Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

High-quality nanopore assemblies of three endangered Neotropical primates

Gabriela Pozo
Universidad San Francisco de Quito
Martina Albuja-Quintana
Universidad San Francisco de Quito
Carolina Sáenz
Universidad San Francisco de Quito
Anthony Fiore
The University of Texas at Austin
Maria de Lourdes Torres
ltorres@usfq.edu.ec

Universidad San Francisco de Quito

Research Article

Keywords: Reference genome assembly, Oxford Nanopore sequencing, Long-read sequencing, Conservation genomics, Genome annotation, Endangered primates, Ateles fusciceps fusciceps, Ateles belzebuth, Cebus aequatorialis, Ecuador

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11284566/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Ecuador harbors exceptional primate diversity but faces severe conservation challenges, with many species threatened by deforestation, hunting, and habitat fragmentation. Among the most threatened are the Ecuadorian white-fronted capuchin (*Cebus aequatorialis*), the brown-headed spider monkey (*Ateles fusciceps fusciceps*), and the white-bellied spider monkey (*Ateles belzebuth*), all of which are at risk of extinction. High-quality reference genomes are essential for monitoring genetic diversity, identifying inbreeding, and guiding evidence-based conservation, yet genomic resources for Ecuadorian mammals remain scarce. Here, we present Oxford Nanopore–based de novo genomic assemblies for *C. aequatorialis* and *A. belzebuth*, together with an improved assembly for *A. f. fusciceps*. Sequencing produced coverages of 21X–32X and genome sizes of 2.6–2.7 Gb. Assemblies reached contig N50 values of 7.6–14.9 Mb, largest contig sizes of 44.9–60.9 Mb, and high completeness (97–98% complete BUSCOs). Approximately 41% of each genome consisted of repetitive elements, similar to proportions reported for other platyrrhine primate genomes assemblies and largely dominated by LINEs and LTR elements. Structural comparisons revealed near-perfect colinearity between the two *Ateles* genomes, with no evidence of major chromosomal rearrangements, whereas extensive structural rearrangements were observed between *Ateles* and *Cebus*, as expected given their deep evolutionary divergence. Gene annotation predicted 24,000–31,000 protein-coding genes in each of the three species. These assemblies provide high-quality genomic references for *C. aequatorialis* and *A. belzebuth*. Beyond addressing a critical gap in genomic information for Neotropical primates, this work establishes a practical and replicable workflow for generating high-quality genomic data in biodiversity-rich yet resource-limited regions. Together, these genomic resources will support future studies of genetic diversity, population history, and adaptive variation, thereby strengthening conservation strategies for Ecuador's most threatened primates.

1. Background

1.1 Conservation context

Ecuador is internationally recognized as one of the most biodiverse countries in the world, functioning as a critical stronghold for the conservation of Neotropical ecosystems [1, 2]. This is particularly evident in its exceptional primate diversity: despite its relatively small geographic area compared to other Neotropical countries such as Brazil, Colombia, Peru and Bolivia, Ecuador harbors 22 formally described taxa, placing it fifth in primate species richness in the region [3]. However, this richness is paralleled by significant conservation challenges: eleven species are currently classified as threatened and nine as near threatened, based on both national and global (e.g., IUCN) assessments [4, 5]. This precarious conservation status reflects escalating anthropogenic pressures on natural habitats, including deforestation, habitat fragmentation, and wildlife exploitation [4, 5]. The present study focuses on three taxa of particular concern: the Ecuadorian white-fronted capuchin (*Cebus aequatorialis*), the brown-headed spider monkey (*Ateles fusciceps fusciceps*), and the white-bellied spider monkey (*Ateles*

belzebuth). All three are at risk of extinction and require immediate, evidence-based conservation actions to prevent further population declines.

The Ecuadorian white-fronted capuchin (*C. aequatorialis*) is currently classified as Critically Endangered both on the IUCN Red List and under Ecuador's national conservation framework [6, 7]. Its distribution spans the coastal lowlands and western slopes of the Andes in Ecuador, extending southward from the Esmeraldas and Guayllabamba rivers into the Tumbes region of northwestern Peru [7, 8]. The principal threats to this species are large-scale habitat loss and fragmentation, primarily driven by agricultural expansion, livestock grazing, and urban encroachment [9–11]. These anthropogenic pressures have significantly reduced forest cover, degraded habitat quality, and contributed to the spatial isolation of populations, factors that collectively disrupt gene flow and compromise the long-term viability of the species [11, 12]. Furthermore, illegal hunting and capture for pet trade continue to exert additional pressure on remaining populations [6]. These threats are particularly concerning given the species' presumed low reproductive rates and dependence on complex social organization, traits characteristic of Neotropical capuchins that hinder rapid population recovery [13].

The brown-headed spider monkey (*A. f. fusciceps*) is classified as Critically Endangered in the IUCN Red List [14] and at the national level in Ecuador [15]. Its distribution in Ecuador is restricted to the Chocó humid forests, coastal deciduous forests, and the western foothills of the Andes. The largest known population is concentrated in Esmeraldas Province, although isolated individuals have also been documented in remnant forest fragments in Manabí Province [4, 15–17]. Although the subspecies is generally considered endemic to Ecuador, a recent phylogeographic study suggests that *A. f. fusciceps* may also occur in southern Colombia, potentially extending its distribution north of the Ecuadorian border [18]. This species demonstrates a pronounced dependence on large, contiguous tracts of undisturbed primary forest, rendering it highly vulnerable to anthropogenic disturbances, particularly deforestation and habitat fragmentation. While some individuals have been observed in degraded or fragmented habitats, these areas are likely ecologically suboptimal and insufficient to sustain viable long-term populations. Although subsistence hunting contributes to localized population pressures, habitat loss and fragmentation represent the most significant threats to the species' survival in the wild [4, 15, 16]. Furthermore, brown-headed spider monkeys have been found to be susceptible to reverse transmission of zoonoses such as SARS-CoV-2, which could lead to spillovers and outbreaks during epidemics or pandemics [19].

The white-bellied spider monkey (*A. belzebuth*) is currently classified as Endangered by both the IUCN Red List and Ecuador's national conservation framework [7, 20]. In Ecuador, its distribution is restricted to the Amazonian Tropical Humid Forest and the Eastern Piedmont Forest, primarily south of the Napo River in the western Amazon [4, 20]. Population assessments indicate a decline of at least 50% over the past three generations, primarily driven by extensive habitat conversion [20]. This species is ecologically sensitive to anthropogenic pressures, including large-scale deforestation, habitat fragmentation, and subsistence hunting [21, 22]. Additionally, mining expansion and associated infrastructure development further exacerbate habitat degradation and reduce landscape connectivity, posing a severe long-term

threat to the viability of remaining populations. Data from Ecuador's Ministry of Environment indicate that the regions with the highest deforestation rates overlap significantly with the range of *A. belzebuth* [23]. Supporting this, Suárez et al. (2013) reported notable population declines in moderately disturbed sectors of Yasuní National Park and local extirpation in heavily disturbed areas [24].

1.2 Value of reference genomes

High-quality reference genomes are now the cornerstone of wildlife genetics and conservation genomics [25]. When short-read resequencing data are mapped to a contiguous assembly, they allow researchers to infer past demographic events, including population bottlenecks and expansions [26], perform genome-wide scans of runs of homozygosity to quantify inbreeding and deleterious load [27], identify loci that show signs of selection that may reflect local adaptation or disease resistance [25], and forensically trace or assign confiscated or traded individuals to their geographic source [28]. All of these are critical inputs for evidence-based management. Yet the global distribution of reference genomes is heavily skewed toward countries of the Global North. A recent meta-analysis found that tropical tetrapods are underrepresented among available reference genomes and that genome-wide sequencing approaches are disproportionately applied to taxa in the Global North [29]. Ecuador provides a clear example of this disparity. In our meta-survey of NCBI GenBank, DNA Zoo, Bat1K and GenomeArk (accessed July/01/2025), we located only a handful of publicly available *de novo* assemblies for mammals native to the country, and just one of those was generated from tissue collected within Ecuador itself [30–33].

Given this scarcity of native reference genomes, population-level studies in most Ecuadorian lineages still rely on heterologous references or reduced-representation markers, both of which limit resolution and can bias demographic inferences [34, 35]. Producing additional, Ecuador-sourced reference genomes will therefore help address an important geographic gap in global genomic resources, enable local researchers to conduct high-resolution analyses without reliance on external data, and accelerate conservation decision-making—for example, by identifying genetically depauperate populations that may require translocations or captive breeding [35, 36]. The assemblies presented here for *A. belzebuth* and *C. aequatorialis*, together with an updated *A. f. fusciceps* reference genome, expand the number of high-contiguity genomes available for Ecuadorian mammals and lay the groundwork for comprehensive, genome-informed conservation across the country's megadiverse fauna.

1.3 Study aims

The present study had two goals. First, we set out to generate high-quality reference genomes for the white-bellied spider monkey (*A. belzebuth*) and the Ecuadorian white-fronted capuchin (*C. aequatorialis*), two primates classified as Endangered and Critically Endangered, respectively. Second, we aimed to improve our earlier *A. f. fusciceps* draft to the same standard, ensuring a coherent genomic framework for comparative and conservation analyses across Ecuadorian atelines and cebids.

To achieve these goals, we designed and implemented a streamlined long-read workflow that can be reproduced in laboratories equipped with Oxford Nanopore Technologies (ONT) devices, a platform now

found in several Ecuadorian field stations and universities. The pipeline is species-agnostic: every step from DNA extraction to annotation relies on open-source tools or broadly available reagents, making it adaptable to not only to other mammals, but also to birds, reptiles, amphibians or invertebrates.

2. Methods

2.1 Sample collection and permits

For *A. f. fusciceps*, a 5 ml blood sample was obtained from a subadult male born at the Hacienda Jambelí Rescue Center (2°46'30.48"S, 79°44'9.51"O; Guayas province, Ecuador). This male was transferred to the TUERI Wildlife Hospital (TUERI-USFQ) for veterinary treatment following injuries sustained at the rescue center. The sample was collected by TUERI-USFQ veterinary staff during a medical examination as part of his care. Genetic data were obtained under Genetic Resources Permit No. MAE-DNB-CM-2019-0126, granted to INABIO by the Ministerio del Ambiente, Agua y Transición Ecológica (MAATE), in accordance with Ecuadorian legislation.

For *C. aequatorialis*, 1ml of blood was collected from an adult male individual housed at the Isla del Tapir Zoo (0°07'01.98"S, 79°27'18"O; Santo Domingo province, Ecuador) during a routine veterinary checkup on May 20, 2023. Sampling was performed by Dr. Carolina Sáenz, and the specimen's genetic data were obtained under Genetic Resources Permit No. MAE-DNB-CM-2019-0126, granted to INABIO by MAATE, in compliance with Ecuadorian law.

For *A. belzebuth*, a blood sample was collected during a veterinary examination of a 5-year-old male individual, born in August 2018 and housed at the AmaZOOnico Wildlife Rescue Center in Tena (Napo province, Ecuador). The sample was collected on January 19–20, 2024, by veterinarians of Fundación Cóndor Andino and transported under Guide No. 01930. Sample collection was conducted under Research Permit No. MAATE-ARSFC-2023-3338, held by Fundación Cóndor Andino. Blood was requested for genomic research during the planned veterinary procedure, and the sample was subsequently donated to INABIO. INABIO confirmed that genomic analyses could be conducted under its framework agreement for access to genetic resources, No. MAE-DNB-CM-2019-0126, following deposition of the sample in its collection.

The project was reviewed by the Committee on Ethics in the Use of Animals in Research and Teaching of Universidad San Francisco de Quito (USFQ) and determined to be exempt from ethics approval (Letter of Endorsement No. 2026-012). This determination was based on the use of samples obtained through standard hospital procedures or from gene banks, with no direct contact with or manipulation of animals as part of the project methodology.

No experimental procedures were performed specifically for this study. All blood samples were collected opportunistically during routine veterinary examinations or clinical treatment by licensed veterinarians. These veterinary procedures were undertaken independently of the genomic study; blood for genomic analysis was obtained during these appointments.

All samples were stored at − 80°C in the Biotechnology & Genomics Laboratory at Universidad San Francisco de Quito (USFQ) prior to DNA extraction and analysis.

2.2 DNA Extraction, library preparation and ONT Sequencing

Genomic DNA of all three species of primates was extracted using the Qiagen DNeasy Blood and Tissue Kit starting with 50 µl of blood preserved in K2 EDTA tubes (QIAGEN, Valencia, CA, USA). DNA was eluted in a final total volume of 60 µl of ultrapure water (combining two 30 µl elutions from the same spin column). Quantity and DNA integrity were evaluated using Qubit Fluorometric Quantitation and a 1% agarose gel electrophoresis, respectively.

Long-read sequencing libraries were constructed using the Ligation Sequencing Kit SQK-LSK112 (Oxford Nanopore Technologies) for *A. f. fusciceps* and *C. aequatorialis*, and the SQK-LSK114 kit (Oxford Nanopore Technologies) for *A. belzebuth*. Sequencing was performed on a MinION with R10.4 flow cells for *A. f. fusciceps* and *C. aequatorialis*, and on a PromethION with R10.4 flow cells for *A. belzebuth*. Protocol and platform differences reflected advancements in ONT technology as genome sequencing progressed. Real-time basecalling was performed using Guppy v5.1.13 (ONT) for the MinION datasets and Dorado v1.0.2 (ONT) for the PromethION dataset to generate raw FASTQ reads for each primate genome.

2.3 Read preprocessing and QC

Raw ONT reads from each individual were processed to remove adapters and low-quality sequences prior to genome assembly. Adapter trimming was performed with Porechop v0.2.4, which identifies and excises sequencing adapters from the ends of reads [37]. Species-specific filtering thresholds were applied with NanoFilt [38]: for *A. f. fusciceps*, reads with a quality score Q < 9 were discarded. For *Ateles belzebuth*, reads with Q < 7 were removed, whereas for *C. aequatorialis*, reads shorter than 2 kb and with Q < 7 were excluded prior to assembly. These thresholds balanced yield and accuracy across datasets while improving downstream assembly quality. Read quality statistics, including N50 read length, total yield, and quality score distributions, were summarized with NanoPlot v1.42.0, providing a visual and quantitative overview of sequencing performance across runs [38].

2.4 De novo assembly, mapping and polishing

De novo genome assemblies were generated with Flye v2.7.1 [39], a long-read assembler optimized for repetitive eukaryotic genomes. Assemblies were constructed directly from the species-specific, adapter- and quality-filtered ONT read sets described in the previous section, thereby minimizing the inclusion of low-quality or excessively short fragments. The Flye assembler was chosen to assemble the genomes of these primate species based on previous experience [33] and Flye's use of the generalized de Bruijn graph approach, which is well suited for genomes containing extensive noncoding regions and repetitive sequences [40, 41].

To improve consensus accuracy, filtered ONT reads from each species were mapped to their respective draft assemblies minimap2 v2.26 [42], and polishing was performed in a species-specific manner. The *A. f. fusciceps* assembly was polished with a a single round of Medaka v1.7.2 (r103 model) to refine per-base accuracy [43], whereas the A. belzebuth and C. aequatorialis assemblies were each polished with two iterations of Racon v1.5.0 [44]. These species-specific pipelines yielded the most complete and contiguous assemblies for each case, and additional polishing rounds did not significantly improve assembly quality.

Polished contigs were subsequently ordered and oriented by aligning them to closely related reference genomes using minimap2. *A. f. fusciceps* contigs were aligned to the *Ateles geoffroyi* assembly (GenBank GCA_015710855.1), *A. belzebuth* contigs to our generated *A. f. fusciceps* assembly, and *C. aequatorialis* contigs to the *C. albifrons* assembly (GCF_013371975.1). This reference-guided step was used solely to order and orient contigs and did not alter the underlying contig sequences.

2.5 Assembly evaluation and annotation

Assembly contiguity and completeness were assessed using both reference-free and reference-guided approaches. QUAST v5.0.2 [45] was employed to compute standard contiguity metrics including total assembly size, N50, L50, and GC content. Gene-space completeness was estimated with BUSCO v5.4.3 [46] using the *primates_odb10* dataset.

Structural genome annotation was carried out with OmicsBox v2.1.0 (BioBam Bioinformatics). Gene prediction was performed using the built-in *de novo* annotation workflow, which integrates ab initio prediction with evidence-based alignment of homologous sequences from related primates. Although MAKER v3.01.03 was initially tested, OmicsBox produced the most consistent and complete gene models across the three assemblies. Consequently, OmicsBox structural annotation was adopted as the final approach for *A. f. fusciceps*, *Ateles belzebuth*, and *C. aequatorialis*.

2.6 Genome analyses and comparison

To assess genome macrosynteny we first extracted the 50 longest scaffolds (≥ 1 Mb) from each assembly with SeqKit v2.6.1 [47], a set that retains long genomic sequences while omitting short, repeat-rich contigs. For the closely related *A. belzebuth* and *A. f. fusciceps* genomes, we aligned scaffolds with NUCmer (MUMmer v4.0.0-rc1) in maximal-unique-match mode (--mum -l 100 -c 500) [48]. The resulting *.delta* file was filtered with delta-filter (-1 -i 95 -l 10000) to keep only one-to-one blocks of ≥ 10 kb and with ≥ 95% identity. Because *C. aequatorialis* diverged from *Ateles* ≥ 18 Myr (average identity ≈ 85%), we used more permissive settings: NUCmer with all-match seeding (--maxmatch -l 40 -c 200), followed by a global best-hit filter (delta-filter -g -i 85 -l 5000). These parameters recover syntenic blocks while suppressing repetitive noise.

Filtered alignments were rendered with mummerplot (--filter --fat; --layout omitted for cross-family plots) and the accompanying *gnuplot* script was standardized to a 4000 × 4000 pixels canvas.

Transposable elements (TEs) were identified and quantified for each primate assembly using a standardized workflow implemented on a Linux high-performance computing cluster (64 CPU cores, 512 GB RAM per node). All tools were installed within a dedicated Conda environment created from the *bioconda* and *conda-forge* channels to ensure reproducibility. RepeatModeler v2.0.5 was used to construct species-specific consensus libraries of repetitive elements [49]. For each genome, a BLAST database was first generated and RepeatModeler was executed with the LTRStruct pipeline enabled, which improves recovery of full-length long terminal repeat (LTR) retrotransposons. The resulting libraries (3–8 Mb) combined novel consensus families with entries matched to Dfam/RepBase under our institutional license.

Repeat masking was subsequently performed with RepeatMasker open-5.0.4, configured with RMBlast v2.14.0 as the search engine. Each genome was soft-masked against its respective custom library using 32 parallel threads. The *-xsmall* option was applied so that masked bases were written in lower-case, allowing downstream gene-prediction tools to ignore repeats without altering sequence coordinates, while *-gccalc* generated GC-content statistics for masked and unmasked regions. This process produced masked FASTA files (*.masked), per-hit annotation tables (*.out), summary reports (*.tbl), and GFF3 tracks (*.gff) suitable for genome browsers.

To obtain genome-wide statistics of repeat coverage, we processed RepeatMasker output files with the utility *buildSummary.pl*. The resulting summary tables were parsed to extract the proportion of each genome annotated as repetitive, as well as subclass-level breakdowns of transposable elements. These detailed counts were collated with *summarizeRM.pl* to generate comparative files across species. The total proportion of masked bases was further extracted into a concise tab-delimited table (*repeat_genome_fraction.tsv*), and subclass contributions were summarized in *TE_breakdown.tsv*.

2.7 Genome-wide SNP discovery and heterozygosity

To estimate genome-wide individual heterozygosity, the reads from each species were mapped to the corresponding *de novo* genome assembly using minimap2 v2.26 with parameters optimized for long-read alignment. Resulting alignments were sorted and indexed with samtools, and single-nucleotide polymorphisms (SNPs) were identified using bcftools v1.17 [50]. Variant calling was performed with bcftools call in multiallelic mode, and raw variant sets were filtered to retain high-confidence sites by applying minimum quality and depth thresholds (QUAL > 30; DP > 10; DP < 100). Only SNPs were retained for downstream analyses, and multiallelic sites were excluded to focus on biallelic variation. Heterozygous SNPs were subsequently extracted for each individual.

Genome-wide individual heterozygosity was calculated as the ratio of heterozygous biallelic SNPs to the total number of callable genomic sites. Callable sites were defined as positions containing unambiguous nucleotides (A, C, G, or T) in the corresponding genome assembly, excluding gaps and ambiguous bases. Because estimates were derived from single individuals per species, heterozygosity values reflect individual-level genetic diversity rather than population-wide variation. Variant call quality was further

assessed using transition/transversion (Ti/Tv) ratios computed with bcftools stats, providing an additional check on the reliability of SNP datasets generated from long-read sequencing data.

2.8 SSR characterization and comparison

To characterize SSR profiles for each species, we used a modified version of the perl script MISA [51, 52] to identify microsatellite repeats with 2 to 6 bp repeat motifs. We also ran MISA on 11 additional published reference genomes for species of *Ateles* and *Cebus* available in GenBank as of 2026-02-01. We then compared SSR profiles from the three genomes generated in our study using nanopore sequencing to those extracted from these additional genomes, most of which were generated using a different sequencing technology (8 using Illumina “sequencing-by-synthesis” short-read sequencing, one using PacBio Single-Molecule Real-Time long-read sequencing).

Additionally, for one species (*A. belzebuth*), we performed a reciprocal comparison of the set of SSRs identified from our genomes generated using nanopore sequencing with the set identified in the reference genome published in GenBank (Accession Number GCA 963573605.1). The modified MISA script and configuration file, data files, code, and documentation used for our comparative analyses are provided in Supplementary Material and on GitHub (https://github.com/difiore/comparative-ssr-analysis).

3. Results

3.1 Sequencing output and assembly statistics

Oxford Nanopore sequencing of *A. f. fusciceps*, *C. aequatorialis*, and *A. belzebuth* produced 52.98 Gb, 86.04 Gb, and 76.04 Gb of data, corresponding to estimated coverages of 20.4X, 31.8X, and 29.2X, respectively. Coverage was calculated based on the predicted genome size of the closely related species *A. geoffroyi* (2.6 Gb; NCBI: JAKFHY000000000). Mean read length ranged from 6.34 kb to 8.01 kb, with mean quality scores between 10.1 and 14.7 (Table 1).

Table 1
Sequencing statistics of the three primate species used in this study

Statistic	***A. f. fusciceps***	***C. aequatorialis***	***A. belzebuth***
Number of bases	52.98 Gb	86.04 Gb	76.04 Gb
Read Count	8.35 million	18.44 million	9.4 million
Coverage	20.4X	31.8X	29.2X
Read N50	13,103	9,635	15,525
Mean Read Length	6.34 kb	4.67 kb	8.01 kb
Mean Q Score	11.2	10.1	14.7

The three Flye assemblies ranged from 2.6 to 2.7 Gb and contained 2,088–3,845 contigs. Assembly contiguity was highest for *C. aequatorialis*, with an N50 of 14.9 Mb (L50 = 55), followed by *A. belzebuth* (N50 = 12.8 Mb; L50 = 57) and *A. f. fusciceps* (N50 = 7.6 Mb; L50 = 97). The largest contigs ranged from 44.9 to 60.9 Mb (Table 2; Fig. 1).

Table 2
Comparative statistics for the draft assemblies of *A. f. fusciceps*, *C. aequatorialis*, and *A. belzebuth* using Flye and polished with Medaka and Racon

Statistic	***A. f. fusciceps***	***C. aequatorialis***	***A. belzebuth***
Contig n°	3,845	2,997	2,088
Total size	2.6 Gb	2.7 Gb	2.6 Gb
N50/N90	7.6 Mb / 1.3 Mb	14.9 Mb / 2.8 Mb	12.8 Mb / 2.5 Mb
Largest contig	44.9 Mb	55.3 Mb	60.9 Mb
L50/L90	97 / 368	55 / 210	57 / 228
Complete BUSCOs	97.4%	97.9%	98%
Single Copy	96.1%	96.3%	96.1%
Duplicated	1.4%	1.6%	1.9%
Fragmented BUSCOs	1.4%	1.2%	1.3%
Missing BUSCOs	1.2%	0.9%	0.7%

BUSCO analysis using the primates_odb10 dataset recovered 97.4–98.0% complete orthologs across the three assemblies. Single-copy complete BUSCOs represented 96.1–96.3%, duplicated BUSCOs 1.4–1.9%, fragmented BUSCOs 1.2–1.4%, and missing BUSCOs 0.7–1.2% (Table 2; Fig. 1).

3.2 Repeat landscape and gene set snapshot

Repeat annotation identified a similar repetitive fraction in all three genomes, accounting for 39.05–39.76% of assembled sequence (Table 3). Retroelements were the most abundant category (27.36–27.64%), including LINEs (12.66–13.51%), LTR elements (9.58–11.30%), and SINEs (2.56–5.40%). Unclassified repeats accounted for 9.15–9.82% of each assembly, whereas DNA transposons, simple repeats, satellites, and small RNAs each represented smaller fractions (Table 3).

Table 3
Summary of repetitive elements identified in the genomes of the three primate species (*A. f. fusciceps*, *C. aequatorialis*, and *A. belzebuth*).

Statistics (%)	*A. f. fusciceps*	*C. aequatorialis*	*A. belzebuth*
Total	39.30	39.76	39.05
Retroelements	27.64	27.57	27.36
LINEs	12.66	12.68	13.51
SINEs	5.40	3.92	2.56
LTR Elements	9.58	10.98	11.30
DNA Transposons	0.67	0.84	0.68
Simple Repeats	0.99	1.16	0.95
Satellites	0.15	0.73	0.00
Small RNA	0.03	0.07	0.04
Unclassified /others	9.59	9.15	9.82

Genome annotation predicted 25,506 protein-coding genes in *A. f. fusciceps*, 30,653 in *C. aequatorialis*, and 24,252 in *A. belzebuth* (Table 4). The annotations contained 169,040–203,316 exons and 143,534–172,663 introns. Mean exons per mRNA ranged from 6.6 to 7.5, while mean gene length ranged from 26,344 to 32,314 bp (Table 4).

Table 4
Summary statistics for the annotated genomes of *A. f. fusciceps*, *C. aequatorialis*, and *A. belzebuth*

Statistic	*A. f. fusciceps*	*C. aequatorialis*	*A. belzebuth*
Number of genes	25506	30653	24252
Number of exons	169040	203316	182212
Number of introns (in CDS)	143534	172663	157960
Mean mRNAs per-gene	1.0	1.0	1.0
Mean exons per-mRNA	6.6	6.6	7.5
Mean introns per-mRNA	5.6	5.6	6.5
Mean gene length (bp)	30053	26344	32314
Mean exon length (bp)	160	162	190
Mean intron length (bp)	5151	4485	4741

3.3 Structural findings

Whole-genome comparison of *A. belzebuth* and *A. f. fusciceps* showed a nearly continuous forward diagonal, with 98% of aligned sequence occurring within colinear blocks ≥ 10 kb (Fig. 2A). No megabase-scale reverse-alignment blocks or translocations were detected, and the few discontinuities in the alignment coincided primarily with scaffold boundaries.

In contrast, comparisons of *C. aequatorialis* with each *Ateles* genome produced a more fragmented alignment pattern containing numerous short forward and reverse segments (Figs. 2B and 2C). Several long *C. aequatorialis* scaffolds aligned to multiple consecutive regions of the *Ateles* assemblies, while off-diagonal and reverse-alignment blocks included segments of approximately 1–3 Mb together with smaller scattered matches.

3.4 Genome-wide patterns of genetic diversity

Genome-wide individual heterozygosity differed among the three reference individuals (Table 5). The *A. belzebuth* individual exhibited the highest heterozygosity, with 9,827,656 heterozygous biallelic SNPs across 2,648,975,690 callable bases ($H = 3.71 \times 10^{-3}$). The *A. f. fusciceps individual* contained 3,797,917 heterozygous biallelic SNPs across 2,639,259,576 callable bases ($H = 1.44 \times 10^{-3}$), whereas *C. aequatorialis* had 2,806,151 heterozygous biallelic SNPs across 2,704,274,059 callable bases ($H = 1.04 \times 10^{-3}$).

Transition/transversion (Ti/Tv) ratios were 2.37 for *A. f. fusciceps*, 1.79 for *C. aequatorialis*, and 2.16 for *A. belzebuth* (Table 5).

Table 5
SNP-based heterozygosity estimates for the three primate genomes

Species	Heterozygous biallelic SNPs	Callable bases (bp)	Heterozygosity (H)	Ti/Tv
A. f. fusciceps	3,797,917	2,639,259,576	1.44×10^{-3}	2.37
C. aequatorialis	2,806,151	2,704,274,059	1.04×10^{-3}	1.79
A. belzebuth	9,827,656	2,648,975,690	3.71×10^{-3}	2.16

3.5 SSR profile comparison and reciprocal SSR detection

Using MISA, the number of SSRs recovered across the 14 *Ateles* and *Cebus* genomes analyzed in this study and from GenBank averaged 70,755 (± 20,094 SD), ranging from 18,179 in *C. albifrons* to 112,240 in *C. aequatorialis* (Additional File 1 - Tables S1 and S2). Total SSR counts did not differ detectably among sequencing technologies ($F_{2,11} = 0.049$, $p = 0.953$; Additional File 2 - Figure S1).

Within both *Ateles* and *Cebus*, 2- to 6-bp 'perfect' SSRs comprised > 91% of all SSRs recovered in every genome. Their length distributions were similar across genomes and sequencing technologies (Fig. 3). The distribution of SSR counts across repeat motif lengths (Additional File 2 - Figure S2), the most

abundant canonical motifs (Additional File 2 - Figure S3), and occupancy of canonical SSR motif space (Additional File 2 - Figure S4) were also broadly comparable among assemblies.

In the reciprocal comparison of the two *A. belzebuth* assemblies, BLAST searches of the 300-bp upstream and downstream flanking regions surrounding 'perfect' 2- to 6-bp SSRs recovered candidate homologous loci in the alternative assembly in > 90% of cases (Additional File 1 - Tables S3a and S3b). Specifically, 95.7% of loci identified in the short-read assembly were recovered in the nanopore assembly, and 93.1% of loci identified in the nanopore assembly were recovered in the short-read assembly.

After applying the stringent reciprocal-matching criteria described in the Methods, 43,652 high-confidence reciprocal SSRs were shared between the two *A. belzebuth* assemblies. These represented approximately 77–80% of the subset of 'perfect' SSR loci for which full-length 300-bp flanking regions could be extracted and both flanks successfully aligned to the alternative genome (~ 55,000 SSRs per assembly) (Fig. 4). The modal absolute length difference among reciprocal SSRs was 0 bp, and the distribution of length differences was strongly centered at zero (Additional File 2 - Figure S5).

4. Discussion

4.1 Sequencing performance and assembly quality

The genomes for the Ecuadorian white-fronted capuchin (*Cebus aequatorialis*), the brown-headed spider monkey (*Ateles fusciceps fusciceps*), and the white-bellied spider monkey (*Ateles belzebuth*) were generated with 20.4–32× ONT coverage and reached high contiguity and gene-space completeness despite differences in sequencing kits and platforms. Variation in read yield and mean quality score likely reflects these technical differences. Comparable ONT-based primate assemblies have been generated at approximately 20–30× coverage, including *Eulemur rufifrons* and *Plecturocebus cupreus* [53, 54], suggesting that this coverage range can provide a practical balance between sequencing effort and assembly quality for large primate genomes. The 2.6–2.7 Gb assembly sizes of *C. aequatorilias, A. belzebuth* and *A. f. fusciceps* are also consistent with available *Ateles* and *Cebus* genomes. The largest contigs in our assemblies (44.9–60.9 Mb) approach the size of small primate chromosomes [55], illustrating the long-range continuity achievable with nanopore reads. Nevertheless, these assemblies remain contig-level rather than chromosome-level references, and large contig size alone does not establish that complete chromosomes were recovered.

4.2 Repeat landscape and genome annotation

Repeat characterization and gene annotation provided additional insights into the genomic composition of the three species. Approximately 39–40% of each assembly was annotated as repetitive sequence, with retroelements forming the dominant category. This overall repeat content is broadly comparable to the substantial transposable-element fractions reported in human, chimpanzee, and rhesus macaque genomes [56, 57]. LINEs and SINEs are prominent components of primate genomes and have contributed to primate genome evolution; LINE-1 elements predate the origin of primates, whereas Alu

elements are characteristic of the primate lineage and have been widely used in evolutionary analyses [56]. The similar repeat fractions observed across the three assemblies therefore support a broadly conserved repeat landscape while also providing species-specific resources for future comparative analyses.

The predicted gene count for *C. aequatorialis* (30,653) was higher than that reported for several other annotated primate genomes, including *Callithrix jacchus* and *Sapajus apella*, and should therefore be interpreted cautiously [58, 59]. Residual indels in ONT-only assemblies and fragmentation of gene models can inflate predicted gene numbers [60]. At the same time, other annotation characteristics, including mean gene length, exon number, and exon length, were broadly comparable among the three species (Table 4). Together with the high BUSCO completeness of the assemblies, these results indicate that the annotations provide useful gene-set resources, while emphasizing that future incorporation of species-specific transcriptomic evidence and higher-accuracy sequence data would improve gene-model resolution.

4.3 Structural conservation and divergence among genomes

Whole-genome comparisons were used to assess patterns of structural conservation and divergence among the three species. The near-continuous colinearity between *A. belzebuth* and *A. f. fusciceps* indicates substantial conservation of genome organization within Ateles. In contrast, comparisons between *C. aequatorialis* and the two *Ateles* assemblies showed more fragmented forward and reverse alignments, consistent with deeper evolutionary separation between Cebidae and Atelidae [61, 62]. The observed off-diagonal and reverse-alignment blocks may reflect lineage-specific inversions or other structural changes accumulated during independent chromosomal evolution [63, 64]. However, because the assemblies have not been ordered and oriented with Hi-C or another long-range scaffolding approach, individual rearrangements should be considered provisional rather than definitive. Chromosome-level scaffolding will be necessary to distinguish biological rearrangements from assembly-boundary effects and to characterize structural evolution with greater confidence.

4.4 Individual heterozygosity and conservation context

Genome-wide heterozygosity was estimated for each reference individual to provide an initial assessment of individual genomic variation in the three taxa. The three reference individuals differed markedly in genome-wide heterozygosity, with the *A. belzebuth* individual showing the highest value and the *C. aequatorialis* individual the lowest (Table 5). Reduced genomic diversity has been documented in primates and other mammals that have experienced long-term population decline, small effective population size, or isolation [35, 65–67]. In this context, the lower values observed in *A. f. fusciceps* and *C. aequatorialis* are notable because both taxa have restricted distributions and are classified as Critically Endangered, whereas the more broadly distributed *A. belzebuth* is classified as Endangered [6, 14, 20]. Habitat loss, fragmentation, and population decline affect all three taxa and can reduce effective population size and genetic diversity over time [6, 14, 20, 68].

These comparisons must nevertheless be interpreted at the individual rather than species level. Each estimate was obtained from a single reference individual, so the values do not measure population-wide heterozygosity and cannot establish differences in genetic diversity among the three species. The Ti/Tv ratios provide an internal quality check on the filtered SNP datasets [69], but population-level resequencing across each species' geographic range will be required to determine whether the differences observed here are representative of broader demographic patterns. The present estimates therefore provide baseline values and hypotheses for future conservation-genomic sampling rather than direct estimates of species-wide diversity.

4.5 SSR profiles and marker-development potential

Genome-wide SSR profiles were compared among the three species and, for *A. belzebuth*, across sequencing technologies to evaluate their consistency and potential for marker development. SSR profiles were highly similar within genera and across sequencing technologies, and sequencing technology had no detectable effect on total SSR counts. The reciprocal *A. belzebuth* comparison further showed that most SSR loci with suitable flanking sequence could be recovered between independently generated nanopore and short-read assemblies, with 43,652 loci meeting the stringent reciprocal criteria. These results support the use of nanopore assemblies for characterizing genome-wide SSR diversity, consistent with the increasing application of long-read sequencing for the characterization of repetitive and tandem-repeat regions [70]. Reciprocal loci that differ in repeat length are particularly useful candidates for future marker development, although their polymorphism should be validated in population samples before they are used for population-genetic inference [71].

4.6 Conservation applications of the reference genomes

The principal conservation value of these assemblies is that they provide species-relevant references for future population resequencing. Reads from wild or captive individuals can be mapped to these genomes to quantify genetic diversity, reconstruct demographic history, identify runs of homozygosity, and investigate deleterious variation [25, 27, 35, 36]. These analyses can help identify isolated or genetically depauperate populations and provide evidence relevant to decisions about connectivity, managed breeding, or translocation. The references can also support development of SNP and SSR marker panels for wildlife forensic applications; although a reference genome alone cannot assign confiscated material to geographic origin, it enables discovery of informative variants that can subsequently be validated for assignment and trafficking investigations [28].

Genome-enabled studies may also identify candidate regions associated with local adaptation or disease-related variation [25]. Such applications are relevant for threatened primates exposed to rapid habitat change and emerging pathogens. However, management decisions should rely on population-level data and ecological information rather than on reference-genome characteristics alone. In this sense, the assemblies presented here are enabling resources: their immediate contribution is to make higher-resolution conservation-genomic analyses feasible for these Ecuadorian primates and to strengthen locally generated genomic capacity.

4.7 Limitations and future directions

Several limitations define the appropriate interpretation of this study. First, the assemblies were generated primarily from ONT long reads and lack Hi-C or optical-mapping data; they are therefore highly contiguous draft assemblies rather than chromosome-level references, and large structural rearrangements cannot yet be resolved with complete confidence. Second, residual ONT sequencing errors, particularly small indels, may persist after polishing and may affect some coding sequences and gene models [60]. Third, the annotation did not incorporate species-specific transcriptomic data, which may contribute to fragmented or inflated gene predictions. Finally, each genome represents a single individual, so the heterozygosity estimates cannot characterize within-species diversity, population structure, connectivity, or genomic load. Future work combining chromosome-level scaffolding, higher-accuracy polishing or complementary sequencing, transcriptomic evidence, and resequencing of multiple populations will allow these reference genomes to support robust analyses of effective population size, inbreeding, demographic history, adaptive variation, and potential genetic rescue [25, 35, 72].

5. Conclusions

This study presents high-quality genome assemblies for two endangered Neotropical primates, the Ecuadorian white-fronted capuchin (*Cebus aequatorialis*) and the white-bellied spider monkey (*Ateles belzebuth*), as well as an improved reference for the brown-headed spider monkey (*Ateles fusciceps fusciceps*: GCA_036589455.2). By applying an Oxford Nanopore–based workflow that is both cost-effective and practical within a primate habitat country in the Global South, we achieved assemblies with high contiguity and completeness, comparable to those generated with more resource-intensive platforms.

These genomic resources fill a critical gap for Ecuadorian biodiversity, where conservation research has historically lacked locally generated references. The assemblies will enable high-resolution analyses of genetic diversity, demographic history, and inbreeding across fragmented populations, provide a foundation for wildlife forensic applications, and support evidence-based management of species currently at risk of extinction.

Beyond their immediate conservation value, our results demonstrate that high-quality genomic resources can be generated locally within biodiversity-rich regions using accessible long-read sequencing technologies. The workflow presented here, spanning DNA extraction, sequencing, assembly, and annotation, provides a practical framework that can be readily adapted by other research groups working in similar settings.

Looking forward, integrating these genomes with additional resequencing datasets, long-range scaffolding approaches, and multi-population sampling will be essential to move from reference genome generation to population-level genomic analyses. These assemblies establish a foundational genomic

resource for Ecuador and the broader Amazon–Chocó region that will support future evolutionary, ecological, and conservation research.

Abbreviations

BUSCO
Benchmarking Universal Single-Copy Orthologs CDS:Coding Sequence Gb:Gigabase Gbp:Gigabase pairs GC:Guanine–Cytosine GFF3:General Feature Format version 3 GPU:Graphics Processing Unit Hi-C:High-throughput Chromosome Conformation Capture HPC:High-Performance Computing kb:kilobase LINE:Long Interspersed Nuclear Element LTR:Long Terminal Repeat Mb:Megabase mRNA:Messenger Ribonucleic Acid N50/N90:Contig length at which 50%/90% of the genome is contained in contigs of that length or longer ONT:Oxford Nanopore Technologies QC:Quality Control QUAST:Quality Assessment Tool for Genome Assemblies RNA:Ribonucleic Acid SINE:Short Interspersed Nuclear Element TE:Transposable Element

Declarations

Ethics approval

All blood samples used in this study were collected opportunistically during routine veterinary examinations or clinical treatment by licensed veterinarians. No experimental treatments were administered specifically for this study. The veterinary procedures were undertaken independently of the genomic study, and blood for genomic analysis was obtained during these appointments.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors’ Contributons

G.P.: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing M.A.-Q.: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing C.S.: Investigation, Writing – original draft

A.D.: Methodology, Formal Analysis, Writing – review & editing M.d.L.T.: Conceptualization, Methodology, Supervision, Writing – review & editing

AI Use Statement

Generative artificial intelligence (AI) tools were used during the preparation of this manuscript to assist with language editing, grammar correction, and improvements to clarity and readability. AI tools were not used for data generation, analysis, interpretation, or the formulation of scientific conclusions. All AI-assisted content was carefully reviewed, verified, and edited by the authors, who assume full responsibility for the accuracy, integrity, and final content of the manuscript.

Funding

This work was supported by Universidad San Francisco de Quito (USFQ) through COCIBA Grants 2022, 2023, 2024, and 2025. Genome sequencing was carried out within the framework of the ORG.one initiative, which provided sequencing consumables.

Author Contribution

G.P. and M.d.L.T. conceptualized the study. G.P., M.A.-Q., A.D., and M.d.L.T. developed the methodology. G.P., M.A.-Q., and C.S. conducted the investigation. G.P., M.A.-Q., and A.D. performed the formal analyses. G.P. acquired funding, and M.d.L.T. supervised the study. G.P., M.A.-Q., and C.S. wrote the original manuscript draft. G.P., M.A.-Q., A.D., and M.d.L.T. reviewed and edited the manuscript.

Acknowledgement

We thank Lizbeth Larreategui, Juan Sebastián Moscoso, and Francis Rodríguez for their contributions in assisting with genome assemblies. We extend our sincere gratitude to Fundación Cóndor Andino, Proyecto Washu and TUERI-USFQ for their support in facilitating the collection of the blood samples. We also wish to thank the members of the Biotechnology and Genomics Laboratory at Universidad San Francisco de Quito (USFQ) for their continuous assistance and collaboration throughout the realization of this project. Finally, we acknowledge the Ministerio del Ambiente, Agua y Transición Ecológica (MAATE) of Ecuador for granting the research and genetic resource permits (MAE-DNB-CM-2019-0126 and MAATE-ARSFC-2023-3338) that made this work possible.

Data Availability

All raw sequencing data and assembled genomes generated in this study have been deposited in NCBI under the following BioProject accessions: Ateles belzebuth (PRJNA1154258), Ateles fusciceps fusciceps (PRJNA1009451), and Cebus aequatorialis (PRJNA1153579). The complete laboratory and bioinformatic pipeline, including DNA extraction, sequencing, assembly, and analysis, is available at

https://dx.doi.org/10.17504/protocols.io.bp2l6e935gqe/v1 and https://github.com/difiore/comparative-ssr-analysis to ensure transparency and reproducibility.

References

1. Mittermeier RA, Turner WR, Larsen FW, Brooks TM, Gascon C. Global Biodiversity Conservation: The Critical Role of Hotspots. In: Zachos FE, Habel JC, editors. Biodiversity Hotspots [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. pp. 3–22. Available from: https://link.springer.com/10.1007/978-3-642-20992-5_1 doi:10.1007/978-3-642-20992-5_1.
2. Antonelli A, Zizka A, Carvalho FA, Scharn R, Bacon CD, Silvestro D, et al. Amazonia is the primary source of Neotropical biodiversity. Proc Natl Acad Sci USA. 2018;115(23):6034–9. 10.1073/pnas.1713819115.
3. Cervera L, Torre S, de la, Jerusalinsky L, Fuentes N, Alfonso-Cortés F, Morelos-Juárez C, et al. Conservation Action Plan for Ecuadorian Primates: Process and Priorities. Primate Conserv. 2017;31:9–15.
4. Tirira D. A field guide to the mammals of Ecuador: including the Galapagos Islands and the Ecuadorian Antarctic zone. Quito, Ecuador: Asociación Ecuatoriana de Mastozoología; 2017.
5. IUCN. The IUCN Red List of Threatened Species. Version 2024-2 [Internet]. 2024. Available from: https://www.iucnredlist.org
6. IUCN. Cebus aequatorialis: Moscoso, de la Torre P, Cornejo S, Mittermeier FM, Lynch RA, J.W., Heymann EW. The IUCN Red List of Threatened Species 2021: e.T4081A191702052 [Internet]. 2015 [cited 2025 Aug 28]. Available from: https://www.iucnredlist.org/species/4081/191702052 10.2305/IUCN.UK.2021-1.RLTS.T4081A191702052.en
7. Tirira DG, Gallo-Viracocha F. Áreas prioritarias para la conservación y vulnerabilidad al cambio climático de *Alouatta palliata aequatorialis* (Atelidae) y *Cebus aequatorialis* (Cebidae) en la provincia de Azuay, Ecuador. Mammalia aequatorialis. 2021;3:37–57.
8. Rylands AB, Mittermeier RA. Ecuadorian white-fronted capuchin (*Cebus aequatorialis*). In: Wilson DE, Mittermeier RA, editors. Handbook of the Mammals of the World. Barcelona: Lynx Edicions; 2013. p. 412.
9. Sierra R. Patrones y factores de deforestación en el Ecuador continental, 1990–2010. Y un acercamiento a los próximos 10 años [Internet]., Quito. Ecuador: Conservación Internacional Ecuador y Forest Trends; 2013. Available from: https://www.conservation.org.ec
10. González-Jaramillo V, Fries A, Rollenbeck R, Paladines J, Oñate-Valdivieso F, Bendix J. Assessment of deforestation during the last decades in Ecuador using NOAA-AVHRR satellite data. Erdkunde. 2016;217–35. 10.3112/erdkunde.2016.03.02.
11. Cervera L, Torre S, de la, Zapata-Ríos G, Alfonso-Cortés F, Álvarez-Solas S, Crowe O, et al. Working Together Towards One Goal: Results of the First Primate Census in Western Ecuador. Primate Conserv. 2018;32:1–8.

12. Estrada A, Garber PA, Rylands AB, Roos C, Fernandez-Duque E, Di Fiore A, et al. Impending extinction crisis of the world's primates: Why primates matter. Sci Adv. 2017;3(1):e1600946. 10.1126/sciadv.1600946.
13. Jack KM, Campos FA. Distribution, Abundance, and Spatial Ecology of the Critically Endangered Ecuadorian Capuchin (*Cebus Albifrons Aequatorialis*). Trop Conserv Sci. 2012;5(2):173–91. 10.1177/194008291200500207.
14. Ateles *fusciceps* ssp. *fusciceps*: Moscoso, P., IUCN, Link A, de la Torre S, Shanee S, Cortes-Ortíz L. The IUCN Red List of Threatened Species 2021: e.T39922A191687911 [Internet]. 2020 [cited 2025 Aug 28]. Available from: https://www.iucnredlist.org/species/39922/191687911 10.2305/IUCN.UK.2021-1.RLTS.T39922A191687911.en
15. Cervera L, Griffith DM. New Population and Range Extension of the Critically Endangered Ecuadorian Brown-Headed Spider Monkey (*Ateles fusciceps fusciceps*) in Western Ecuador. Trop Conserv Sci. 2016;9(1):167–77. 10.1177/194008291600900109.
16. Morelos Juarez C, Tapia A, Cervera L, Alfonso-Cortez F, Fuentes N, Araguillin E et al. Distribución actucología y estrategias para la conservación de un primate críticamente amenazado (*Ateles fusciceps fusiceps*) en el Ecuador. In: Distribución actucología y estrategias para la conservación de un primate críticamente amenazado [Internet]. University of Sussex; 2018. Available from: https://hdl.handle.net/10779/uos.23457710.v1
17. Gallo-Viracocha F, Urgilés-Verdugo C, Fuentes N, Alfonso-Cortes F, Zurita-Arthos L, Torres TC, et al. Distribution, conservation, and vulnerability to climate change of the Ecuadorian Brown-headed Spider Monkey (Primates: Atelidae): Distribución, conservación y vulnerabilidad al cambio climático del mono araña de cabeza marrón ecuatoriano (Primates: Atelidae). Mamm aequatorialis. 2022;4:39–52. 10.59763/mam.aeq.v4i.50.
18. Rivera Román ES. Filogeografía de uno de los primates más amenazados del mundo, el mono araña de cabeza café (*Ateles fusciceps*) [Tesis de Maestría en Ciencias Biológicas] [Internet]. [Bogotá, Colombia]: Universidad de los Andes; 2024. Available from: https://repositorio.uniandes.edu.co/server/api/core/bitstreams/5199afd7-cc13-4054-ac55-593da4161a4f/content
19. Carvajal M, Saenz C, Fuentes N, Guevara R, Muñoz E, Prado-Vivar B et al. SARS-CoV-2 infection in brown-headed spider monkeys (*Ateles fusciceps*) at a wildlife rescue center on the coast of Ecuador—South America. Medina RA, editor. Microbiol Spectr. 2024;12(4):e02741-23. 10.1128/spectrum.02741-23
20. IUCN, Ateles belzebuth. Link, A., Palacios, E., Stevenson, P.R., Boubli, J.P., Mittermeier, R.A., Shanee, S., Urbani, B., de la Torre, S., Cornejo, F.M., Moscoso, P., Mourthé, Í., Muniz, C.C. & Rylands, A.B.: The IUCN Red List of Threatened Species 2021: e.T2276A191684587 [Internet]. 2019 [cited 2025 Aug 28]. Available from: https://www.iucnredlist.org/species/2276/191684587 10.2305/IUCN.UK.2021-1.RLTS.T2276A191684587.en

21. Di Fiore A, Link A, Dew JL. Diets of wild spider monkeys. In: Campbell CJE, editor. Spider Monkeys: Behavior, Ecology and Evolution of the Genus *Ateles*. Cambridge University Press; 2008. pp. 81–137. (Cambridge Studies in Biological and Evolutionary Anthropology).
22. Link A, Milich K, Di Fiore A. Demography and life history of a group of white-bellied spider monkeys (*Ateles belzebuth*) in western Amazonia. Am J Primatol. 2018;80(8):e22899. 10.1002/ajp.22899.
23. Ministerio del Ambiente. Deforestación del Ecuador continental periodo 2014–2016 [Internet]. Ministerio del Ambiente. 2017. Available from: http://certificacionpuntoverde.ambiente.gob.ec/libraries/EAlfresco.php/?doc=5708eb09-80c7-4c92-aca0-21dfa0ee711b
24. Suárez E, Zapata-Ríos G, Utreras V, Strindberg S, Vargas J. Controlling access to oil roads protects forest cover, but not wildlife communities: a case study from the rainforest of Yasuní Biosphere Reserve (Ecuador). Anim Conserv. 2013;16(3):265–74. 10.1111/j.1469-1795.2012.00592.x.
25. Theissinger K, Fernandes C, Formenti G, Bista I, Berg PR, Bleidorn C, et al. How genomics can help biodiversity conservation. Trends Genet. 2023;39(7):545–59. 10.1016/j.tig.2023.01.005.
26. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493–6. 10.1038/nature10231.
27. Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9(10):1205–18. 10.1111/eva.12414 . PubMed PMID: 27877200; PubMed Central PMCID: PMC5108213.
28. Ogden R, Linacre A. Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Sci International: Genet. 2015;18:152–9. 10.1016/j.fsigen.2015.02.008.
29. Linck EB, Cadena CD. A Latitudinal Gradient of Reference Genomes. Mol Ecol. 2024;34(23):e17551. 10.1111/mec.17551.
30. Consortium DZ. Chromosome-length genome assembly of the capybara (*Hydrochoerus hydrochaeris*) [Internet]. 2022. Available from: https://www.dnazoo.org/assemblies/hydrochoerus_hydrochaeris
31. Consortium DZ. Chromosome-length genome assembly of the spectacled bear (*Tremarctos ornatus*) [Internet]. 2021. Available from: https://www.dnazoo.org/assemblies/tremarctos_ornatus
32. Simmons NB, Ingala MR, Pieri M, O'Toole BP, Gray JL, Philge P, et al. The genome sequence of *Molossus alvarezi* González-Ruiz, Ramírez-Pulido and Arroyo-Cabrales, 2011 (Chiroptera, Molossidae). Wellcome Open Res. 2024;9:522. 10.12688/wellcomeopenres.22726.1.
33. Pozo G, Albuja-Quintana M, Larreátegui L, Gutiérrez B, Fuentes N, Alfonso-Cortés F et al. First whole-genome sequence and assembly of the Ecuadorian brown-headed spider monkey (*Ateles fusciceps fusciceps*), a critically endangered species, using Oxford Nanopore Technologies. G3 Genes|Genomes|Genetics. 2024;14(3):jkae014. 10.1093/g3journal/jkae014
34. Lin MJ, Iyer S, Chen NC, Langmead B. Measuring, visualizing, and diagnosing reference bias with biastools. Genome Biol. 2024;25(1):101. 10.1186/s13059-024-03240-8.

35. Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, et al. The era of reference genomes in conservation genomics. Trends Ecol Evol. 2022;37(3):197–202. 10.1016/j.tree.2021.11.008.
36. Brandies P, Peel E, Hogg CJ, Belov K. The Value of Reference Genomes in the Conservation of Threatened Species. Genes. 2019;10(11):846. 10.3390/genes10110846.
37. Wick R. Porechop: adapter trimmer for Oxford Nanopore reads [Internet]. 2017. Available from: https://github.com/rrwick/Porechop
38. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Berger B, editor. Bioinformatics. 2018;34(15):2666–9. 10.1093/bioinformatics/bty149
39. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6. 10.1038/s41587-019-0072-8.
40. Li Z, Chen Y, Mu D, Yuan J, Shi Y, Zhang H, et al. Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-bruijn-graph. Brief Funct Genomics. 2011;11(1):25–37. 10.1093/bfgp/elr035.
41. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348–65. 10.1038/s41587-021-01108-x.
42. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. Bioinformatics. 2018;34(18):3094–100. 10.1093/bioinformatics/bty191
43. Oxford Nanopore Technologies. medaka: Sequence correction provided by ONT Research. GitHub repository [Internet]. GitHub. 2018. Available from: https://github.com/nanoporetech/medaka
44. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46. 10.1101/gr.214270.116.
45. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5. 10.1093/bioinformatics/btt086.
46. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Kelley J, editor. Molecular Biology and Evolution. 2021;38(10):4647–54. 10.1093/molbev/msab199
47. Shen W, Le S, Li Y, Hu F, SeqKit:. A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. Zou Q, editor. PLoS ONE. 2016;11(10):e0163962. 10.1371/journal.pone.0163962
48. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. Darling AE, editor. PLoS Comput Biol. 2018;14(1):e1005944. 10.1371/journal.pcbi.1005944
49. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020;117(17):9451–7. 10.1073/pnas.1921046117.

50. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2):giab008. 10.1093/gigascience/giab008.
51. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). Theor Appl Genet. 2003;106(3):411–22. 10.1007/s00122-002-1031-0.
52. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Valencia A, editor. Bioinformatics. 2017;33(16):2583–5. 10.1093/bioinformatics/btx198
53. Hauff L, Rasoanaivo NE, Razafindrakoto A, Ravelonjanahary H, Wright PC, Rakotoarivony R, et al. *De Novo* Genome Assembly for an Endangered Lemur Using Portable Nanopore Sequencing in Rural Madagascar. Ecol Evol. 2025;15(1):e70734. 10.1002/ece3.70734.
54. Pfeifer SP, Baxter A, Savidge LE, Sedlazeck FJ, Bales KL. De Novo Genome Assembly for the Coppery Titi Monkey (*Plecturocebus cupreus*): An Emerging Nonhuman Primate Model for Behavioral Research. Eyre-Walker A, editor. Genome Biology and Evolution. 2024;16(5):evae108. 10.1093/gbe/evae108
55. Brown TA. Chapter 1: The Human Genome. In: Genomes [Internet]. 2nd ed. Wiley-Liss (now Garland Science); 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21134/
56. Konkel MK, Walker JA, Batzer MA. LINEs and SINEs of primate evolution. Evolutionary Anthropology: Issues News Reviews. 2010;19(6):236–49. 10.1002/evan.20283.
57. Navarro FCP, Galante PAF. A Genome-Wide Landscape of Retrocopies in Primate Genomes. Genome Biol Evol. 2015;7(8):2265–75. 10.1093/gbe/evv142.
58. Harris RA, Raveendran M, Warren W, LaDeana HW, Tomlinson C, Graves-Lindsay T, et al. Whole Genome Analysis of SNV and Indel Polymorphism in Common Marmosets (Callithrix jacchus). Genes. 2023;14(12):2185. 10.3390/genes14122185.
59. Byrne H, Webster TH, Brosnan SF, Izar P, Lynch JW. Signatures of adaptive evolution in platyrrhine primate genomes. Proc Natl Acad Sci USA. 2022;119(35):e2116681119. 10.1073/pnas.2116681119.
60. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 2018;19(1):90. 10.1186/s13059-018-1462-9.
61. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MAM, et al. A Molecular Phylogeny of Living Primates. PLoS Genet. 2011;7(3):e1001342. 10.1371/journal.pgen.1001342.
62. Ruiz-Herrera A, Farré M, Robinson TJ. Molecular cytogenetic and genomic insights into chromosomal evolution. Heredity (Edinb). 2012;108(1):28–36. 10.1038/hdy. 2011.102 PubMed PMID: 22108627; PubMed Central PMCID: PMC3238115.
63. Ruiz-Herrera A, García F, Giulotto E, Attolini C, Egozcue J, Ponsà M, et al. Evolutionary breakpoints are co-localized with fragile sites and intrachromosomal telomeric sequences in primates. Cytogenet Genome Res. 2005;108(1–3):234–47. doi:10.1159/000080822 PubMed PMID: 15545736.

64. Springer MS, Meredith RW, Gatesy J, Emerling CA, Park J, Rabosky DL, et al. Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix. PLoS ONE. 2012;7(11):e49521. 10.1371/journal.pone.0049521.

65. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. Nature. 2013;499(7459):471–5. 10.1038/nature12228.

66. Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. 2015;348(6231):242–5. 10.1126/science.aaa3952.

67. Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, et al. Genomic Flatlining in the Endangered Island Fox. Curr Biol. 2016;26(9):1183–9. 10.1016/j.cub.2016.02.062.

68. Frankham R. Genetics and conservation biology. CR Biol. 2003;326(S1):22–9. 10.1016/S1631-0691(03)00023-4.

69. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8. 10.1038/ng.806.

70. De Roeck A, De Coster W, Bossaerts L, Cacace R, De Pooter T, Van Dongen J, et al. NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. Genome Biol. 2019;20(1):239. 10.1186/s13059-019-1856-3.

71. Cai Y, Yu H, Liu H, Jiang C, Sun L, Niu L, et al. Genome-wide screening of microsatellites in golden snub-nosed monkey (*Rhinopithecus roxellana*), for the development of a standardized genetic marker system. Sci Rep. 2020;10(1):10614. 10.1038/s41598-020-67451-2.

72. Supple MA, Shapiro B. Conservation of biodiversity in the genomics era. Genome Biol. 2018;19(1):131. 10.1186/s13059-018-1520-3.

Figures

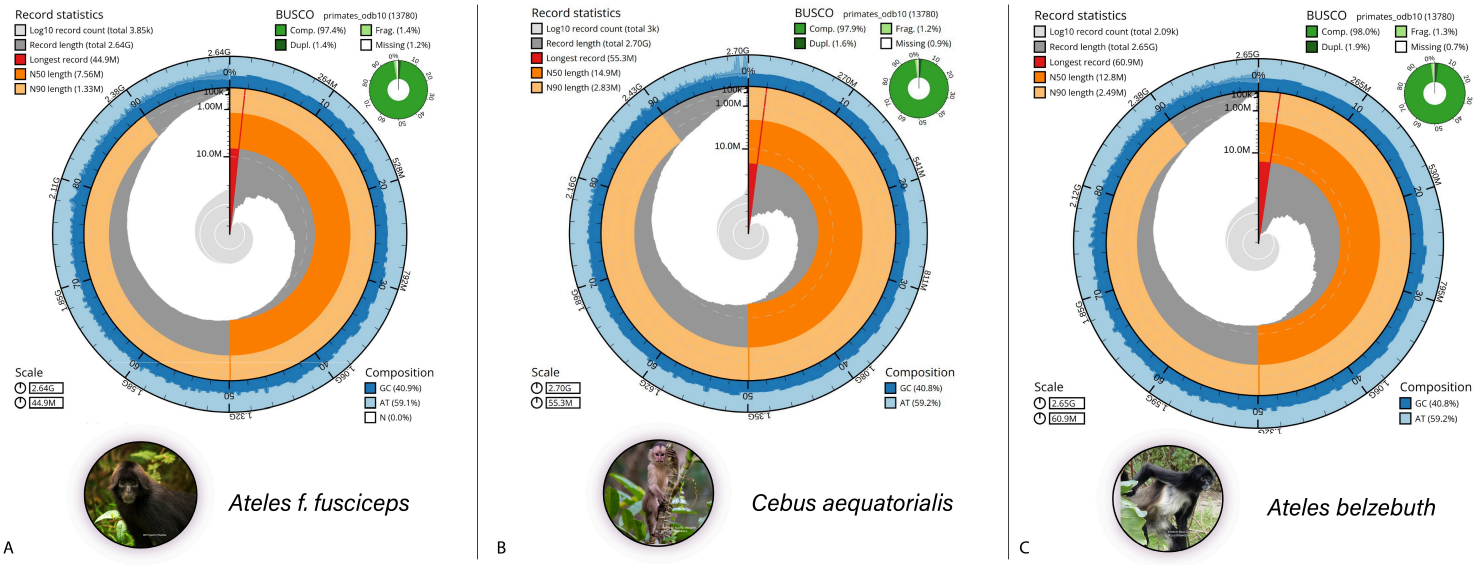


Figure 1

Genome assemblies' metrics of **A)** *Ateles fusciceps fusciceps*, **B)** *Cebus aequatorialis*, and **C)** *A. belzebuth* produced by Flye. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The plot also shows the N50 ad N90 values, as well as the GC, AT, and N compositions. A summary of the complete, duplicated, fragmented, and missing BUSCOs (primates_odb10) is represented in the right-hand corner of each Snailplot.

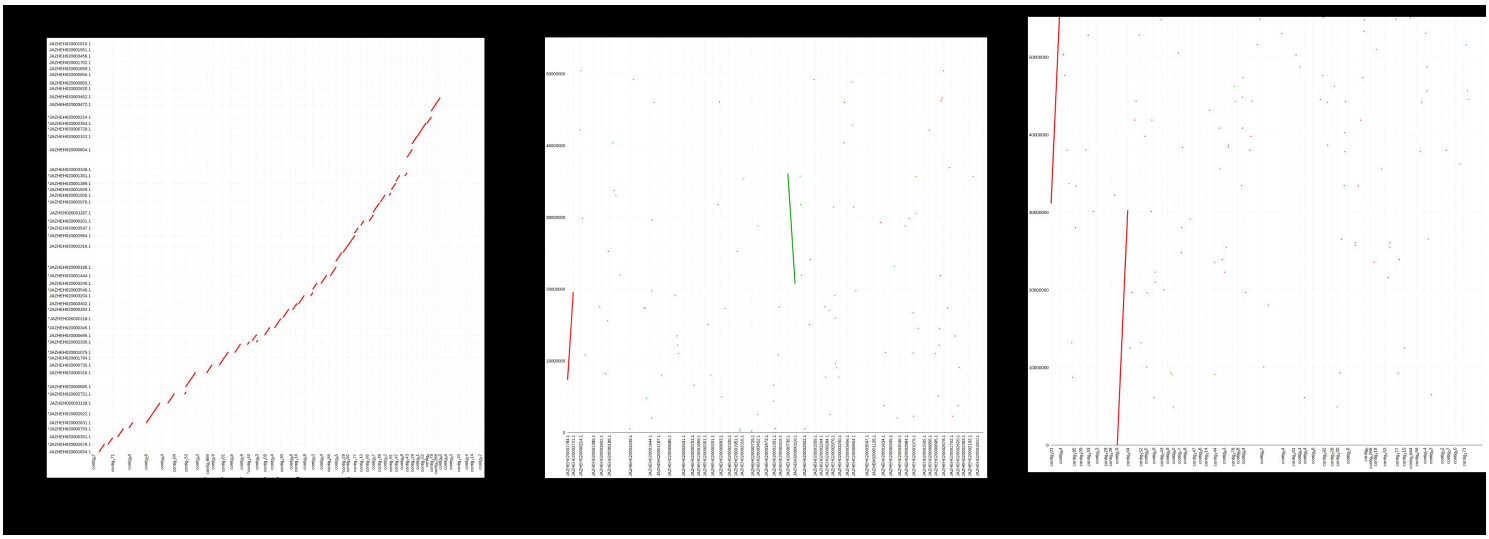


Figure 2

Whole-genome dot plots generated with MUMmer. Forward hits are in red, reverse hits are in green, and axes are labelled with species names. (A) Alignment of *Ateles belzebuth*(reference) against *Ateles fusciceps fusciceps* (query). (B) Alignment of *Cebus aequatorialis* (query) against *A. belzebuth* (reference). (C) Alignment of *C. aequatorialis* (query) against *A. f. fusciceps*(reference). Forward matches are shown in red and reverse matches in green.

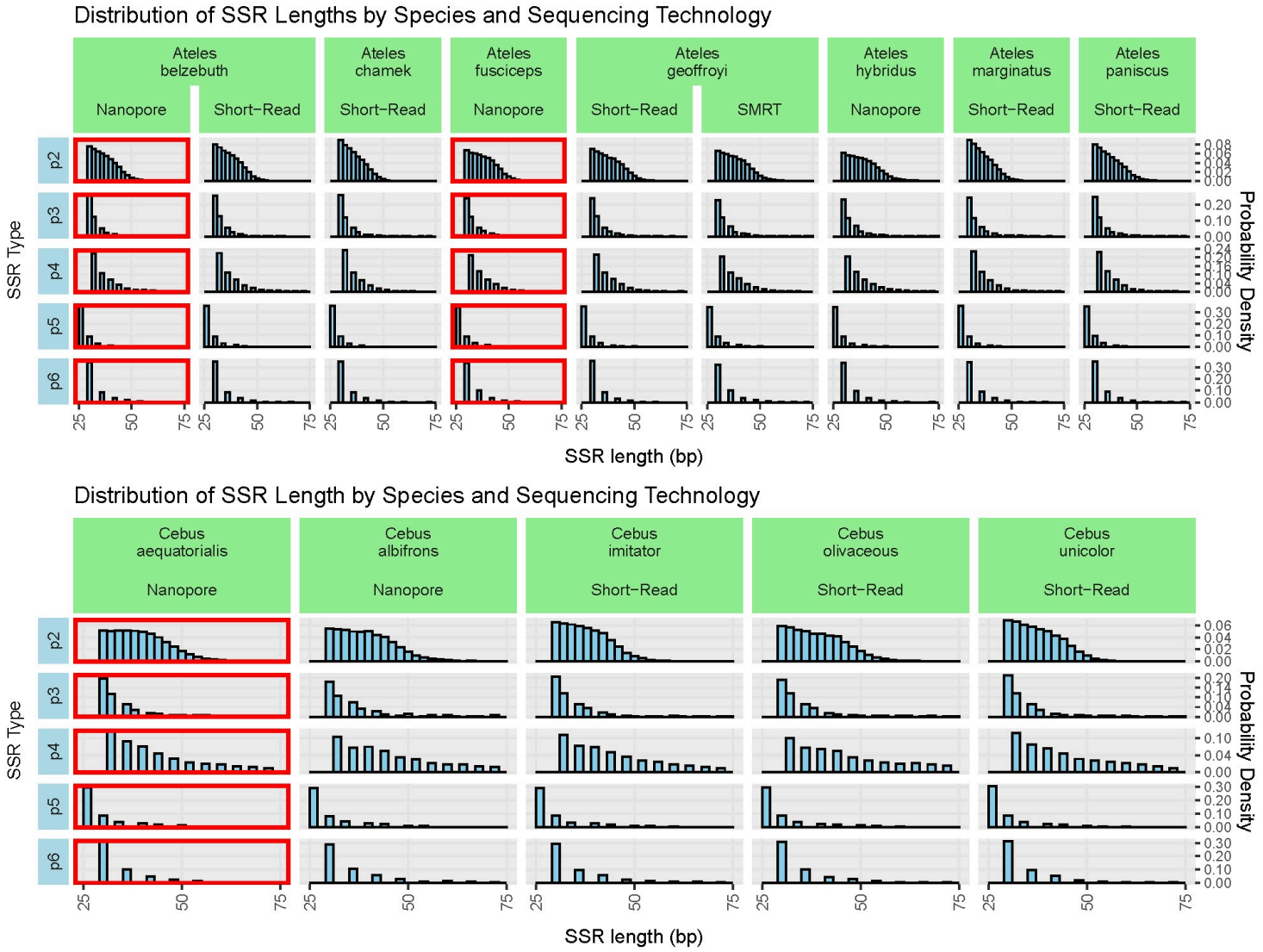


Figure 3

Length distribution of “perfect” SSRs with 2 to 6-bp repeat motifs across 9 *Ateles* and 5 *Cebus* genomes. Distributions are truncated at 75 bp because, in all genomes, more than 96% of “perfect” SSRs fall at or below this length. Red boxes indicate genomes generated using nanopore sequencing in this study. SSR length distributions are highly consistent within genera and are also similar between independent assemblies using different sequencing technologies for two species: *A. belzebuth* (nanopore and short-read sequencing) and *A. geoffroyi* (SMRT and short-read technology).

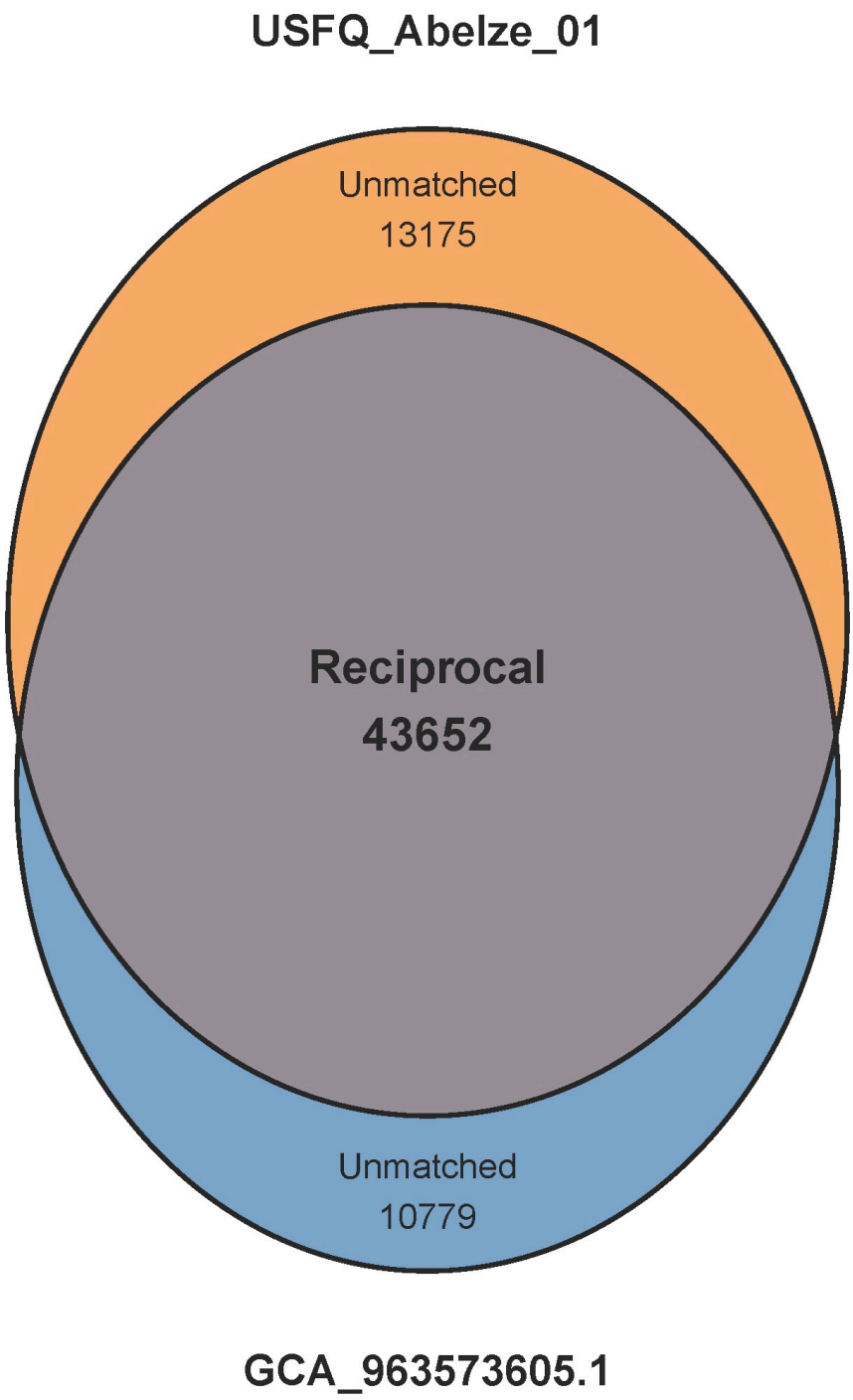


Figure 4

Venn diagram illustrating "perfect" SSRs with 2- to 6-base pair repeat motifs detected in in the two *A. belzebuth* genome assemblies: the Oxford Nanopore assembly generated in this study (Genome 1; USFQ_Abelze_01) and the published GenBank assembly (Genome 2; GCA_963573605.1). Each ellipse shows the total number of SSRs in the respective genome, and the overlapping region represents reciprocal SSRs identified under stringent matching criteria. Reciprocal SSRs were defined by successful

BLAST of both upstream and downstream flanking regions of SSRs from one genome to the alternative genome assembly, detection of an equivalent repeat motif between the BLAST hits, and mutual overlap of the SSR regions. Multi-hit loci – SSRs in one genome mapping to more than one locus in the other genome -- are included. Percentages of reciprocal SSRs relative to the total SSR count in each genome are indicated in the overlap region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1Tables.pdf
- AdditionalFile2Figures.pdf