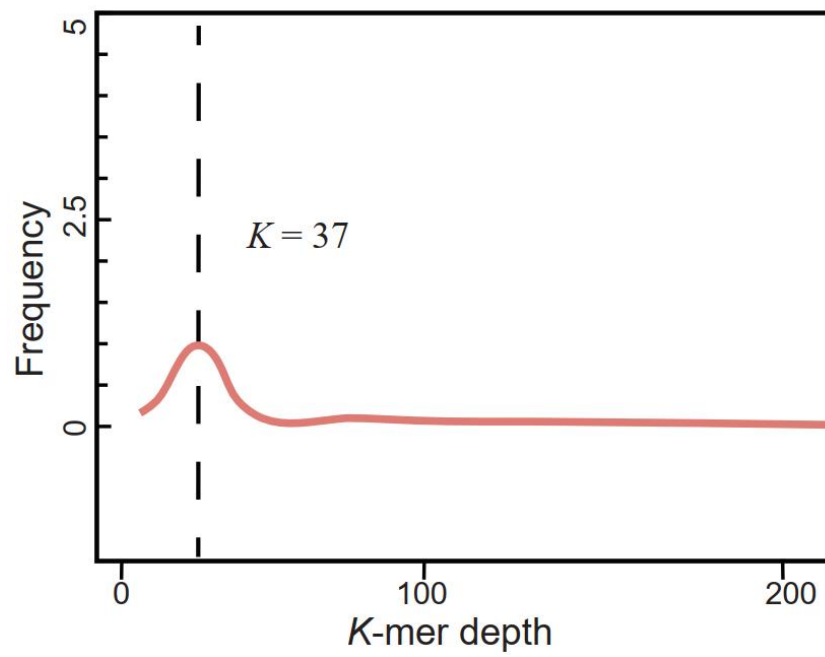
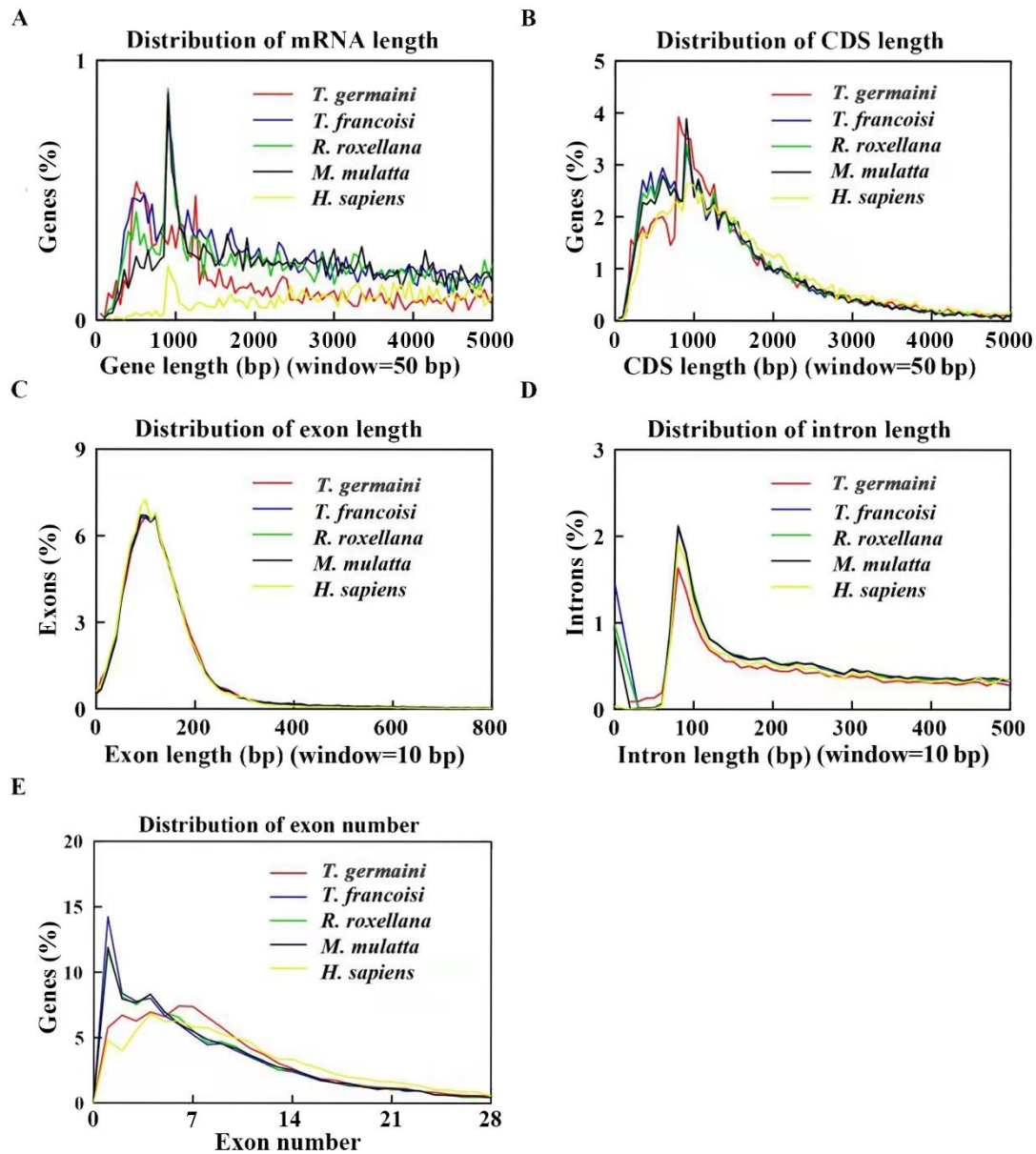


Supplementary Materials



Supplementary Figure S1 *K-mer* frequency distribution in *Trachypithecus germaini*.



Supplementary Figure S2 Comparison of gene features of *T. germaini*, *T. francoisi*, *R. roxellana*, *M. mulatta*, and *H. sapiens* genomes

A. Distribution of mRNA length. **B.** Distribution of CDS length. **C.** Distribution of exon length. **D.** Distribution of intron length. **E.** Distribution of exon number.

Supplementary Table S1 Sequenced reads of *T. germaini* genome

Platform	Read length (bp)	Clean reads		Coverage (X)
		Number	Bases (Gb)	
MGISEQ-2 00 0	150	1 386 970 69 6	208.05	68

Supplementary Table S2 *K-mer* frequency and genome size evaluation of *T. germaini*

k-mer	k-mer_num	pkdepth	genome_size	used_base	depth (X)	heterozygosity (%)
17	112 934 474 246	37	3 052 283 087	137 799 724 500	45.15	0.21

Notes: pkdepth indicates peak depth estimated from k-mer distribution.

Supplementary Table S3 Comparison of reference genomes among four primates

Species	Genome size (Gb)	N50 (Mb)	BUSCO (%)	Gene number	Reference
<i>T. germaini</i>	2.91	55.9*	94.7	20 332	This study
<i>T. francoisi</i>	2.87	16.3*	91.5	25 421	(Liu et al., 2020)
<i>R. roxellana</i>	3.04	5.72*/144.56§	94.0	22 497	(Wang et al., 2019)
<i>M. mulatta</i>	2.95	8.19*/13.64§	92.7	20 389	(He et al., 2019)

Notes: * represents the contig N50; § represents the scaffold N50

Supplementary Table S4 Mapping ratio of short reads mapped to assembled genome

Library	Insert size (bp)	Total reads	Mapped reads	Mapped ratio (%)	PE mapped reads	PE mapped ratio (%)
Clean_69A	350	1 389 249 035	1 384 525 981	99.66	1 360 346 090	98.15

Notes: PE indicates paired-end.

Supplementary Table S5 Evaluation of *T. germaini* genome assembly completeness by Benchmarking Universal Sing-Copy Orthologs (BUSCO) with mammalia_odb9 dataset

Category	Number	Ratio (%)
Complete BUSCOs (C)	3 884	94.7
Complete and single-copy BUSCOs (S)	3 820	93.1
Complete and duplicated BUSCOs (D)	64	1.6
Fragmented BUSCOs (F)	93	2.3
Missing BUSCOs (M)	127	3.0
Total BUSCO groups searched	4 104	

Supplementary Table S6 Repeats in assembled *T. germaini* genome

Type	Repeat region size	% in the genome
Trf	85 013 697	2.91
Repeatmasker	982 893 830	33.70
Proteinmask	375 224 754	12.86
De novo	1 196 635 527	41.02
Total	1 441 938 057	49.43

Supplementary Table S7 Transposable elements (TEs) in assembled *T. germaini* genome

	Repbased TEs	TE proteins	De novo	Combined TEs	
Type	Length (bp)	Length (bp)	Length (bp)	Length (bp)	% in genome
DNA	39 835 178	12 919 408	64 009 607	76 042 883	2.61
LINE	485 842 042	316 204 234	622 951 892	670 889 784	22.99
SINE	313 675 838	0	333 558 428	401 074 689	13.75
LTR	146 023 877	46 206 963	191 936 232	224 237 659	7.69
Other	80	0	0	80	0.000003
Unknown	0	0	7 485 885	7 485 885	0.25
Total	982 893 830	375 224 754	1 170 494 674	1 388 090 070	47.59

Supplementary Table S8 Predicted protein-coding genes in *T. germaini* genome.

	Gene Set	Gene Number	Average mRNA length (bp)	Average cds length (bp)	Average exon length (bp)	Average exon number	Average intron length (bp)
<i>De novo</i>	Augustus	19 445	56 794.41	1 394.79	168.47	8.27	7 610.94
	Genescan	19 280	68 662.86	1 988.68	170.98	11.63	6 271.72
	<i>T. francoisi</i>	18 658	38 269.46	1 522.01	177.14	8.59	4 840.37
Homolog	<i>R. roxellana</i>	18 909	23 550.72	1 296.76	204.2	6.35	4 159.42
	<i>M. mulatta</i>	24 030	24 625.21	1 344.47	202.44	6.64	4 126.97
	<i>H. sapiens</i>	14 307	23 602.73	1 335.81	194.51	6.86	3 794.83
EVM		20 332	69 942.61	1 767.06	164.56	10.73	6 998.22

Supplementary Table S9 Evaluation on completeness of predicted protein-coding genes in *T. germaini* genome by BUSCO with mammalia_odb9 dataset

Category	Number	Ratio (%)
Complete BUSCOs (C)	3857	93.9
Complete and single-copy BUSCOs (S)	3810	92.8
Complete and duplicated BUSCOs (D)	47	1.1
Fragmented BUSCOs (F)	120	2.9
Missing BUSCOs (M)	127	3.2
Total BUSCO groups searched	4104	

Supplementary Table S10 Gene annotation in assembled *T. germaini* genome

	Number	Percent(%)
Total	20 332	
InterPro	14 814	73.16
GO	10 870	53.68
KEGG	12 776	63.09
Swissprot	15 981	78.92
TrEMBL	16 211	80.06
NR	16 627	82.11
Unannotated	3 576	17.66
Annotated	16 756	82.41

Supplementary Table S11 Annotated non-coding RNAs in *T. germaini* genome

Type	Category	Copy(w)	Average length (bp)	Total length (bp)	% of genome
miRNA		547	90.15	49 311	0.00169
tRNA		365	75.14	27 425	0.00094
rRNA	rRNA	237	174.84	41 437	0.0014
	18S	26	233	6058	0.00021
	28S	188	181.20	34 066	0.0012
	5.8S	6	120.66	724	0.000025
	5S	17	34.64	589	0.000020
snRNA	snRNA	2 095	111.28	233 142	0.0080
	CD-box	690	101.39	69 963	0.0024
	HACA-box	274	135.66	37 171	0.0013
	splicing	945	118.68	112 161	0.0038

REFERENCES

- He YX, Luo X, Zhou B, Hu T, Meng XY, Audano PA, et al. 2019. Long-read assembly of the Chinese rhesus macaque genome and identification of ape-specific structural variants. *Nature Communications*, **10**(1): 4233.
- Liu ZJ, Zhang LY, Yan ZZ, Ren ZJ, Han FM, Tan XX, et al. 2020. Genomic mechanisms of physiological and morphological adaptations of limestone langurs to karst habitats. *Molecular Biology and Evolution*, **37**(4): 952–968.
- Wang L, Wu JW, Liu XM, Di DD, Liang YH, Feng YF, et al. 2019. A high-quality genome assembly for the endangered golden snub-nosed monkey (*Rhinopithecus roxellana*). *GigaScience*, **8**(8): giz098.