A high-quality, long-read genome assembly of the endangered ring-tailed lemur (Lemur catta)

Marc Palmada-Flores, Joseph D. Orkin, Bettina Haase, Jacquelyn Mountcastle, Mads F. Bertelsen, Lukas F. K. Kuderna, Erich D. Jarvis, Tomas Marques-Bonet

Context

The strepsirrhines are a remarkably diverse radiation of primates that includes more than one-quarter of all recognized primate species [1]. The vast majority of strepsirrhines (103 species) are members of the Lemuroidea, colloquially known as “lemurs,” and endemic to Madagascar. Despite their geographic isolation, the lemur radiation is exceptionally diverse, including both the smallest living primate (Microcebus berthae) and one of the largest (the recently extinct subfossil lemur, Archaeolemur fontaniyoni) [2, 3].

Of particular interest, both ecologically and in the public imagination, are ring-tailed lemurs (Lemur catta, NCBI:txid9447; Fig. 1). Ring-tailed lemurs are medium-bodied, ecologically flexible members of the Lemuridae family and the sole member of the genus Lemur [4–6]. In contrast to most other Lemuridae, L. catta predominantly inhabit the dry and seasonal forests of southern Madagascar [7]. They consume an omnivorous diet mostly of fruit and leaves and engage in a multi-male multi-female social structure with a polygynandrous mating system [7]. Ring-tailed lemurs are under severe conservation pressure; they are classified as “endangered” by the International Union for Conservation of Nature (IUCN) [8], resulting primarily from deforestation, hunting, and capture for the pet trade. A recent population census has revealed a dramatic population decline, with as few as 2,200 individuals remaining [9]. Of further concern, the species is distributed across a highly fragmented range with only 8 populations of ≥100 individuals remaining [9]. Despite this near-term population decline, a recent microsatellite analysis indicates that the genetic diversity of L. catta populations could be exceptionally high, with evidence of genetic isolation by distance throughout their geographic range [6].

From a genomic perspective, relatively little is known about ring-tailed lemurs (and strepsirrhines more broadly). Genome assemblies have been published for 18 strepsirrhine species, but none of these assemblies has a contig N50 value >1 Mb, and only 3 of them are >100 kb [10]. Recently, an L. catta genome (Lem-
Cat_v1_BIUU) was assembled by the Zoonomia consortium [11], given that it is derived from Illumina short reads, its metrics and application are still limited compared to the genome quality of recent highly contiguous assemblies [12]. This general lack of genomic resources remains a limitation for the comparative and population genomics of lemurs.

Here, we present a new high-quality genome assembly of \textit{L. catta} (mLemCat1) that conforms to the standards of the Vertebrate Genomes Project (VGP). mLemCat1 was assembled with a combination of Pacific Biosciences (PacBio) continuous long reads (CLR data), Optical Mapping Bionano reads, Arima HiC data, and 10X linked-reads. Our new assembly will allow for a deep assessment of the genome biology and conservation genomics of endangered ring-tailed lemurs. Additionally, given the paucity of high-contiguity strepsirrhine assemblies, it will allow major advances in the genomics across the Lemuridae family.

**Data Description**

Bionano: San Diego/CA/USA; Qubit: Thermo Fisher Scientific: Waltham/MA/USA; PacBio: Menlo Park/CA/USA; Sage Science: Beverly/MA/USA; 10xgenomics: Pleasanton/CA/USA; Illumina: San Diego/CA/USA; Arima Genomics: San Diego/CA/USA; KAPA (Roche) Basel Switzerland.

**Library preparation and sequencing**

A sample of spleen tissue was collected post-mortem from a male at the Copenhagen Zoo (Denmark) in 2015 and immediately flash-frozen (ZIMS Global Accession No. GAN: DKL15-03 323). We isolated 30 μg of ultra-high molecular weight DNA (uHMW) from 35 mg of flash-frozen spleen tissue using the agarose plug Bionano Genomics protocol for animal tissue (DNA isolation fibrous tissue protocol No. 30071C). uHMW DNA quality was assessed by a Pulsed-Field Gel assay and quantified with a Qubit 2 Fluorometer (Supplementary Fig. S1).

A 10 μg quantity of uHMW DNA was sheared using a 26G blunt-end needle (PacBio protocol PN 101–181-000 Version 05). A large-insert PacBio library was prepared using the Pacific Biosciences Express Template Prep Kit v2.0 (No. 100–938-900) following the manufacturer protocol. The library was then size selected (>20 kb) using the Sage Science BluePippin Size-Selection System. A total of 23 PacBio 1M v3 (#101–531-000) smrtcells were sequenced on the Sequel instrument (PacBio Sequel System, RRID:SCR_017989) (sequencing kit 3.0 No. 101–597-800) with a 10-hour movie and 2 hours pre-extension time. Unfragmented uHMW DNA was used to generate a linked-reads library on the 10X Genomics Chromium (Genome Library Kit & Gel Bead Kit v2 PN-120258, Genome Chip Kit v2 PN-120257, i7 Multiplex Kit PN-120262). This 10X library was sequenced on an Illumina Novaseq (Illumina NovaSeq 6000 Se- quencing System, RRID:SCR_016387) 54 150-bp PE lane. uHMW DNA was labeled for Bionano Genomics optical mapping (Bionano Irys system, RRID:SCR_016754) using the Bionano Prep Direct Label and Stain (DLS) Protocol (30206E) and run on 1 Saphyr (Saphyr, RRID:SCR_017992) instrument chip flow cell. Hi-C preparation was performed by Arima Genomics using the Arima-HiC kit (P/N: AS10008) and an Illumina-compatible library was generated using the KAPA Hyper Prep kit (P/N: KK8504). This library was then sequenced on an Illumina HiSeq X Ten (Illumina HiSeq X Ten, RRID:SCR_016385) (150 bp PE) at ~60× coverage following the manufacturer’s protocols. Assuming a genome size of 3.21 Gb from the Goat database [13], the present genome (mLemCat1) has been produced with 86.43× of 10X linked-reads data, 66.68× of Arima data, 154.57× of Bionano data, and 62.88× of PacBio data.

**De novo assembly**

The genome was assembled following the VGP standard pipeline v1.6 [12], and the specific parameter settings are available on the VGP GitHub repository (Additional File 1). Specifically, contigs were generated using FALCON (FALCON, RRID:SCR_018804) [14] and FALCON-Unzip [15], producing primary and alternate assemblies. We used purge_dups (purge_dups, RRID:SCR_021173) [16] to identify false duplications caused by regions of high heterozygosity. Purged contigs were removed from the primary assembly and added to the alternate assembly. We then scaffolded the primary assembly using the Bionano Irys and Arima HiC data to produce a high-quality, contiguous and complete assembly of mLemCat1.
Table 1: Genome quality metrics for the mLemCat1 genome assembly compared to previous assembly and standards

| Quality category | Quality metric | VGP standard | mLemCat1 | LemCat_v1_BIUU |
|------------------|----------------|--------------|----------|----------------|
| Continuity       | No. scaffolds  | 23–480       | 141      | 575,427        |
|                  | Scaffold N50 (Mb) | 90.982       | 0.216    | 0.216          |
|                  | Largest scaffold (Mb) | 285.823       | 2.320    | 2.320          |
|                  | No. contigs | 518          | 580,026  |                |
|                  | Contig N50 (Mb) | 1–25         | 10.570   | 0.158          |
|                  | Largest contig (Mb) | 40.360       | 1.312    |                |
|                  | Gaps/Gb | 75–1,500    | 179.5    | 2,001.3        |
|                  | Span (Gb) | 2.122       | 2.298    |                |
| Structural accuracy | False duplications (%) | 0.2–5.0 | 0.39 |                |
| Base accuracy    | Base pair QV | 39–43       | 44.45    |                |
|                  | k-mer completeness (%) | 87–98 | 91.45 |                |
| Functional completeness | Genes (BUSCOs [S]) (%) | 82–98 | 93.8 | 84.6          |

No.: Number of S: single-copy genes. BUSCOs database is vertebrata_odb10 (n = 3,354) and the BUSCO version is: V4.0.6.

*For example, mitochondrial.

Figure 2: BUSCO assessment results. Comparison between mLemCat1 and LemCat_v1_BIUU Lemur catta assemblies using the Primates_ODB10 database (n = 13,780). The new mLemCat1 assembly shows a 7.3% increase in complete single-copy orthologous genes.

Assembly using 10X linked-read data with Scaff10X V2.0 [17], Bionano optical maps with Bionano Solve V2.1 [18], and Arima Hi-C data with Salsa V2.2 [19]. We assembled the mitochondrial genome separately using MitoVGP [20] with PacBio and 10X data. The primary scaffolds, alternate contigs, and mitochondrial assembly were polished simultaneously. We first performed polishing and gap filling with the original PacBio data using Arrow [14], followed by 2 rounds of short-read polishing using the 10X linked-read data. Specifically, 10X data were mapped to the assembly using Longranger V2.1.3 [21] and polishing was done with FreeBayes (FreeBayes, RRID:SCR_010761) [22]. All computing was performed on the DNAnexus (DNAnexus, RRID:SCR_011884) cloud platform.

Genome quality assessment

Compared to the currently available short-read L. catta genome (LemCat_v1_BIUU) [11], the new mLemCat1 assembly has higher contiguity values, fewer scaffolds, and a slightly smaller assembly size (Table 1). We generated basic contiguity assembly metrics for both assemblies using QUAST V5.0.2 (QUAST, RRID:SCR_001228) [23], which are presented in Table 1. The assembly has a total scaffold size of 2.122 Gb within 141 scaffolds. The mLemCat1 contig and scaffold N50 values are 10.570 and 90.982 Mb, representing 20.41- and 421.21-fold increases, respectively, compared with the LemCat_v1_BIUU assembly. In comparison with the human genome assembly (hg38), the L95 and N95 statistics (L95 = 24; N95 = 46.710 Mb for hg38; L95 = 28; N95 = 21.924 Mb for mLemCat1) are similar, given the expected chromosomes for both (22 autosomes + 2 sexual chromosomes in human, and 27 autosomal + 2 sex chromosomes for L. catta) [24]. Further comparison can be found in Supplementary Table S1. The overall GC content of this assembly is 40.48%.

The mLemCat1 assembly has a high level of accuracy and completeness that conforms to the proposed standards of the VGP [12]. We assessed the base and structural accuracies of the assembly with Merqury V1.1, using a Meryl V1.7 database [25] based on 130.708 Gb (84× coverage) of 10X linked-reads. The base pair QV of the primary assembly is 44.33, which exceeds the VGP standard. The k-mer completeness is 91.45%. We classified the structural accuracy using the false duplications percentage calculated in the false_duplications.sh script from Merqury V1.1. The assembly is estimated to have 0.39% false duplications based on the percentage of k-mers found in unexpected copy numbers.

To assess the functional completeness of the assembly, we recovered BUSCO genes from both mLemCat1 and the existing Illumina-based assembly (LemCat_v1_BIUU) (Fig. 2). Specifically, we conducted a gene completeness assessment using BUSCO (BUSCO, RRID:SCR_015008) V4.0.6 [26], setting human as the reference species in the --augustus_species parameter, and
using the primates_OrthoDB10 database, which comprises a total of 13,780 genes. Of the 13,780 possible BUSCOs, we identified 12,138 single-copy (88.1%), 100 duplicates (0.7%), and 188 fragmented genes (1.4%) in mLemCat1, leaving 9.8% of BUSCOs missing. In contrast, we could only recover 11,132 single-copy BUSCOs (80.8%) from LemCat_v1_BIUU, with 15.3% of BUSCO genes missing. We also ran the same analyses, but changed the database to vertebrata_odb10 and obtained even better results for mLemCat1 compared to LemCat_v1_BIUU, which we report in Table 1. The present assembly (mLemCat1) can be useful to create synteny plots between the present species and others, such as humans (Supplementary Fig. S2) because it has N50 statistics comparable to other high-quality catarrhine genomes [32, 33], although there are fewer satellites (0.30%), simple repeats (0.68%), and low-complexity elements (0.13%) (Supplementary Table S3).

Mitogenome of L. catta

We assembled a gapless mitochondrial genome with a span of 17,086 bp using both PacBio CLR (long reads) and 10X data (short reads) using MitoVGP V2.2 with additional parameters “-f 18000 -v LENIENT”, as described in the Additional File 2: Table S3 of Formenti et al. [20], and annotated the assembly using the MITOS2 web server [30]. With the annotation results we plotted a map of the mitochondrion with GenomeVx [31] (Supplementary Fig. S3). Thirteen main protein-coding genes have been annotated in this new mitogenome including nad1, nad2, nad3, nad4, nad4L, nad5, nad6, cox1, cox2, cox3, atp6, atp8, and cob.

Analysis of the repeatome

To assess the structure and variety of repeat elements in the L. catta genome, we analyzed mLemCat1 with RepeatMasker (RepeatMasker, RRID:SCR_012954) V4.1.2-p1. Non-default settings included the use of sensitive mode, the query assumed species set to primates, nhmmscan 3.3.2 (Nov 2020), and FamDB: HMM-Dfam_3.3, without the exclusion of simple repeats. In total, 50.32% of the bases in the L. catta genome (mLemCat1) are masked as interspersed repeats, including long and short interspersed nuclear elements, long terminal repeats, and DNA elements (Fig. 3A, Supplementary Table S3). In comparison with the previous Illumina-only assembly (LemCat_v1_BIUU) we observed minor differences in the structure and

**Figure 3:** (A) Percentages of elements in the L. catta genome (mLemCat1) masked by RepeatMasker. (B) Percentage of Alus masked in primate long-read assemblies. (C) Spider plots of the total number of different Alu-like elements masked in each genome assembly. Lemurs have fewer AluS elements than anthropoid primates. Axis values represent 1,000× events: FAM (Fossil Alu Monomer), FLAM (Free Left Alu Monomers), FRAM (Free Right Alu Monomer), AluJ (oldest), AluS (intermediate), AluY (youngest), Alu (non-specified). LINE: long interspersed nuclear element; LTR: long terminal repeat; SINE: short interspersed nuclear element [31].
variety of repeat elements (Fig. 4). The new long-read–based assembly has 1.31% more interspersed repeats (50.32% vs 49.01%), and a higher percentage of sequence in each repeat subtype, except for satellites, simple repeats, low-complexity elements, and ERV classes I and II. We also observed both a lower percentage of sequence and a smaller number of Alu events in mLemCat1. Additionally, the total number of masked bases is lower in the new assembly, but they represent a higher percentage of the sequence, owing to mLemCat1 having a shorter span.

Alus are the most abundant repeat elements in the human genome, and differences in their rates, distribution, and proliferation could have led to distinct functional changes in multiple primate lineages [34]. Alu elements have been present since the earliest stages of primate evolution, are frequently located in gene-rich regions, and may play an important role in gene regulation [35–38]. To compare the Alu repeat landscape of L. catta with those of other highly contiguous primate assemblies, we ran the same version of RepeatMasker as above adding the -alu option. The genomes used for the comparison were long-read–based assemblies, including human (hg38), chimpanzee (panTro6), western gorilla (gorGor6), Sumatran orangutan (ponAbe3), rhesus macaque (rheMac10), common marmoset (calJac4), and gray mouse lemur (Mmur_3.0) (Supplementary Table S4).

We identified substantially fewer Alu elements in the lemur genomes (L. catta and Microcebus murinus) than those of the catarhines, with the fewest being found in the L. catta genome (3.66% of repeat elements) (Fig. 3B, Supplementary Table S5). In contrast to the other primates assessed, for which AluS elements are most abundant, AluJ is the most common element in mLemCat1 (54.17% of Alu events). Both lemurs have fewer AluS events than the anthropoids and fewer AluY events than the catarhines, consistent with previous reports of the expansion of these 2 families after the Catarrhini-Strepsirrhini split [39]. The fact that the common marmoset has the highest number of AluS elements (Fig. 3C) confirms that the burst that started before the Catarrhini and Platyrrhini parvorders diverged continued with different activity in both lineages after their split. Recent Alu activity (AluY events) is most abundant in catarhines, particularly the rhesus macaque, which, when compared to great apes (Fig. 3C), has a higher overall percentage of Alus (Fig. 3B).

**Conclusion**

We have assembled a new high-quality genome reference for the ring-tailed lemur (L. catta) that satisfies the VGP quality assembly standards. Compared to pre-existing genomic resources, the new assembly has higher contiguity and completeness and contains more single-copy complete BUSCO genes with fewer fragmented or missing genes. Additionally, we analyzed the L. catta repeatome and observed substantially fewer Alu events compared to other
high-quality primate assemblies. This assembly illustrates how long-reads and further scaffolding data such as HiC or optical mappings can drastically improve the contiguity and completeness of an assembly, which also allows for improved analysis of structural variation. We suggest that this new assembly will be an excellent resource for the mammalian genomics community, with particular value for the conservation genomics of lemurs (Fig. 1).

Data Availability

The raw sequencing data and assembly are available via NCBI BioProject: PRJNA562215. mLemCat1 assembly and the raw reads used to generate it can be accessed at GenomeArk [41]. The complete mitogenome of mLemCat1 is available in Genomark as mLemCat1.MT.20190820.fasta.gz [41]. Specific command line parameters are available in Additional File 1. The supporting datasets are available in the GigaScience database (GigaDB) [40].

Additional Files

Additional File 1: Links to the websites with the assembly pipeline specifics used to create Lemur catta (mLemCat1) genome assembly and command lines used to perform the different analyses.

Supplementary Figure S1: Pulse Field Gel assay (Sage Pippin Pulse) used for quality control of the ultra-High Molecular Weight DNA

Supplementary Figure S2: A chromosomal overall synteny plot between Lemur catta (mLemCat1 assembly, vertical axis) and Homo sapiens (hg38 assembly, horizontal axis)

Supplementary Figure S3: Representation of the Lemur catta mitogenome (mLemCat1.MT.20190820)

Supplementary Table S1: Continuity metrics of human (hg38) and Lemur catta (mLemCat1) assemblies

Supplementary Table S2: Comparison of scaffold N50 and assembly size of the latest primate genomes published in GigaScience

Supplementary Table S3: Repetitive elements identified by RepeatMasker in the L. catta genome assembly (mLemCat1)

Supplementary Table S4: Primate assembly lengths and respective quantifications of their Alu content

Supplementary Table S5: Alu-monomers and Alu subfamilies count in each assembly

Abbreviations

bp: base pairs; BUSCO: Benchmarking Universal Single-Copy Orthologs; Gb: gigabase pairs; IUCN: International Union for Conservation of Nature; kb: kilobase pairs; Mb: megabase pairs; NCBI: National Center for Biotechnology Information; PacBio: Pacific Biosciences; SMRT: single-molecule real-time; uHMW: ultra-high molecular weight; VGP: Vertebrate Genomes Project.

Competing Interests

"LFKK is currently an employee of Illumina Inc." All other authors declare no competing interests.

Funding

The project that gave rise to these results received the support of 2 fellowships from “la Caixa” Foundation (ID 100010434) and from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 847648. The fellowship codes are LCF/BQ/BPI20/11760004 and LCF/BQ/DR20/11790032. T.M.B. is supported by funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 864203), BFU2017-86471-P (MINECO/FEDER, UE), “Unidad de Excelencia Maria de Maeztu,” funded by the AEI (CEX2018-000792-M), Howard Hughes International Early Career, NIH 1R01HG010898-01A1, and Secretaria d’Universitats i Recerca and CERCA Programme del Departament d’Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880).

Authors’ Contributions

M.F.P and J.D.O analyzed the data; J.M. and B.H. generated the data; B.H. generated the draft assembly; M.F.B. collected the samples. M.F.P and J.D.O wrote the manuscript with contributions from all authors. T.M.B., E.D.J., O.F., and L.F.K.K. designed the research.

References

1. Estrada, A, Garber, PA, Rylands, AB, et al. Impending extinction crisis of the world’s primates: why primates matter. Sci Adv 2017;3(1):e1600946.
2. Orkin, JD, Kuderna, LFK, Marques-Bonet, T. The diversity of primates: from biomedicine to conservation genomics. Annu Rev Anim Biosci 2021;9(1):103–24.
3. Fleagle, JG. Primate Adaptation and Evolution. Academic; 2013. doi:10.1016/C2009-0-01979-5.
4. Sussman, RW. Demography and social organization of free-ranging Lemur catta in the Beza Mahafaly Reserve, Madagascar. Am J Phys Anthropol 1991;84(1):43–58.
5. Cameron, A, Gould, L. Fragment-adaptive behavioural strategies and intersite variation in the ring-tailed lemur (Lemur catta) in south-central Madagascar. In: LK Marsh, C Chapman, eds. Primates in Fragments: Complexity and Resilience. New York, NY: Springer. 2013:227–43.
6. Chandrashekar, A, Knierim, JA, Khan, S, et al. Genetic population structure of endangered ring-tailed lemurs (Lemur catta) from nine sites in southern Madagascar. Ecol Evol 2020;10(15):8030–43.
7. Sauther, ML, Sussman, RW, Gould, L. The socioecology of the ringtailed lemur: Thirty-five years of research. Evol Anthropol 1999;8(4):120–32.
8. Lafleur, M, Gould, L. Lemur catta. The IUCN Red List of Threatened Species. 2020. doi:10.2305/IUCN.UK.2020-2. RLTS.T11496A115565760.en. https://www.iucnredlist.org/species/11496/115565760. [Accessed 7 March 2022].
9. Lafleur, M, Clarke, TA, Reuter, K, et al. Rapid decrease in populations of wild ring-tailed lemurs (Lemur catta) in Madagascar. Folia Primatol (Basel). 2016;87(5):320–30.
10. Kuderna, LF, Esteller-Cucala, P, Marques-Bonet, T. Branching out: what omics can tell us about primate evolution. Curr Opin Genet Dev 2020;62:65–71.
11. Zoonomia Consortium. A comparative genomics multitool for scientific discovery and conservation. Nature 2020;587(7833):240–5.
12. Rhie, A, McCarthy, SA, Fedrigo, O, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature 2021;592(7856):737–46.
13. Challis, RJ, Kumar, S, Stevens, L, et al. GenomeHubs: simple containerized setup of a custom Ensembl database and web server for any species. Database 2017;2017:doi:10.1093/database/bax039.
14. Chin, C-S, Alexander, DH, Marks, P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 2013;10(6):563–9.
15. Chin, C-S, Peluso, P, Sedlazeck, FJ, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 2016;13(12):1050–4.
16. Guan, D, McCarthy, SA, Wood, J, et al. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 2020;36(9):2896–8.
17. Ning, Z, Harry, E. Scaff10X. 2021. https://github.com/wtsi-hpg/Scaff10X.
18. Bionano Genomics, Inc. Software Downloads. 2021. https://bionanogenomics.com/support/software-downloads/.
19. Ghurye, J, Rhie, A, Walenz, BP, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol 2019;15(8):e1007273.
20. Formenti, G, Rhie, A, Balacco, J, et al. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biol 2021;22(1):120.
21. Bishara, A, Liu, Y, Weng, Z, et al. Read clouds uncover variation in complex regions of the human genome. Genome Res 2015;25(10):1570–80.
22. Garrison, E, Marth, G. Haplotype-based variant detection from short-read sequencing. 2012. arXiv:1207.3907.
23. Gurevich, A, Saveliev, V, Vyahhi, N, et al. Merqury: reference-free quality assessment tool for genome assemblies. Bioinformatics 2013;29(8):1072–5.
24. Cardone, MF, Ventura, M, Tempesta, S, et al. Analysis of chromosome conservation in Lemur catta studied by chromosome paints and BAC/PAC probes. Chromosoma 2002;111(5):348–56.
25. Rhie, A, Walenz, BP, Koren, S, et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 2020;21(1):245.
26. Simão, FA, Waterhouse, RM, Ioannidis, P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015;31(19):3210–2.
27. Roodgar, M, Babveyh, A, Nguyen, LH, et al. Chromosome-level de novo assembly of the pig-tailed macaque genome using linked-read sequencing and HiC proximity scaffolding. Gigascience 2020;9(7):doi:10.1093/gigascience/giaa069.
28. Batra, SS, Levy-Sakin, M, Robinson, J, et al. Accurate assembly of the olive baboon (Papio anubis) genome using long-read and Hi-C data. Gigascience 2020;9(12):doi:10.1093/gigascience/giaa134.
29. Wang, L, Wu, J, Liu, X, et al. A high-quality genome assembly for the endangered golden snub-nosed monkey (Rhinopithecus roxellana). Gigascience 2019;8(8):doi:10.1093/gigascience/giz098.
30. Donath, A, Jühling, F, Al-Arab, M, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res 2019;47(20):10543–52.
31. Conant, GC, Wolfe, KH. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 2008;24(6):861–2.
32. He, Y, Luo, X, Zhou, B, et al. Long-read assembly of the Chinese rhesus macaque genome and identification of ape-specific structural variants. Nat Commun 2019;10(1):4233.
33. Kronenberg, ZN, Fiddes, IT, Gordon, D, et al. High-resolution comparative analysis of great ape genomes. Science 2018;360(6393):eaar6343.
34. Batzer, MA, Deininger, PL, Hellmann-Blumberg, U, et al. Standardized nomenclature for Alu repeats. J Mol Evol 1996;42(1):3–6.
35. Liu, GE, Alkan, C, Jiang, L, et al. Comparative analysis of Alu repeats in primate genomes. Genome Res 2009;19(5):876–85.
36. Urrutia, AO, Ocaña, LB, Hurst, LD. Do Alu repeats drive the evolution of the primate transcriptome? Genome Biol 2008;9(2):R25.
37. Oliver, KR, Greene, WK. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. Mob DNA 2011;2(1):8.
38. Kurosaki, T, Ueda, S, Ishida, T, et al. The unstable CCTG repeat responsible for myotonic dystrophy Type 2 originates from an AluSx element insertion into an early primate genome. PLoS One 2012;7(6):e38379.
39. Konkel, MK, Walker, JA, Batzer, MA. LINEs and SINEs of primate evolution. Evol Anthropol 2010;19(6):236–49.
40. Palmada-Flores, M, Orkin, JD, Haase, B, et al. Supporting data for “A high-quality, long-read genome assembly of the endangered ring-tailed lemur (Lemur catta).” GigaScience Database 2022, http://dx.doi.org/10.5524/102199.
41. GenomeArk, Vertebrate Genomes Project, Lemur catta, ring-tailed lemur, https://vgp.github.io/genomeark/Lemur_catta/. [Accessed 27 July 2021].