80% alignment coverage (Table 1 and Supplementary Material 2), with a better RSEM-EVAL score than that of the other two assemblies (RSEM-EVAL score = -9,831,876,879). We predicted the likely protein coding sequences from these transcripts and annotated them with BLASTP. We also predicted 81,400 coding sequences from 185,831 transcripts with ≥ 300 nt. From these sequences, 79.8% (64,929) showed a significant (e -value $< 1e-10$) BLAST hit against an Euarchontoglires protein SwissProt dataset; the vast majority of hits being with human proteins, the most complete and curated protein dataset. As for full-length or nearly full-length transcripts, 24,016 of them cover at least 80% of the SwissProt proteins (HSP coverage).

Expression Profiles and Transcriptome Annotation

Transcript expression profiles were estimated by comparing annotations (SwissProt database) between the two cell samples, considering only transcripts with TPM estimates ≥ 1 , accounting for 45,276, with 38.4% of them expressed in both samples (Figure 1A). Our transcriptome assembling strategy was highly geared to assemble a high rate of complete or almost-complete transcripts, with a median HSP coverage of 88.2% (Figure 1C), including low expressed transcripts, since 50% of all transcripts were expressed below 10 TPM. On the other hand, transcripts exclusively expressed in one cell sample showed lower HSP coverage (median = 57.4% and 60.1% in blood and fibroblasts, respectively) and expression level estimates (2.2 and 2.3 median TPM, in blood and fibroblasts, respectively) (Figure 1C). When considering only transcripts with homology to innate immune system proteins (Figures 1B,D), a similar pattern was observed though with a higher median TPM and a higher number of transcripts for innate immunity genes in blood cells, as expected, depicting its complexity in comparison to the fibroblast transcriptome.

Functional enrichment analysis using the Reactome database was subsequently performed to identify pathways that were significantly over-represented in each sample. Enrichments

were mainly found in pathways related to the immune system (enrichment ratio 3.9, FDR = 0) and hemostasis in blood cells (enrichment ratio 3.0, FDR = 5.7E-07) (Table 2) and those associated with collagen and extracellular matrix metabolism (enrichment ratio 5.3, FDR = 4.7E-12) in fibroblasts (Table 3). Using the deconvolution method with CIBERSORTx we found 13 out of 22 cell signatures in the blood sample (Supplementary Figure 1).

Classification of Known and Novel Protein-Coding Genes

In addition to the functional annotation of the CDS carried out with the Supraprimates SwissProt database, searches were performed in the eggNOG and Pfam databases (Supplementary Material 3). From the 64,929 CDS with a significant BLAST hit against a Euarchontoglires protein, 34,084 sequences showed identity and HSP coverage over 60%, and these were considered homologs (orthologs or recent paralogs) to Supraprimates proteins. Of the remaining sequences (30,845), 94% (29,056) were associated with a protein domain or a KEGG Orthology (KO) group (Supplementary Material 4) and were indicated as potential novel or specific protein-coding genes of *B. arachnoides*. In addition, 4,537 sequences without BLAST hits, showed hits with a protein domain or KO, increasing the number of potential novel isoforms or genes to 33,593. The top three COG categories of these sequences were 'signal transduction mechanisms,' 'function unknown,' and 'transcription' (Supplementary Figure 2). The vast majority of the predicted peptides (11,274 from 11,934) which were not associated with any element of the databases searched herein were shorter than 200 amino acids, either too fragmented for proper identification or spuriously predicted (Supplementary Material 4).

Identification of Gene Duplication and Species-Specific Orthogroups

From the total 81,400 ORF's predicted from the transcriptome sequences, 58,138 have TMP estimates ≥ 1 , and using CD-HIT with a threshold of 95% of similarity, we reduce the sequence redundancy to 37,987 predicted protein sequences. Adding this to the proteomes of the ten other selected primates,

TABLE 1 | Summary of *Brachyteles arachnoides* transcriptome assembly, CDS prediction, annotation, and completeness according to BUSCO.

	All transcripts	Transcripts ≥ 300 nt	Coding sequences (CDS)
No of Transcripts	284,283	185,831	81,400
No of BLAST Hits	94,853	83,864	64,929
Full Length ($\geq 80\%$)			24,016
BUSCO Statistic			
Complete BUSCOs	4,479 (72.3%)	4,479 (72.3%)	4,335 (70.0%)
Complete – single-copy BUSCOs	1,450 (23.4%)	1,450 (23.4%)	1,785 (28.8%)
Complete – duplicated BUSCOs	3,029 (48.9%)	3,029 (48.9%)	2,550 (41.2%)
Fragmented BUSCOs	465 (7.5%)	450 (7.3%)	504 (8.1%)
Missing BUSCOs	1,248 (20.2%)	1,263 (20.4%)	1,353 (21.9%)

Coding sequences were predicted by Transdecoder, with full length annotation performed using BLAST.

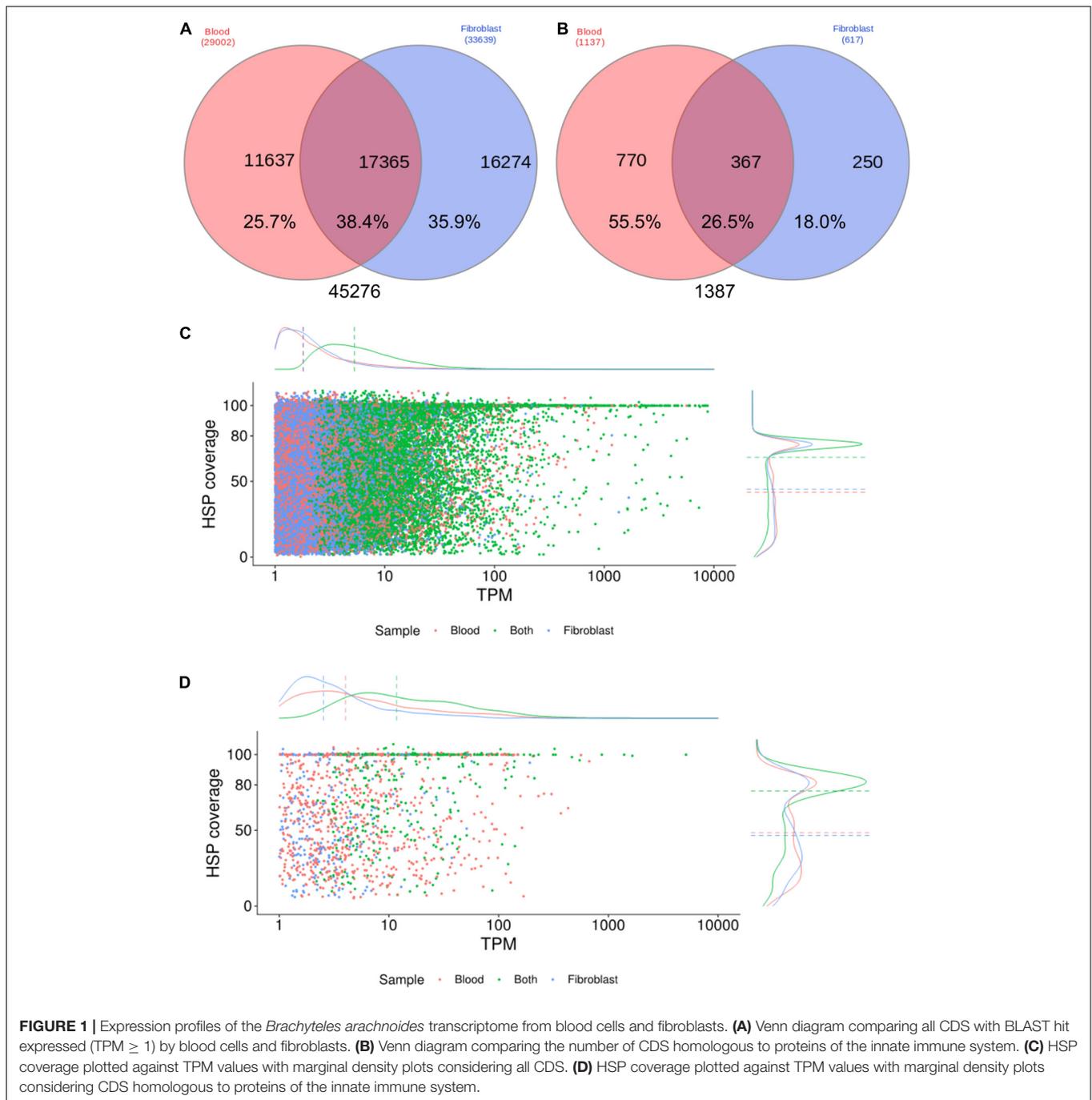


FIGURE 1 | Expression profiles of the *Brachyteles arachnoides* transcriptome from blood cells and fibroblasts. **(A)** Venn diagram comparing all CDS with BLAST hit expressed (TPM ≥ 1) by blood cells and fibroblasts. **(B)** Venn diagram comparing the number of CDS homologous to proteins of the innate immune system. **(C)** HSP coverage plotted against TPM values with marginal density plots considering all CDS. **(D)** HSP coverage plotted against TPM values with marginal density plots considering CDS homologous to proteins of the innate immune system.

we end up with 245,988 sequences for orthogroups inference. The comparative analysis of the 11 primate species assigned 92.9% of the sequences in orthogroups (**Supplementary Material 5**). The southern muriqui showed the lowest percentage of genes assigned to an orthogroup (**Figure 2**), this might be a result of spurious ORF's prediction from non-coding RNAs present in the transcriptome. However, the absolute number of assigned genes of *B. arachnoides* is comparable to the other species (**Supplementary Material 5**). Of 21,369 inferred orthogroups, 10,534 (49.3%) have all species present (**Figure 2**), further

validating the completeness and quality of the transcriptome presented here. This analysis was able to identify 4,101 single-copy orthogroups representing a core set of conserved ortholog genes among primates. In addition to these conserved genes, 1,230 orthogroups were species-specific to *B. arachnoides*, due to in-paralogs grouping. Although we know that these in-paralogous sequences are the result of duplications after speciation, this high number might be due to the use of a transcriptome instead of a reference genome, which is absent for this species.

TABLE 2 | Enrichment analysis of the 500 transcripts with the highest TPM estimates in blood cells.

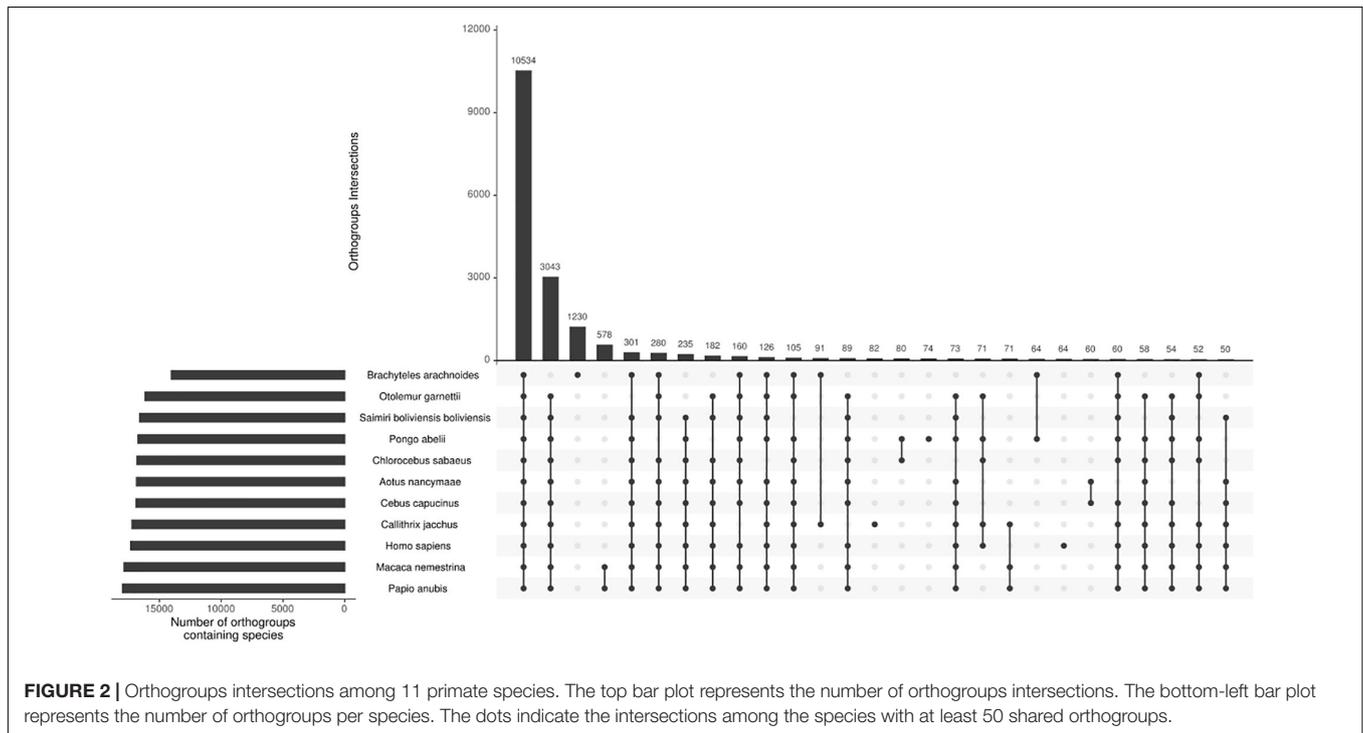
Gene set	Description	Size of gene set	Overlap	Expect	Enrichment ratio	Adj. p-value
R-HSA-168249	Innate immune system	535	88	22.6	3.9	0
R-HSA-109582	Hemostasis	279	35	11.8	3.0	5.7E-07
R-HSA-372790	Signaling by GPCR	252	33	10.7	3.1	5.7E-07
R-HSA-375276	Peptide ligand-binding receptors	42	10	1.8	5.6	5.6E-04
R-HSA-202403	TCR signaling	69	12	2.9	4.1	1.8E-03
R-HSA-76005	Response to elevated platelet cytosolic Ca ²⁺	61	11	2.6	4.3	2.4E-03
R-HSA-202430	Translocation of ZAP-70 to Immunological synapse	6	4	0.3	15.8	2.5E-03
R-HSA-2029480	Fcγ receptor (FCGR) dependent phagocytosis	57	10	2.4	4.1	5.4E-03
R-HSA-2871809	FCER1 mediated Ca ²⁺ mobilization	20	6	0.8	7.1	5.8E-03
R-HSA-449147	Signaling by Interleukins	237	23	10.0	2.3	5.9E-03

Reactome database was used for enrichment.

TABLE 3 | Enrichment analysis of the 500 transcripts with the highest TPM estimates in fibroblasts.

Gene set	Description	Size of gene set	Overlap	Expect	Enrichment ratio	Adj. p-value
R-HSA-1474244	Extracellular matrix organization	128	31	5.8	5.3	4.7E-12
R-HSA-8948216	Collagen chain trimerization	14	9	0.6	14.1	3.2E-07
R-HSA-8951671	RUNX3 regulates YAP1-mediated transcription	7	5	0.3	15.7	3.8E-04
R-HSA-381426	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	52	11	2.4	4.6	1.5E-03
R-HSA-8874081	MET activates PTK2 signaling	16	6	0.7	8.2	3.9E-03
R-HSA-446353	Cell-extracellular matrix interactions	11	5	0.5	10.0	5.5E-03

Reactome database was used for enrichment.



The OrthoFinder also performs the identification of gene duplication events across the studied lineages analyzing the gene trees of each orthogroup and comparing with the species tree.

This analysis inferred 3,877 orthogroups with a duplication event on the terminal branch of southern muriqui, of which 3,642 shared orthologs with human proteome. This data constitutes

a rich and useful resource that can be further used for a variety of evolutionary genomics projects aiming recent gene family expansions. This analysis also detected well-supported 45 duplication events in the common ancestor of NWM. Among these events we identified an expansion of *APOBEC3G* in Platyrrhini clade, including three copies in *A. nancymaae* (Nancy Ma's night monkey), nine copies in *C. jacchus* (common marmoset), four copies in *C. capucinus* (White-faced capuchin), five copies in *S. boliviensis boliviensis* (Bolivian squirrel monkey) and five copies in *B. arachnoides*. On the other hand just one ortholog was identified in the Old World monkeys, Great Apes, and Human genomes (Figure 3). In fact, Trinity has assembled 12 transcripts coding for *APOBEC3G*, of which 9 presented TPM estimates greater than 1 and these were clustered in five representative sequences that seem to have originated from different genes. A maximum likelihood phylogenetic tree using the aligned *APOBEC3G* amino acid sequences revealed a great cluster of NWM sequences separated from Catarrhini sequences (Figure 3). However, the tree reconstruction was not able to recover the relation among NWM *APOBEC3G* sequences, since the canonical genes of *A. nancymaae* and *C. jacchus* appeared in different branches.

Adaptive Selection Analysis

Identification of genes under adaptive selection in the *B. arachnoides* lineage was performed with an initial set of 4,533 transcripts successfully identified as having homologous sequences in *C. jacchus*, *S. boliviensis* and *H. sapiens*. This analysis identified 30 genes in *B. arachnoides* (adjusted p -value < 0.05) which, according to GO, were associated with the biological processes and molecular functions listed in Supplementary Material 3. These positively selected genes belong to a wide variety of categories, such as the innate immune system (*APOBEC3G*, *OAS2*, and *CEACAM1*), the immune response (*CLEC17A*, *ITGAM*, *SLC15A2*, *ECM1*, *HLA-DRB1*, and *GZMB*), ion transport (*SLC39A13*), rRNA processing and translation (*BMS1*, *DDX52*, and *MRPS14*), and DNA binding proteins (*TAF9B*, *ZSWIM8*, and *ZBTB44*). The protein alignments with domain information and the reconstructed gene trees of each positively selected CDS using the 11 different primate species are displayed in the Supplementary Figures 3, 4, respectively. Main lineages presented a monophyletic grouping in most trees, with exceptions being caused in each case by the anomalous position of a single species.

In order to identify sites under positive selection, FEL was used for estimating rates of non-synonymous (dN) and synonymous (dS) nucleotide substitutions (ω) at each site in a sequence alignment comprising 11 primate species. Among the 30 genes under adaptive selection in *B. arachnoides* (Supplementary Material 6), significant estimates of positive selection with $\omega > 1.0$ (LRT, adj- $p < 0.05$) were found at 132 sites (Supplementary Material 7). Two genes related to the innate immune response (*OAS2* and *CEACAM1*) showed the highest number of sites under positive selection, with 13 and 10 sites with $\omega > 1.0$, respectively. Most *OAS2* sites were located in *OAS* domain 1, while two *CEACAM1* sites occurred at a signal peptide region, three at the immunoglobulin variable-region-like

(IgV-like) domain, three other distributed along two of the three immunoglobulin constant-region type-2-like (IgC2-like) domains and the last two at the initial portion of the cytoplasmic domain (Figure 4). A third gene related to the innate immune response (*APOBEC3G*), showed seven sites inside the cytidine and deoxycytidylate deaminase domain. Interestingly, 13 of the 27 remaining genes under adaptive selection encoded membrane or extracellular matrix proteins, nine of which at the plasma membrane (Supplementary Material 6).

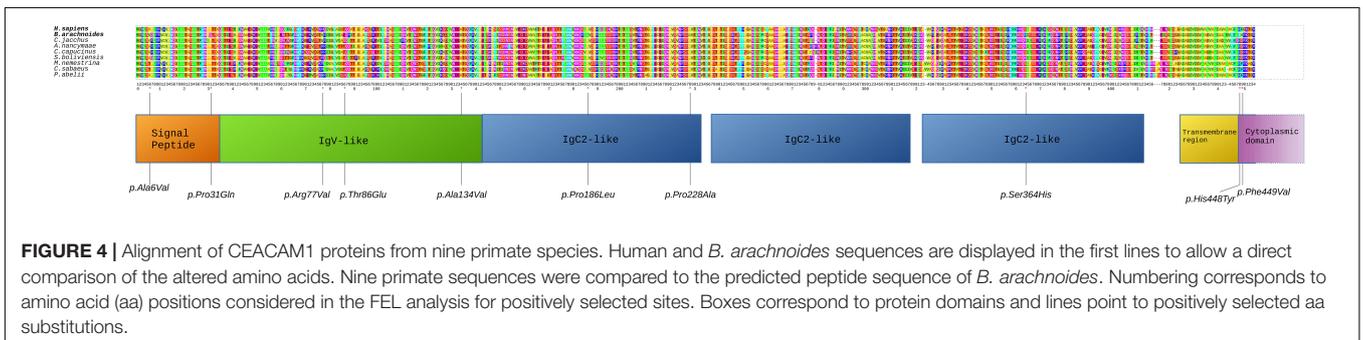
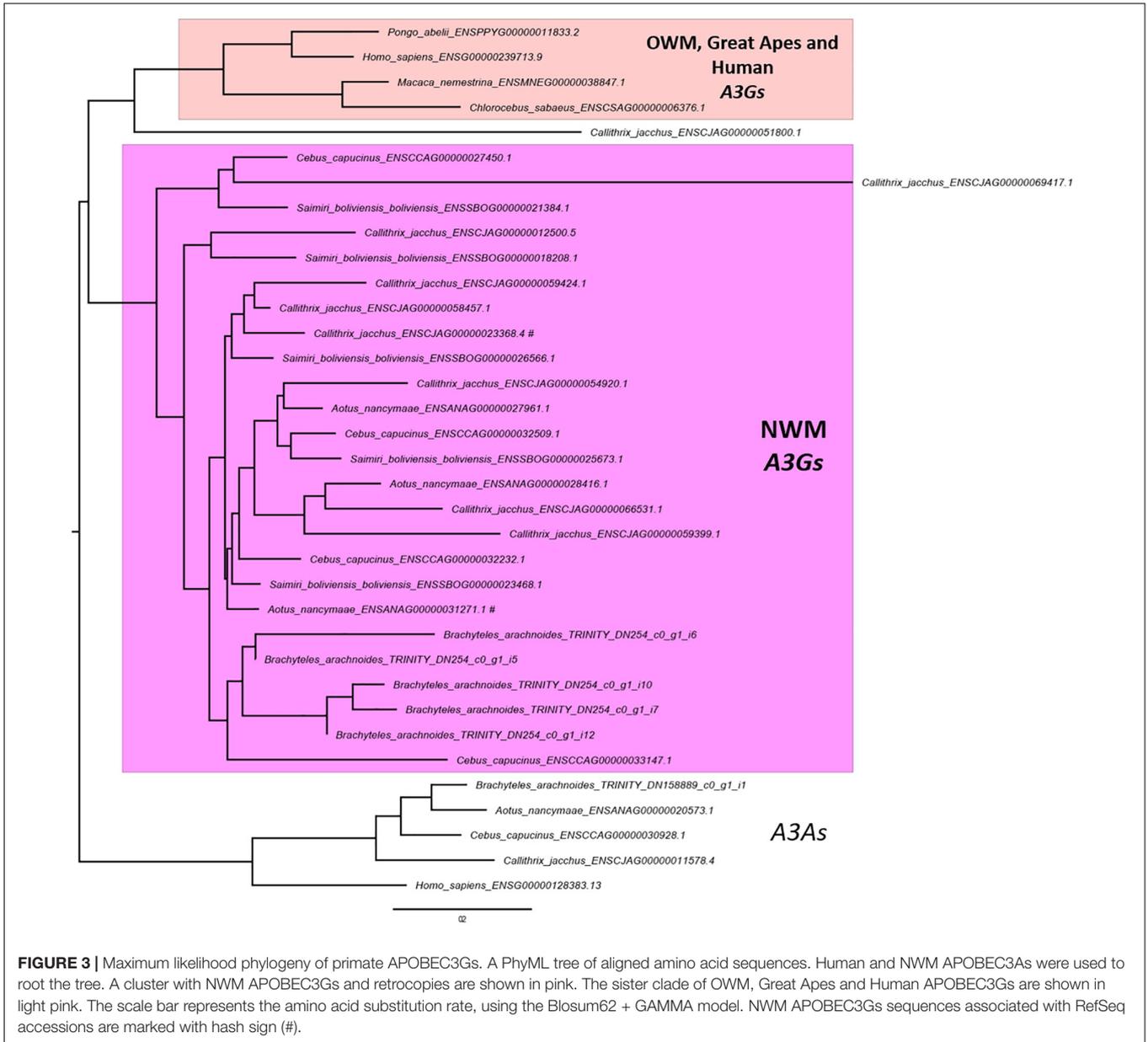
Molecular Modeling of APOBEC3G and CEACAM1

In order to evaluate the possible impact of sites under positive selection on *APOBEC3G* and *CEACAM1*, we used a 3D modeling approach to assess their structures. The *APOBEC3G* structures of *H. sapiens*, shown as a ribbon diagram (Figure 5D), and *B. arachnoides* were superposed. The analysis presented a RMSD value of 0.469 (Figure 5C), and both proteins have a similar conformation due to the modeling by satisfaction of spatial restraints (Supplementary Figure 6). However, when we compare the electrostatic potential profile of both models, we can observe a more electrophilic region in the center of human monomers of *APOBEC3G* (Figure 5B) in comparison to the southern muriqui protein, which has a more neutral profile in the same region (Figure 5A). The superposition of *CEACAM1* models of *H. sapiens* and *B. arachnoides* presented a higher RMSD value (3.426) (Supplementary Figure 5) in comparison to *APOBEC3G* models, and this high value is a consequence of a coil region, located on aminoacid residues ranging from 421 to 455 (Supplementary Figure 5). This same region in the *H. sapiens* model presents a small beta sheet structure, probably a consequence of residues His449 and Phe450. Overall, both *CEACAM1* models have only beta sheet folding and a similar structure (Supplementary Figure 6). As it occurred with *APOBEC3G*, the electrostatic potential profile of *CEACAM1* shows some differences in the comparison of the two proteins (Supplementary Figure 5), mainly on residues 77 and 346.

DISCUSSION

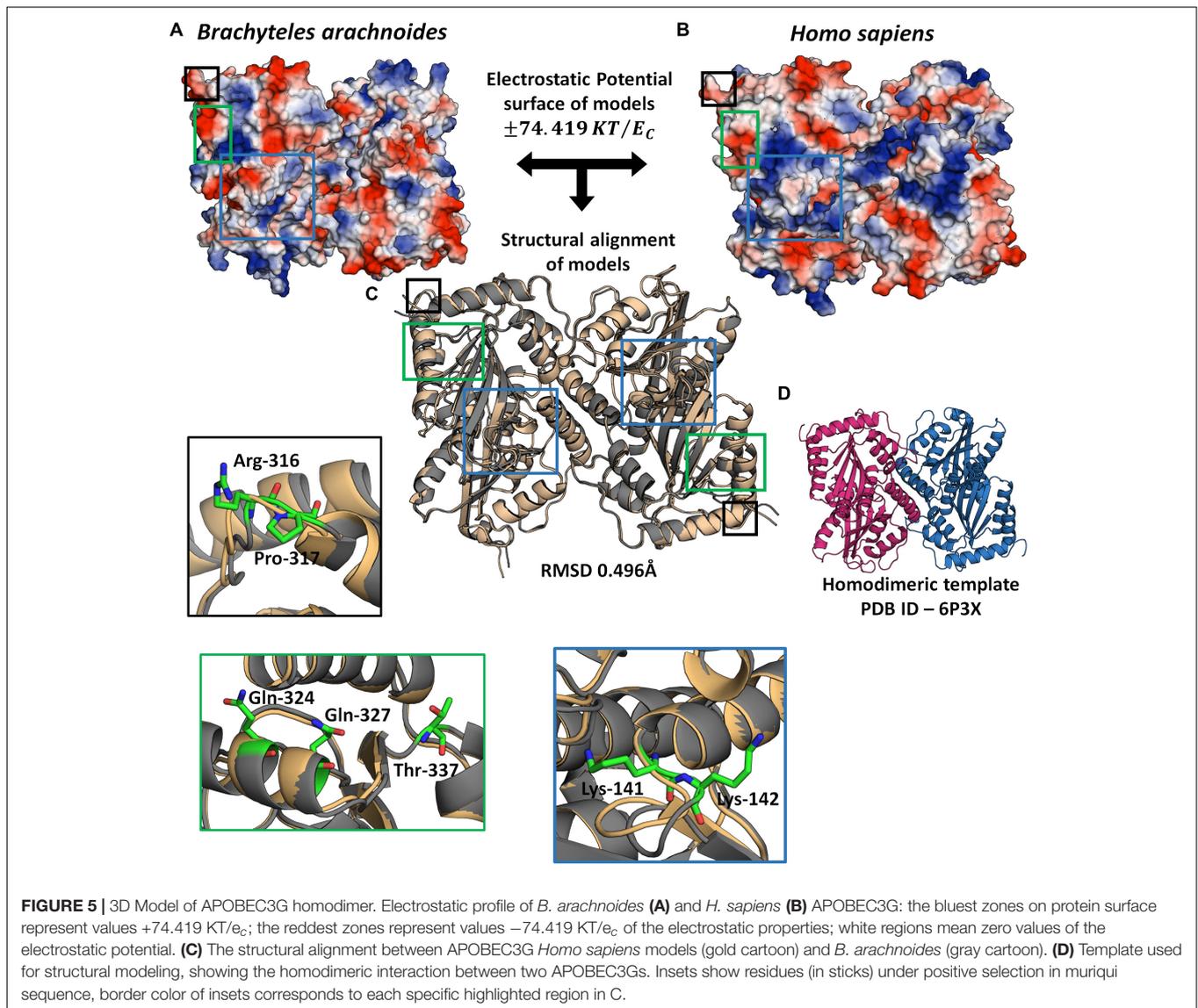
We present herein, for the first time, a high-quality transcriptome assembly report on the critically endangered New World monkey, *B. arachnoides*, providing valuable genetic data of a species whose reference genome is still unavailable, as well as the first transcriptome report of a representative member of the family Atelidae. This work also provides the first transcriptome analysis of both blood cells and fibroblasts of any New World monkey species, assembling and annotating more than 24,000 full-length protein-coding sequences from these cell samples. It is also relevant to highlight the richness and completeness of this dataset, as the blood sample presented here showed 13 out of 22 cell signatures and the comparative analysis showed a high number of inferred orthogroups across all primate lineages studied. The Human Protein Atlas² (June 8th, 2020) examined

²<https://www.proteinatlas.org/humanproteome/tissue>



19,670 protein coding genes, and presents their expression in many different tissues. In the case of blood, a total of 16,063 genes, corresponding to 81.7% of the total, are detected by RNA-seq.

This number agrees with previous publications indicating that blood transcriptome presents a rich source of information in itself (Newman et al., 2015).



A comparison with other *de novo* assemblies of New World monkeys showed that the N50 length of *B. arachnoides* contigs was similar to that in *C. jacchus* assemblies (2,441–2,935 bp) (Maudhoo et al., 2014) but higher than that in *Cebus apella* (1,809 bp) (Boddy et al., 2017). Comparison of transcriptional profiles between cell samples showed a higher number of coding sequences expressed exclusively in fibroblasts than in blood cells. Fibroblast genes were enriched in fewer pathways than blood genes, pointing to the higher transcript complexity resulting from the different cell types present in blood.

Sequencing and *de novo* assembly of *B. arachnoides* transcriptome provided information on several genes despite the lack of a reference genome which was useful for testing evolutionary hypotheses and selection. Analyses of 4,528 sets of *B. arachnoides* orthologs using a branch-site model revealed evidence of positive selection in 30 genes, at 132 amino acid sites, several of them within host defensive genes. This probably

reflects the selective pressures imposed by pathogen infections on proteins encoded by immune system genes which have been targets of positive selection and rapid evolutionary adaptation (Barreiro and Quintana-Murci, 2010; Daugherty and Malik, 2012; Akkaya and Barclay, 2013; Rausell and Telenti, 2014).

In this study, the *OAS2* gene presented the highest number of sites under positive selection. It encodes a 2'-5'-oligoadenylate synthase 2 whose homodimer may bind to viral double-stranded RNA and catalyze synthesis of 2'-5' oligoadenylates, leading to activation of RNase L and RNA degradation (Williams and Kerr, 1978; Sarkar et al., 1999). In humans, the *OAS* family comprises four paralogs (*OAS1*, *OAS2*, *OAS3*, and *OASL*) distinguished by the number of *OAS* units, which is the number of NTase and *OAS1*-C domains they contain as a result of genomic tandem duplications (*OAS1*- one unit, *OAS2*- two units, and *OAS3*- three units) (Hancks et al., 2015; Mozzi et al., 2015). These authors postulated that these domain duplications and fusions might

be adaptive, making *OAS2* and *OAS3* encoded proteins more resistant to pathogen-encoded inhibitors than the one encoded by *OAS1*, making *OAS1* a more frequent target of positive selection than *OAS2*. In this study, however, analysis of the almost complete, assembled sequence of the *OAS1* ortholog in *B. arachnoides* did not show evidence of positive selection.

In humans, several studies have shown associations of single nucleotide polymorphisms located within the *OAS* genes and increased risk of infection by viruses of the genus *Flavivirus*, including West Nile virus, dengue virus and tick-borne encephalitis virus (Lim et al., 2009; Barkhash et al., 2010; Thamizhmani and Vijayachari, 2014). Another relevant zoonotic disease caused by a *Flavivirus* is the yellow fever. Monath (2012) reviewing the risks and benefits of taking the live, attenuated yellow fever 17D vaccine, reported associations of a vaccine-associated viscerotropic disease and polymorphisms in innate immune genes, including *OAS1* and *OAS2* genes. From 2016 to 2018, Brazil has experienced the worst outbreak of yellow fever in the last 80 years, with high numbers of humans and non-human primates deaths (Possas et al., 2018). Although *Alouatta* is reported to be the most susceptible species to yellow fever virus, as reviewed by Strier et al. (2019) and Silva et al. (2020) reported reductions of 10–26% of two populations of *B. hypoxanthus* in areas with yellow fever outbreaks, both being Atelid species. Our results on positive selection sites identified in *OAS2* suggest that variations within *OAS* genes can be related to a high susceptibility to *Flavivirus*, as seen in humans, and this transcriptome is a valuable resource for future studies aiming to unveil this association.

CEACAM1, a gene belonging to the carcinoembryonic antigen (CEA) gene family, was also found to be under positive selection. It encodes the carcinoembryonic antigen-related cell adhesion molecule 1, a glycoprotein involved in cell adhesion, neovascularization, insulin homeostasis, and T-cell regulation, and plays an important role as a target for pathogen interaction and internalization (Hauck et al., 2006; Kuespert et al., 2006). *CEACAM1* has rapidly evolved in vertebrates, including humans, in a host, species-specific manner (Kammerer et al., 2004, 2007; Kammerer and Zimmermann, 2010; Voges et al., 2010; Zimmermann and Kammerer, 2016). One of the protein domains that was identified as harboring sites under positive selection was the immunoglobulin variable-region-like (IgV-like) domain, which was reported to be responsible for functional differences regarding bacterial binding (Voges et al., 2010), correlating with the restricted pathogenicity of the *CEACAM1*-targeting microorganisms.

In addition to our findings on adaptive selection in genes related to the immune response, we identified the *APOBEC3G* gene to be potentially under positive selection. The protein encoded by this gene belongs to the *APOBEC3* family of cytidine deaminases which binds to single-stranded DNA and converts cytosine to uracil (Navaratnam and Sarwar, 2006). This protein can inhibit retroviral infection and endogenous retrotransposons, such as *Alu* elements (Hulme et al., 2007; Mbisa et al., 2010; Hultquist et al., 2011) and is classified as a host restriction factor, an important protective component of the innate immune system (Chiu and Greene, 2008). *APOBEC3G* has been subject to

strong positive selection throughout primate evolution, possibly reflecting a long-term evolutionary conflict between retroviruses and their primate hosts (Sawyer et al., 2004; Bulliard et al., 2009; Krupp et al., 2013). In the comparative analysis, we also found several copies of this gene, which is in agreement with a recent published paper that found that NWM genomes, specifically Cebidae family, have accumulated *APOBEC3G* copies derived from retrotransposition, moreover these retrocopies can be transcribed and retain functional domains (Yang L. et al., 2020). Our findings show for the first time an Atelid species with *APOBEC3G* expansion and extend the knowledge that this might have occurred in the common ancestor of NWM.

Though, the seven sites herein identified to be under positive selection in *APOBEC3G* are present inside the cytidine and deoxycytidylate deaminase domain (InterPro ID: IPR002125), they are located downstream of two zinc-coordinating motifs (Navarro et al., 2005) and outside the intermonomeric contact region identified by Bulliard et al. (2009) and these sites did not affect the structure of the molecular model. However, the differences in the electrostatic profile might have a functional impact, though it is speculative. A recent work showed that mutating the residues in the dimeric interface was not sufficient to cancel the HIV-1 restriction (Yang H. et al., 2020). In fact, cloned NWM *APOBEC3G* retrocopies restrict HIV-1 but not LINE-1 (Yang L. et al., 2020), showing that differences in the sequence as consequence of accelerated evolution do not affect the antiviral activity and the positive selection on *APOBEC3G* gene might have occurred through a viral-driven evolutionary pressure. Nevertheless, more functional studies are needed for defining the nature of these underlying evolutionary forces.

Interestingly, several genes under positive selection encoded membrane or extracellular matrix proteins, some of which localized in the plasma membrane. Scarce available information on their evolution and divergence notwithstanding, they were associated with immunological functions, leading us to suggest that they might be under similar evolutionary pressures to the above mentioned genes. Such was the case of C-type lectin domain family 17, member A (*CLEC17A*), also known as Prolactin, a glycan-binding receptor that may be involved in innate immunity as a pattern recognition receptor and whose expression was reported in association with proliferating B cells in germinal centers (Graham et al., 2009; Angulo et al., 2019). Integrin alpha-M (*ITGAM*) was identified as the major cell receptor of neutrophils in the recognition of pathogenic fungi that may evade the complement system (Losse et al., 2010). Solute carrier family 15, member 2 (*SLC15A2*) is responsible for the intake of oligopeptides of 2–4 amino acids, and was implicated in the uptake optimization of bacterial peptides into the cytosol of macrophages, enhancing the production of proinflammatory cytokines (Hu et al., 2018). Extracellular matrix protein 1 (*ECM1*) is a soluble protein described in the promotion of follicular helper cell differentiation and antibody production (He et al., 2018). Our findings corroborate those of Nielsen et al. (2005) who reported a high number of immune-related genes targeted by positive selection in the genomes of humans and chimpanzees as a consequence of strong pathogen-driven evolutionary pressure. Future functional studies of these genes and their encoded

proteins may help to elucidate their physiological implications in *B. arachnoides*.

CONCLUSION

Transcriptome assembly of the southern muriqui (*B. arachnoides*), a critically endangered New World monkey, has proven to be a valuable genetic resource of a species whose reference genome is presently unavailable. It also provides the first transcriptome account of an Atelid primate. The high throughput approach for analyzing transcriptome data identified 30 candidate genes under positive selection with strong statistical evidence, and an important gene expansion event, shedding light on the evolutionary pressures modeling the primate genome, mainly on genes involved in the immune response. Furthermore, we presented a comparative analysis with other primates genomes, extending the knowledge of NWM genomic evolution. This work generated an abundant catalog of complete sequences that may be used to further elucidate several other aspects of genome evolution in primates as well as unveil the evolutionary diversification genes involved in the immune response. Further functional studies are required to provide an understanding of the nature of these underlying evolutionary processes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, SRR10907267-SRR10907271 and <https://www.ncbi.nlm.nih.gov/>, GJIV00000000.

REFERENCES

- Akkaya, M., and Barclay, A. N. (2013). How do pathogens drive the evolution of paired receptors? *Eur. J. Immunol.* 43, 303–313. doi: 10.1002/eji.201242896
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Angulo, C., Sanchez, V., Delgado, K., and Reyes-Becerril, M. (2019). C-type lectin 17A and macrophage-expressed receptor genes are magnified by fungal β -glucan after *Vibrio parahaemolyticus* infection in *Totoaba macdonaldi* cells. *Immunobiology* 224, 102–109. doi: 10.1016/j.imbio.2018.10.003
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041. doi: 10.1073/pnas.181342398
- Barkhash, A. V., Perelygin, A. A., Babenko, V. N., Myasnikova, N. G., Pilipenko, P. I., Romaschenko, A. G., et al. (2010). Variability in the 2'-5'-oligoadenylate synthetase gene cluster is associated with human predisposition to tick-borne encephalitis virus-induced disease. *J. Infect. Dis.* 202, 1813–1818. doi: 10.1086/657418
- Barreiro, L. B., and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* 11, 17–30. doi: 10.1038/nrg2698
- Bentz, A. B., Thomas, G. W. C., Rusch, D. B., and Rosvall, K. A. (2019). Tissue-specific expression profiles and positive selection analysis in the tree swallow (*Tachycineta bicolor*) using a de novo transcriptome assembly. *Sci. Rep.* 9:15849. doi: 10.1038/s41598-019-52312-4
- Boddy, A. M., Harrison, P. W., Montgomery, S. H., Caravas, J. A., Raghanti, M. A., Phillips, K. A., et al. (2017). Evidence of a conserved molecular response to selection for increased brain size in primates. *Genome Biol. Evol.* 9, 700–713. doi: 10.1093/gbe/evx028
- Boehm, M. K., and Perkins, S. J. (2000). Structural models for carcinoembryonic antigen and its complex with the single-chain Fv antibody molecule MFE23. *FEBS Lett.* 475, 11–16. doi: 10.1016/S0014-5793(00)01612-4
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bulliard, Y., Turelli, P., Röhrig, U. F., Zoete, V., Mangeat, B., Michielin, O., et al. (2009). Functional analysis and structural modeling of human APOBEC3G reveal the role of evolutionarily conserved elements in the inhibition of human immunodeficiency virus type 1 infection and Alu transposition. *J. Virol.* 83, 12611–12621. doi: 10.1128/JVI.01491-09
- Chaves, P. B., Magnus, T., Jerusalinsky, L., Talebi, M., Strier, K. B., Breves, P., et al. (2019). Phylogeographic evidence for two species of muriqui (genus *Brachyteles*). *Am. J. Primatol.* 81:e23066. doi: 10.1002/ajp.23066
- Chiu, Y.-L., and Greene, W. C. (2008). The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* 26, 317–353. doi: 10.1146/annurev.immunol.26.021607.090350

ETHICS STATEMENT

The animal study was reviewed and approved by IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Brazil; permanent license number 111375-1).

AUTHOR CONTRIBUTIONS

DM, AL, RS, and NS performed the data analysis. DM wrote the manuscript. DM, AL, AC, MM, HS, and MB contributed with data interpretation and manuscript preparation. CF performed the experiments and Illumina sequencing. HS and MB conceived and designed the work and reviewed the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by CNPq grant 302970/2016-9 to HS.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Alcides Pissinatti, head of the Centro de Primatologia do Rio de Janeiro, for providing the samples, the Bioinformatics Core Facility (INCA-RJ) for their support and Dr. Thiago Parente for valuable comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00831/full#supplementary-material>

- Darc, M., Schrago, C. G., Soares, E. A., Pissinatti, A., Menezes, A. N., Soares, M. A., et al. (2012). Molecular evolution of $\alpha 4$ integrin binding site to lentiviral envelope proteins in new world primates. *Infect. Genet. Evol.* 12, 1501–1507. doi: 10.1016/j.meegid.2012.05.009
- Daugherty, M. D., and Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* 46, 677–700. doi: 10.1146/annurev-genet-110711-155522
- de Oliveira, E. H. C., Neusser, M., Pieczarka, J. C., Nagamachi, C., Sbalqueiro, I. J., and Müller, S. (2005). Phylogenetic inferences of *Atelinae* (Platyrrhini) based on multi-directional chromosome painting in *Brachyteles arachnoides*, *Ateles paniscus paniscus* and *Ateles b. marginatus*. *Cytogenet. Genom. Res.* 108, 183–190. doi: 10.1159/000080814
- DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. Available online at: <http://www.pymol.org> (accessed June 4, 2020).
- Di Fiore, A., Chaves, P. B., Cornejo, F. M., Schmitt, C. A., Shanee, S., Cortés-Ortiz, L., et al. (2015). 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. Phylogenet. Evol.* 82(Pt B), 495–510. doi: 10.1016/j.ympev.2014.03.028
- Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., et al. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522–W525. doi: 10.1093/nar/gkm276
- Dumas, F., and Mazzoleni, S. (2017). Neotropical primate evolution and phylogenetic reconstruction using chromosomal data. *Eur. Zool. J.* 84, 1–18. doi: 10.1080/11250003.2016.1260655
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Fam, B. S. O., Reales, G., Vargas-Pinilla, P., Paré, P., Viscardi, L. H., Sortica, V. A., et al. (2019). AVPR1b variation and the emergence of adaptive phenotypes in Platyrrhini primates. *Am. J. Primatol.* 81:e23028. doi: 10.1002/ajp.23028
- Fletcher, W., and Yang, Z. (2010). The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol.* 27, 2257–2267. doi: 10.1093/molbev/msq115
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- George, R. D., McVicker, G., Diederich, R., Ng, S. B., MacKenzie, A. P., Swanson, W. J., et al. (2011). Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Res.* 21, 1686–1694. doi: 10.1101/gr.121327.111
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Graham, S. A., Jégouzo, S. A. F., Yan, S., Powlesland, A. S., Brady, J. P., Taylor, M. E., et al. (2009). Prolectin, a glycan-binding receptor on dividing B cells in germinal centers. *J. Biol. Chem.* 284, 18537–18544. doi: 10.1074/jbc.M109.012807
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hancks, D. C., Hartley, M. K., Hagan, C., Clark, N. L., and Elde, N. C. (2015). Overlapping patterns of rapid evolution in the nucleic acid sensors cGAS and OAS1 suggest a common mechanism of pathogen antagonism and escape. *PLoS Genet.* 11:e1005203. doi: 10.1371/journal.pgen.1005203
- Hauck, C. R., Agerer, F., Muenzner, P., and Schmitter, T. (2006). Cellular adhesion molecules as targets for bacterial infection. *Eur. J. Cell Biol.* 85, 235–242. doi: 10.1016/j.ejcb.2005.08.002
- He, L., Gu, W., Wang, M., Chang, X., Sun, X., Zhang, Y., et al. (2018). Extracellular matrix protein 1 promotes follicular helper T cell differentiation and antibody production. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8621–8626. doi: 10.1073/pnas.1801196115
- Hu, Y., Song, F., Jiang, H., Nuñez, G., and Smith, D. E. (2018). SLC15A2 and SLC15A4 mediate the transport of bacterially derived Di/tripeptides to enhance the nucleotide-binding oligomerization domain-dependent immune response in mouse bone marrow-derived macrophages. *J. Immunol.* 201, 652–662. doi: 10.4049/jimmunol.1800210
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Hulme, A. E., Bogerd, H. P., Cullen, B. R., and Moran, J. V. (2007). Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* 390, 199–205. doi: 10.1016/j.gene.2006.08.032
- Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L., et al. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J. Virol.* 85, 11220–11234. doi: 10.1128/JVI.05238-11
- International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945. doi: 10.1038/nature03001
- Kaján, G. L., Doszpoly, A., Tarján, Z. L., Vidovszky, M. Z., and Papp, T. (2020). Virus-host coevolution with a focus on animal and human DNA viruses. *J. Mol. Evol.* 88, 41–56. doi: 10.1007/s00239-019-09913-4
- Kammerer, R., Popp, T., Härtle, S., Singer, B. B., and Zimmermann, W. (2007). Species-specific evolution of immune receptor tyrosine based activation motif-containing CEACAM1-related immune receptors in the dog. *BMC Evol. Biol.* 7:196. doi: 10.1186/1471-2148-7-196
- Kammerer, R., Popp, T., Singer, B. B., Schlender, J., and Zimmermann, W. (2004). Identification of allelic variants of the bovine immune regulatory molecule CEACAM1 implies a pathogen-driven evolution. *Gene* 339, 99–109. doi: 10.1016/j.gene.2004.06.023
- Kammerer, R., and Zimmermann, W. (2010). Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. *BMC Biol.* 8:12. doi: 10.1186/1741-7007-8-12
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Koehler, A. B., Pereira, L. C. M., Nicola, P. A., Angelo, A. C., and Weber, K. S. (2005). The southern muriqui, *Brachyteles arachnoides*, in the State of Paraná: current distribution, ecology, and the basis for a conservation strategy. Neotropical primates. *A Newsletter of the Neotropical Section of the IUCN/SSC Primate Specialist Group*, Gland: IUCN, Supp. 13, 67–72.
- Kosakovsky Pond, S. L., and Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105
- Kosiol, C., Vinar, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., et al. (2008). Patterns of positive selection in six Mammalian genomes. *PLoS Genet.* 4:e1000144. doi: 10.1371/journal.pgen.1000144
- Krupp, A., McCarthy, K. R., Ooms, M., Letko, M., Morgan, J. S., Simon, V., et al. (2013). APOBEC3G polymorphism as a selective barrier to cross-species transmission and emergence of pathogenic SIV and AIDS in a primate host. *PLoS Pathog.* 9:e1003641. doi: 10.1371/journal.ppat.1003641
- Kuespert, K., Pils, S., and Hauck, C. R. (2006). CEACAMs: their role in physiology and pathophysiology. *Curr. Opin. Cell Biol.* 18, 565–571. doi: 10.1016/j.ceb.2006.08.008
- Lan, Y., Sun, J., Xu, T., Chen, C., Tian, R., Qiu, J.-W., et al. (2018). De novo transcriptome assembly and positive selection analysis of an individual deep-sea fish. *BMC Genom.* 19:394. doi: 10.1186/s12864-018-4720-z
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12:323. doi: 10.1186/1471-2105-12-323

- Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J. A., Stewart, R., et al. (2014). Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biol.* 15:553. doi: 10.1186/s13059-014-0553-5
- Lim, J. K., Lisco, A., McDermott, D. H., Huynh, L., Ward, J. M., Johnson, B., et al. (2009). Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. *PLoS Pathog.* 5:e1000321. doi: 10.1371/journal.ppat.1000321
- Losse, J., Zipfel, P. F., and Józsi, M. (2010). Factor H and factor H-related protein 1 bind to human neutrophils via complement receptor 3, mediate attachment to *Candida albicans*, and enhance neutrophil antimicrobial activity. *J. Immunol.* 184, 912–921. doi: 10.4049/jimmunol.0901702
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10. doi: 10.14806/ej.17.1.200
- Maudhoo, M. D., Ren, D., Gradnigo, J. S., Gibbs, R. M., Lubker, A. C., Moriyama, E. N., et al. (2014). De novo assembly of the common marmoset transcriptome from NextGen mRNA sequences. *Gigascience* 3:14. doi: 10.1186/2047-217X-3-14
- Mbisa, J. L., Bu, W., and Pathak, V. K. (2010). APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. *J. Virol.* 84, 5250–5259. doi: 10.1128/JVI.02358-09
- Monath, T. P. (2012). Review of the risks and benefits of yellow fever vaccination including some new analyses. *Expert Rev. Vacc.* 11, 427–448. doi: 10.1586/erv.12.6
- Mozzi, A., Pontremoli, C., Furni, D., Clerici, M., Pozzoli, U., Bresolin, N., et al. (2015). OASes and STING: adaptive evolution in concert. *Genome Biol. Evol.* 7, 1016–1032. doi: 10.1093/gbe/evv046
- Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., et al. (2015). Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32, 1365–1371. doi: 10.1093/molbev/msv035
- Navaratnam, N., and Sarwar, R. (2006). An overview of cytidine deaminases. *Int. J. Hematol.* 83, 195–200. doi: 10.1532/IJH97.06032
- Navarro, F., Bollman, B., Chen, H., König, R., Yu, Q., Chiles, K., et al. (2005). Complementary function of the two catalytic domains of APOBEC3G. *Virology* 333, 374–386. doi: 10.1016/j.virol.2005.01.011
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi: 10.1038/nmeth.3337
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782. doi: 10.1038/s41587-019-0114-2
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170. doi: 10.1371/journal.pbio.0030170
- Olival, K. J., Hosseini, P. R., Zambrana-Torrel, C., Ross, N., Bogich, T. L., and Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature* 546, 646–650. doi: 10.1038/nature22975
- Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., et al. (2011). A molecular phylogeny of living primates. *PLoS Genet.* 7:e1001342. doi: 10.1371/journal.pgen.1001342
- Perlman, S. (2020). Another decade, another coronavirus. *N. Engl. J. Med.* 382, 760–762. doi: 10.1056/NEJMe2001126
- Perry, G. H., Melsted, P., Marion, J. C., Wang, Y., Bainer, R., Pickrell, J. K., et al. (2012). Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* 22, 602–610. doi: 10.1101/gr.130468.111
- Pond, S. L. K., and Muse, S. V. (2005). “Hyphy: hypothesis testing using phylogenies” in *Statistical Methods in Molecular Evolution*, ed. R. Nielsen (New York, NY: Springer-Verlag), 125–181. doi: 10.1007/0-387-27733-1_6
- Possas, C., Lourenço-de-Oliveira, R., Tauil, P. L., Pinheiro, F. D. P., Pissinatti, A., Cunha, R. V. D., et al. (2018). Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. *Mem. Inst. Oswaldo Cruz* 113, e180278. doi: 10.1590/0074-02760180278
- Rausell, A., and Telenti, A. (2014). Genomics of host-pathogen interactions. *Curr. Opin. Immunol.* 30, 32–38. doi: 10.1016/j.coi.2014.06.001
- Ribeiro, I. P., Menezes, A. N., Moreira, M. A. M., Bonvicino, C. R., Seuánez, H. N., and Soares, M. A. (2005a). Evolution of cyclophilin A and TRIM5 retrotransposition in New World primates. *J. Virol.* 79, 14998–15003. doi: 10.1128/JVI.79.23.14998-15003.2005
- Ribeiro, I. P., Schrago, C. G., Soares, E. A., Pissinatti, A., Seuánez, H. N., Russo, C. A. M., et al. (2005b). CCR5 chemokine receptor gene evolution in New World monkeys (*Platyrrhini*, Primates): implication on resistance to lentiviruses. *Infect. Genet. Evol.* 5, 271–280. doi: 10.1016/j.meegid.2004.07.009
- Sarkar, S. N., Ghosh, A., Wang, H. W., Sung, S. S., and Sen, G. C. (1999). The nature of the catalytic domain of 2'-5'-oligoadenylate synthetases. *J. Biol. Chem.* 274, 25535–25542. doi: 10.1074/jbc.274.36.25535
- Sawyer, S. L., Emerman, M., and Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* 2:E275. doi: 10.1371/journal.pbio.0020275
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026
- Schneider, H., Canavez, F. C., Sampaio, I., Moreira, M. A., Tagliaro, C. H., and Seuánez, H. N. (2001). Can molecular data place each neotropical monkey in its own branch? *Chromosoma* 109, 515–523. doi: 10.1007/s004120000106
- Schneider, H., Schneider, M. P., Sampaio, I., Harada, M. L., Stanhope, M., Czelusniak, J., et al. (1993). Molecular phylogeny of the New World monkeys (*Platyrrhini*, primates). *Mol. Phylogenet. Evol.* 2, 225–242. doi: 10.1006/mpev.1993.1022
- Schrago, C. G., Menezes, A. N., Moreira, M. A. M., Pissinatti, A., and Seuánez, H. N. (2012). Chronology of deep nodes in the neotropical primate phylogeny: insights from mitochondrial genomes. *PLoS One* 7:e51699. doi: 10.1371/journal.pone.0051699
- Silva, N. I. O., Sacchetto, L., de Rezende, I. M., Trindade, G. D. S., LaBeaud, A. D., de Thoisy, B., et al. (2020). Recent sylvatic yellow fever virus transmission in Brazil: the news from an old disease. *Virology J.* 17, 9. doi: 10.1186/s12985-019-1277-7
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32, 1342–1353. doi: 10.1093/molbev/msv022
- Soares, E. A., Menezes, A. N., Schrago, C. G., Moreira, M. A. M., Bonvicino, C. R., Soares, M. A., et al. (2010). Evolution of TRIM5alpha B30.2 (SPRY) domain in new world primates. *Infect. Genet. Evol.* 10, 246–253. doi: 10.1016/j.meegid.2009.11.012
- Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., et al. (2010). Adaptation genomics: the next generation. *Trends Ecol. Evol.* 25, 705–712. doi: 10.1016/j.tree.2010.09.002
- Strier, K. B., Possamai, C. B., Tabacow, F. P., Pissinatti, A., Lanna, A. M., Rodrigues de Melo, F., et al. (2017). Demographic monitoring of wild muriqui populations: criteria for defining priority areas and monitoring intensity. *PLoS One* 12:e0188922. doi: 10.1371/journal.pone.0188922
- Strier, K. B., Tabacow, F. P., de Possamai, C. B., Ferreira, A. I. G., Nery, M. S., de Melo, F. R., et al. (2019). Status of the northern muriqui (*Brachyteles hypoxanthus*) in the time of yellow fever. *Primates* 60, 21–28. doi: 10.1007/s10329-018-0701-8
- Talebi, M., Bastos, A., and Lee, P. C. (2005). Diet of southern muriquis in continuous Brazilian atlantic forest. *Int. J. Primatol.* 26, 1175–1187. doi: 10.1007/s10764-005-6463-3
- Talebi, M., Melo, F., Rylands, A. B., Ferraz, D. S., Ingberman, B., Mittermeier, R. A., et al. (2019). *The IUCN Red List of Threatened Species 2019. The IUCN Red List of Threatened Species*. Available online at: <https://dx.doi.org/10.2305/IUCN.UK.2019-2.RLTS.T2993A17927228.en> (accessed January 30, 2020).
- Talebi, M., and Soares, P. (2005). Conservation research on the southern muriqui (*Brachyteles arachnoides*) in São Paulo State, Brazil. *Neotropical primates. A Newsletter of the Neotropical Section of the IUCN/SSC Primate Specialist Group*, Gland: IUCN, Supp. 13, 53–59.

- Talebi, M. G., and Lee, P. C. (2010). Activity patterns of *Brachyteles arachnoides* in the largest remaining fragment of Brazilian Atlantic forest. *Int. J. Primatol.* 31, 571–583. doi: 10.1007/s10764-010-9414-6
- Thamizhmani, R., and Vijayachari, P. (2014). Association of dengue virus infection susceptibility with polymorphisms of 2'-5'-oligoadenylate synthetase genes: a case-control study. *Braz. J. Infect. Dis.* 18, 548–550. doi: 10.1016/j.bjid.2014.03.004
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Unni, S., Huang, Y., Hanson, R. M., Tobias, M., Krishnan, S., Li, W. W., et al. (2011). Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* 32, 1488–1491. doi: 10.1002/jcc.21720
- Vargas-Pinilla, P., Paixão-Côrtes, V. R., Paré, P., Tovo-Rodrigues, L., Vieira, C. M., de, A. G., et al. (2015). Evolutionary pattern in the OXT-OXTR system in primates: coevolution and positive selection footprints. *Proc. Natl. Acad. Sci. U.S.A.* 112, 88–93. doi: 10.1073/pnas.1419399112
- Voges, M., Bachmann, V., Kammerer, R., Gophna, U., and Hauck, C. R. (2010). CEACAM1 recognition by bacterial pathogens is species-specific. *BMC Microbiol.* 10:117. doi: 10.1186/1471-2180-10-117
- Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 45, W130–W137. doi: 10.1093/nar/gkx356
- Webb, B., and Sali, A. (2016). Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* 54, 5.6.1–5.6.37. doi: 10.1002/cpbi.3
- Williams, B. R., and Kerr, I. M. (1978). Inhibition of protein synthesis by 2'-5' linked adenine oligonucleotides in intact cells. *Nature* 276, 88–90. doi: 10.1038/276088a0
- Yang, H., Ito, F., Wolfe, A. D., Li, S., Mohammadzadeh, N., Love, R. P., et al. (2020). Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. *Nat. Commun.* 11:632. doi: 10.1038/s41467-020-14377-y
- Yang, L., Emerman, M., Malik, H. S., and McLaughlin, R. N. (2020). Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. *eLife* 9:e058436. doi: 10.7554/eLife.58436
- Yang, L., Wang, Y., Zhang, Z., and He, S. (2014). Comprehensive transcriptome analysis reveals accelerated genic evolution in a Tibet fish, *Gymnodiptychus pachycheilus*. *Genome Biol. Evol.* 7, 251–261. doi: 10.1093/gbe/evu279
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Zimmermann, W., and Kammerer, R. (2016). Coevolution of paired receptors in *Xenopus carinoembryonic* antigen-related cell adhesion molecule families suggests appropriation as pathogen receptors. *BMC Genom.* 17:928. doi: 10.1186/s12864-016-3279-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Moreira, Lamarca, Soares, Coelho, Furtado, Scherer, Moreira, Seuánez and Boroni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.