Journal of Heredity, 2023, **114**, 271–278
https://doi.org/10.1093/jhered/esad015
Advance access publication 3 March 2023
Genome Resources

American Genetic Association

OXFORD

Genome Resources

A chromosome-scale high-contiguity genome assembly of the cheetah (*Acinonyx jubatus*)

Sven Winter[1,*], René Meißner[1], Carola Greve[2], Alexander Ben Hamadou[2], Petr Horin[3,4], Stefan Prost[5,6,7,8], Pamela A. Burger[1]

[1]Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Vienna, Austria,
[2]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany,
[3]Department of Animal Genetics, University of Veterinary Sciences, Brno, Czech Republic,
[4]Central European Institute of Technology, University of Veterinary Sciences Brno (CEITEC Vetuni), Brno, Czech Republic,
[5]Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland,
[6]Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Vienna, Austria,
[7]Natural History Museum Vienna, Central Research Laboratories, Vienna, Austria,
[8]South African National Biodiversity Institute, National Zoological Garden, Pretoria, South Africa

[*]Corresponding authors: sven.winter@vetmeduni.ac.at (S.W.), pamela.burger@vetmeduni.ac.at (P.B.)

Corresponding Editor: Klaus-Peter Koepfli

Abstract

The cheetah (*Acinonyx jubatus*, SCHREBER 1775) is a large felid and is considered the fastest land animal. Historically, it inhabited open grassland across Africa, the Arabian Peninsula, and southwestern Asia; however, only small and fragmented populations remain today. Here, we present a de novo genome assembly of the cheetah based on PacBio continuous long reads and Hi-C proximity ligation data. The final assembly (*VMU_Ajub_asm_v1.0*) has a total length of 2.38 Gb, of which 99.7% are anchored into the expected 19 chromosome-scale scaffolds. The contig and scaffold N50 values of 96.8 Mb and 144.4 Mb, respectively, a BUSCO completeness of 95.4% and a k-mer completeness of 98.4%, emphasize the high quality of the assembly. Furthermore, annotation of the assembly identified 23,622 genes and a repeat content of 40.4%. This new highly contiguous and chromosome-scale assembly will greatly benefit conservation and evolutionary genomic analyses and will be a valuable resource, e.g., to gain a detailed understanding of the function and diversity of immune response genes in felids.

Key words: conservation genomics, Felidae, Hi-C, PacBio, proximity-ligation.

Introduction

The cheetah (*Acinonyx jubatus*, SCHREBER 1775) is a large carnivore of the cat family Felidae, in which it forms the tribe Acinonychini together with the puma (*Puma concolor*), and the jaguarundi (*Herpailurus yagouaroundi*) (Durant *et al.* 2021). The cheetah is known as the fastest land animal, as it reaches speeds of up to 105 km/h (Sharp 1997). Historically, it occurred in open grasslands across Africa, the Arabian Peninsula, and southwestern Asia (Durant *et al.* 2017). At present, it only inhabits small fractions of its former range resulting in small and fragmented populations (Durant *et al.* 2017). The cheetah, as a species, is currently considered "vulnerable" on the International Union for Conservation of Nature (IUCN) Red List of threatened species, with two subspecies *A. j. venaticus* (Iran) and *A. j. hecki* (Northwest Africa), being listed as "critically endangered" (Belbachir 2008; Durant *et al.* 2021; Farhadinia *et al.* 2017). The use and need for genomic analyses to support conservation decisions and management has increasingly been recognized, e.g., by the IUCN Cat Specialist Group (IUCN Cat Specialist Group 2021). Recently, conservation and evolutionary genomic analyses of the cheetah based on short-read sequences have been published (Dobrynin *et al.* 2015; Prost *et al.* 2022). However, in-depth conservation or evolutionary genomic analyses, such as the inference of runs of homozygosity (ROH) or analyses of mutational load and genetic health, greatly benefit from highly continuous reference genomes. As such, highly continuous reference genomes provide information aside from commonly used single nucleotide polymorphism or microsatellite data to evaluate a threatened species' fitness (Wold *et al.* 2021) and inbreeding status (Humble *et al.* 2022) and play crucial roles in the development of management actions in conservation (Brandies *et al.* 2019). Especially ROH analyses benefit from highly continuous reference genomes. Since such a reference genome is currently unavailable for the cheetah, we sequenced and assembled a chromosome-level genome for this threatened species, with a much-improved continuity than previously reported genome assemblies.

Received December 6, 2022; Accepted February 28, 2023

Material and Methods

Sequencing and assembly

High molecular weight genomic DNA was extracted from the blood of a 14-year-old female cheetah named "Pintada" (GAN:27869175) from Lisbon Zoo, Portugal, using the PacBio Nanobind CBB kit (PacBio, Menlo Park, CA, USA). The blood was drawn during routine veterinary procedures and immediately frozen at −20 °C. DNA concentration and molecule length were evaluated using the Qubit dsDNA BR Assay kit on the Qubit Fluorometer (Thermo Fisher Scientific) and the Genomic DNA Screen Tape on the Agilent 2200 TapeStation system (Agilent Technologies), respectively.

Two sequencing libraries were prepared, one PacBio continuous long read (CLR) library using the SMRTbell Express Prep Kit 2.0 (PacBio) and one short read library using the NEBNext Ultra II FS DNA Library Kit for Illumina (New England BioLabs Inc., Ipswich, MA, USA). The long-read library was then sequenced on the PacBio Sequel IIe system in CLR mode using the Sequel II Binding Kit 2.2 (PacBio). The short-read library with an insert size of 350 bp was sequenced on the NovaSeq6000 platform (Illumina, Inc., San Diego, CA, USA), generating 150 bp paired-end reads.

After receiving the data from the Sequel IIe run, we converted the PacBio subreads from BAM to FASTQ format using BAM2fastx v.1.3.0, a PacBio Secondary Analysis Tool (https://github.com/PacificBiosciences/pbbioconda; see Table 1 for a list of software used in this study). We then used Flye v.2.9 (RRID: SCR_017016) (Kolmogorov *et al.* 2019) to assemble the reads into a contig-level genome assembly. Flye was run with the options for raw PacBio reads and the default parameters, including one iteration of long-read polishing.

The short-read data were first trimmed using fastp v.0.20.1 (RRID: SCR_016962) (Chen *et al.* 2018) with base correction and low complexity filter enabled to remove sequencing adaptors and polyG stretches at the end of reads. We also employed a 4 bp sliding window to detect regions of poor quality (Phred score <15). Reads were removed if they fit into one of the following categories: reads below 36 bp length, reads with >40% low-quality bases, or reads with 5 or more undetermined bases (Ns). The trimmed reads were then mapped to the assembly using bwa-mem v.0.7.17 (RRID: SCR_010910) (Li 2013). The resulting mapping file was then sorted by position in the assembly, converted to BAM format, and indexed using samtools v.1.9 (RRID: SCR_002105) (Li *et al.* 2009). The mapped short reads were used to further improve the base-level accuracy of the assembly with one iteration of short-read polishing with pilon v.1.23 (RRID: SCR_014731) (Walker *et al.* 2014).

To anchor the polished contigs into chromosome-scale scaffolds, we utilized previously generated proximity ligation (Hi-C) data for the same subspecies from the DNAZoo (www.dnazoo.org, accession numbers: SRR8616936, SRR8616937). First, we followed the Arima Hi-C mapping

Table 1. Software and versions used to generate the *Acinonyx jubatus* assembly.

Pipeline Step	Software	Version
PacBio BAM to FASTQ	BAM2fastx	v.1.3.0
Short-read trimming and filtering	fastp	v.0.20.1
assembly & long-read polishing	Flye	v.2.9
Mapping of short reads	BWA-MEM	v.0.7.17
Mapping of long reads	minimap2	v.2.17
Sort & index BAM	samtools	v.1.9
Short-read polishing	pilon	v.1.23
Hi-C data processing for scaffolding	Arima Genomics mapping pipeline	Commit 2e74ea4
Scaffolding	YaHS	v.1.1
Hi-C contact map generation	JuicerTools	v.1.22.01
Manual editing of Hi-C contact map	Juicebox	v.1.11.08
Generate long-read subsets	seqtk	v.1.3
Gap-closing	TGS-GapCloser	v.1.1.1
Assembly statistics	Quast	v.5.0.2
Gene set completeness	BUSCO	v.5.3.1
Mapping statistics	QualiMap	v.2.2.1
Assembly completeness	Merqury	v.1.1
Contamination analyses	BLASTN	v.2.11.0+
	BlobToolKit	v.3.2.4
Mitochondrial genome assembly	GetOrganelle	v.1.7.5
Repeat library	RepeatModeler	v.2.0.1
Repeat masking	RepeatMasker	v.4.1.0
Homology-based gene prediction	GeMoMa pipeline	v.1.7.1
Functional annotation	BLASTP	v.2.11.0+
	InterProScan	v.5.50.84
Synteny analyses	JupiterPlot	v.3.8.1

pipeline used by the Vertebrate Genome Project (https://github.com/VGP/vgp-assembly/blob/master/pipeline/salsa/arima_mapping_pipeline.sh) for filtering and mapping of the data to the assembly. In short, the pipeline mapped the reads to the assembly using bwa-mem v.0.7.17 (RRID: SCR_010910) (Li 2013) and filtered the mapped reads with samtools v.1.14 (RRID: SCR_002105) (Li *et al.* 2009) based on multiple parameters such as mapping quality, read quality, and CIGAR strings. Subsequently, duplicated reads were marked and removed using the Picard v. 2.26.10 (RRID: SCR_006525) (Broad Institute 2019) tool "MarkDuplicates". The mapped and filtered reads were then used in YaHS v.1.1 (RRID: SCR_022965) (Zhou *et al.* 2022) for proximity-ligation-based scaffolding. Hi-C contact maps were generated with JuicerTools v.1.22.01 (RRID: SCR_017226) (Durand et al. 2016) and used for manual curation of the scaffolded assembly in Juicebox v.1.11.08 (RRID: SCR_021172) (Durand et al. 2016). Furthermore, we run TGS-GapCloser v.1.1.1 (RRID: SCR_017633) (Xu *et al.* 2020) for two iterations to close gaps in the assembly and increase its contiguity. Each iteration of gap-closing utilized a random subset of approximately 25% of the long-read data to reduce computational requirements, as well as the short reads for polishing. The long-read subsets were generated from the complete dataset with seqtk v. 1.3 (RRID: SCR_018927) (Li 2018b) using the random number generator seeds 11 and 18.

To evaluate the quality and completeness of the assembly, we generated assembly statistics with Quast v.5.0.2 (RRID: SCR_001228) (Gurevich *et al.* 2013), ran a gene set completeness analysis with BUSCO v.5.3.1 (RRID: SCR_015008) (Manni *et al.* 2021) using the *carnivora_odb10* dataset and compared the results to the previously available chromosome-scale assembly from DNAZoo (Dobrynin *et al.* 2015; Dudchenko *et al.* 2017), which is based on an earlier draft genome (GCA_001443585.1; Dobrynin *et al.* 2015), and the currently best reference assembly *Aci_jub_2* from GenBank (GCA_003709585.1). We also evaluated the mapping rate of both the short and long reads with QualiMap v.2.2.1 (RRID: SCR_001209) (Okonechnikov *et al.* 2016) after mapping the reads to our assembly with bwa-men v.0.7.17 (RRID: SCR_010910) (Li 2013) and minimap2 v.2.17 (RRID: SCR_018550) (Li 2018a), respectively. In addition, we analyzed the completeness, base-level error rate, and quality value (QV) of the assembly based on a k-mer size of 21 using Merqury v.1.1 (RRID: SCR_022964) (Rhie *et al.* 2020). We further evaluated potential contamination with BlobToolKit v.3.2.4 (RRID: SCR_017618) (Laetsch and Blaxter, 2017) utilizing the generated mapping files and a BLASTN v.2.11.0+ (RRID: SCR_001598) (Zhang *et al.* 2000) search against the NCBI Nucleotide database.

In addition to the nuclear genome, we also assembled the mitochondrial genome from the short reads with GetOrganelle v1.7.5(RRID: SCR_022963) (Jin *et al.* 2020).

Annotation

For increased accuracy during gene prediction, repeat regions in the assembly were first masked. RepeatModeler v. 2.0.1 (RRID: SCR_015027) (Flynn *et al.* 2020) was used to generate a de novo repeat library, which was then combined with the Felidae repeat dataset (July 2022) from RepBase (RRID: SCR_021169) (Bao *et al.* 2015). This custom repeat library was then used to annotate and mask the repeats in the genome using RepeatMasker v.4.1.0 (http://www.repeatmasker.org/RMDownload.html,RRID: SCR_012954). We hard-masked all interspersed repeats and soft-masked simple repeats.

Genes were predicted using the homology-based gene prediction with MMseqs2 (RRID: SCR_022962) (Steinegger and Söding 2017) as an alignment tool implemented in the GeMoMa pipeline v.1.7.1 (RRID: SCR_017646) (Keilwagen et al. 2016, 2018). We used the following nine mammalian genomes and associated annotations as references: *Homo sapiens* (GCF_000001405.40), *Mus musculus* (GCF_000001635.27), *Lynx canadensis* (GCF_007474595.2), *Canis lupus familiaris* (GCF_014441545.1), *Prionailuris bengalensis* (GCF_016509475.1), *Leopardus geoffroyi* (GCF_018350155.1), *Felis catus* (GCF_018350175.1), *Panthera tigris* (GCF_018350195.1), and *Panthera leo* (GCF_018350215.1).

Functional annotation of the predicted proteins was conducted by a BLASTP v.2.11.0+ (RRID: SCR_001010) (Zhang *et al.* 2000) search with an e-value cutoff of 10^{-6} against the Swiss-Prot database (RRID: SCR_002380; release 2021-02). Furthermore, we annotated gene ontology (GO) terms, domains, and motifs using InterProScan v.5.50.84 (RRID: SCR_005829) (Jones *et al.* 2014; Quevillon *et al.* 2005). The completeness of the predicted proteins was evaluated with BUSCO v.5.3.1 (RRID: SCR_015008) (Manni *et al.* 2021) using the *carnivora_odb10* dataset.

Synteny between feline genomes

We analyzed synteny between *VMU_Ajub_asm_v1.0*, the previously published cheetah assembly *Aci_jub_2* (GCA_003709585.1), the chromosome-scale cheetah assembly from DNAZoo (Dobrynin *et al.* 2015; Dudchenko *et al.* 2017), the tiger *Panthera tigris* (GCF_018350195.1), two assemblies of the domestic cat *Felis catus* (Fca126: GCF_018350175.1, Bredemeyer *et al.* 2021; Fca9.1: GCA_000181335.5), the leopard cat *Prionailurus bengalensis* (GCA_016509475.2; Bredemeyer *et al.* 2021), and the Canada lynx *Lynx canadensis* (GCA_007474595.2; Rhie *et al.* 2021) using JupiterPlot v.3.8.1(RRID: SCR_022961) (Chu 2018). The closely related leopard cat was used as a reference to identify homologous cheetah chromosomes, as the chromosome structure based on G-banding is very similar to the cheetah (O'Brien *et al.* 2006).

Results and discussion

Genome sequencing and assembly

Sequencing generated 341 Gb of long-read PacBio data or approximately 136-fold coverage with a mean subread length of 10,308.5 bp and approximately 18-fold (45.5 Gb) of short-read Illumina data.

After assembly with Flye, polishing with pilon, proximity-ligation scaffolding with YaHS (Fig. 1A), and two iterations of gap-closing, the final assembly (*VMU_Ajub_asm_v1.0*) had a total length of 2.38 Gb in 198 scaffolds (including the mitochondrial genome) and a scaffold and contig N50 of 144.4 Mb and 96.8 Mb, respectively (Table 2A, Fig. 1C). The largest 19 scaffolds (>40 Mb), representing the expected haploid chromosome number of the cheetah (2n = 38) (O'Brien *et*

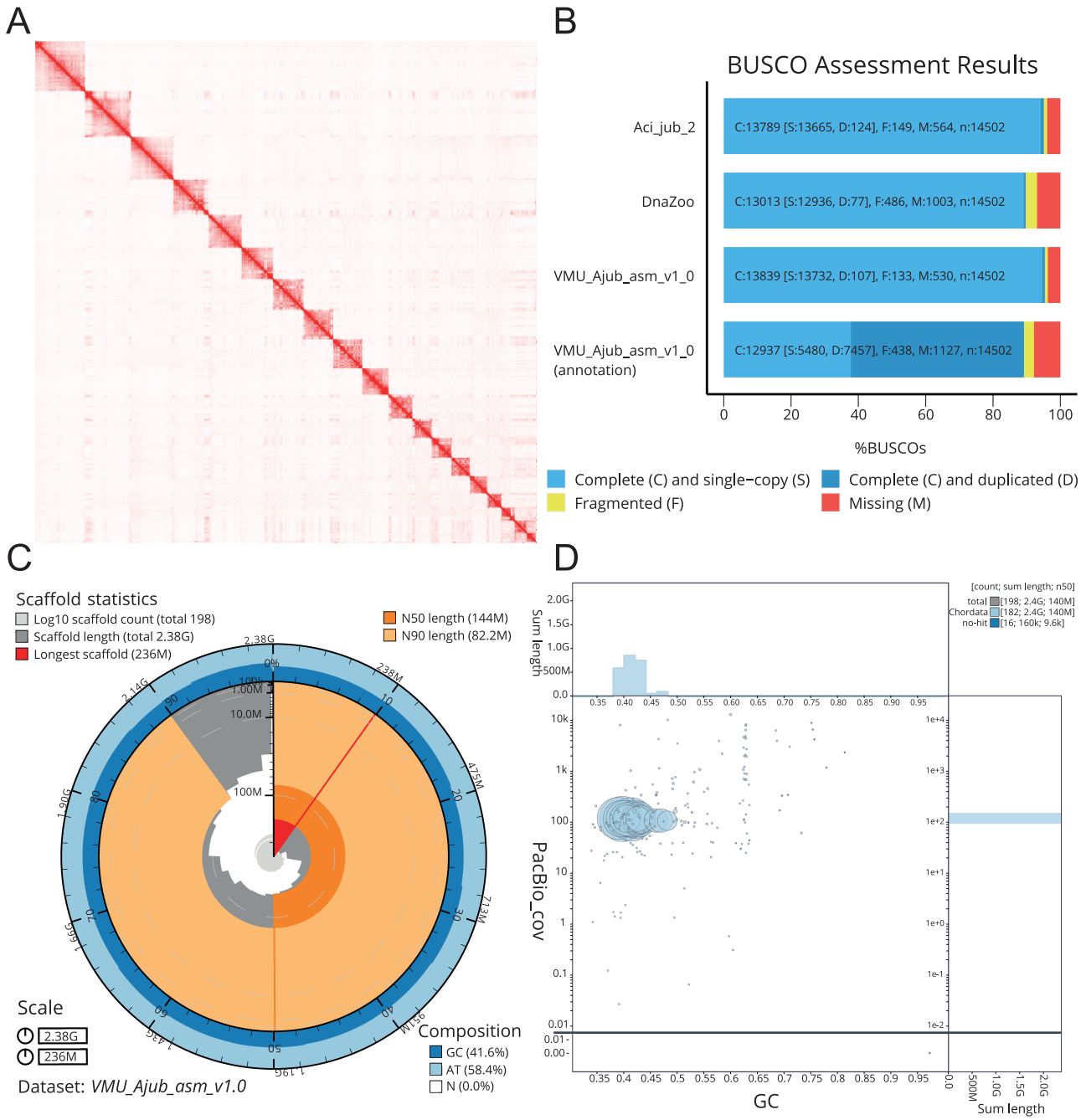

Fig. 1. Assembly quality assessment of *VMU_Ajub_asm_v1.0*. A) Hi-C contact density map depicting the 19 distinct chromosome-level scaffolds. B) BUSCO gene set completeness analyses for the assembly, annotation (predicted proteins), and previously available assemblies (Aci_jub_2/ DNAZoo) for comparison. C) SnailPlot summarizing assembly statistics. D) BlobPlot analysis comparing GC content (*x* axis), sequencing depth of PacBio reads (*y* axis), and taxonomic assignment of contigs (colors) show no evidence of contamination.

al. 2006), span 99.7% of the total assembly length, resulting in a scaffold L50 of seven. *VMU_Ajub_asm_v1.0* is highly contiguous and reflects a major improvement in contiguity compared with the previously available cheetah genome assembly *Aci_jub_2* and the chromosome-scale one from DNA Zoo (Dobrynin *et al.* 2015; Dudchenko *et al.* 2017), as evidenced by a 569-fold and 3,007-fold larger contig NG50 (96.8 Mb vs. 170 kb vs. 32.2 kb), respectively (Table 2A). The high quality and completeness of *VMU_Ajub_asm_v1.0* were also highlighted by a BUSCO completeness score of 95.4%, an increase of 0.3% and 5.7% compared with the previously available assemblies (Fig. 1B), and Merqury k-mer-based completeness of 98.4%, with an error rate of 0.013% (QV = 38.7). Furthermore, no evidence of contamination was visible in a BlobPlot (Fig. 1D).

Annotation

Repeat annotation.

A repeat content of 40.4% or 960.5 Mb of the sequence of VMU_Ajub_asm_v1.0 was identified by the repeat annotation (Table 2B). Long Interspersed Nuclear Elements (LINEs) were the most abundant, spanning nearly one-quarter (24.7%) of the genome, followed by Short Interspersed Nuclear Elements

Table 2. Assembly statistics of *VMU_Ajub_asm_v1.0* in comparison to the previously available cheetah assemblies *Aci_jub_2* and DNAZoo (A) and repeat content of *VMU_Ajub_asm_v1.0* (B)

A

	Scaffold-level			Contig-level		
	VMU_Ajub_asm_v1.0	***Aci_jub_2***	**DNAZoo**	***VMU_Ajub_asm_v1.0***[a]	***Aci_jub_2***[a]	**DNAZoo**[a]
No. of Scaffolds/contigs	198	3,220	13,047	220	27,346	163,014
No. of Scaffolds/contigs (>1 KBP)	197	3,220	2,247	219	26,646	130,184
L50	7	15	7	9	3,937	19,040
LG50[b]	7	16	7	9	4,307	21,646
N50 (BP)	144,444,042	48,500,042	144,637,309	96,827,784	179,924	35,115
NG50[b] (BP)	144,444,042	47,062,725	144,637,309	96,827,784	170,063	32,192
Max. Scaffold/contig length (BP)	235,669,126	120,246,179	235,519,777	218,150,482	1,559,821	419,587
Total length (BP)	2,377,450,114	2,384,851,327	2,373,338,770	2,377,445,714	2,374,148,075	2,328,406,444
GC (%)	41.59	41.55	41.29	41.59	41.55	41.29
No. of N's	4,400	10,703,252	42,858,800	0	0	128,625
No. of N's per 100 KBP	0.19	448.8	1807.45	0	0	5.52

B

Type of element	Number of elements	Length (bp)	Percentage of assembly
Sines	887,856	129,629,255	5.45%
Lines:	1,755,353	586,984,696	24.69%
L1/Line1	1,558,262	535,696,713	22.53%
LTR **elements**	371,149	128,445,131	5.40%
DNA **transposons**	384,830	73,338,402	3.08%
Unclassified	10,953	2,596,934	0.11%
Small RNA	612,496	89,904,390	3.78%
Satellites	3,697	921,773	0.04%
Simple repeats	748,674	32,528,191	1.37%
Low complexity	91,753	5,217,093	0.22%
TOTAL	4,868,672	960,526,883	40.40%

[a]Broken into contigs at gaps with a length of ≥10 N's. Statistics for these columns are based on contigs, the remaining columns are based on Scaffolds.
[b]Based on an estimated reference length of 2,503,680,000 BP calculated from a C-value of 2.56 PG (genomesize.com)

(SINEs) and Long Terminal Repeat (LTR) Elements with 5.45% and 5.4%, respectively. The remaining repeat classes, such as DNA Transposons, small RNA, simple repeats, etc., each accounted for less than 4%.

Gene annotation.

The homology-based gene prediction with GeMoMa identified 23,622 genes in *VMU_Ajub_asm_v1.0* with a median gene length of 7,857.5 bp spanning 468.4 Mb of the total assembly length. A BUSCO score of 89.2% of identified complete Carnivora orthologous genes suggest high annotation completeness (Fig. 1B). InterProScan functionally annotated 66,775 out of the 67,405 predicted proteins (99.1%) and assigned at least one Gene Ontology (GO) term to 50,735 proteins (75.3%). In addition, more than 96.9% (65,333) of the predicted proteins were identified from the Swiss-Prot database.

Synteny between feline genomes

JupiterPlots showed high levels of synteny between *VMU_Ajub_asm_v1.0* and other felid species, as expected by the identical chromosome numbers (2n = 38) and the conserved nature of Felidae genomes (O'Brien *et al.* 2006) (Fig. 2). Therefore, we were able to assign chromosome names to the 19 chromosome-scale scaffolds of *VMU_Ajub_asm_v1.0* according to the karyotype format commonly used for felids (O'Brien *et al.* 2006). We based the naming on the homologous regions of the leopard cat genome (Fig. 2C), whose chromosomes have very similar G-banding patterns as the cheetah chromosomes (O'Brien *et al.* 2006; Wurster-Hill and Gray 1973). Comparing *VMU_Ajub_asm_v1.0* and the previous DNAZoo assembly (Fig. 2A) found structural differences only in the form of one translocation in chromosome D2 (scaffold 14) and an inversion in the smallest chromosome E4 (scaffold 19). However, both differences could potentially be scaffolding or assembly errors in one of the assemblies, despite both utilizing Hi-C data for scaffolding from the same individual. We found even fewer differences comparing *VMU_Ajub_asm_v1.0* with *Aci_jub_2* (Supplementary Figure 1), which was expected, as *Aci_jub_2* is not chromosome-scale allowing for smaller scaffolds or contigs to be placed in syntenic positions. The most

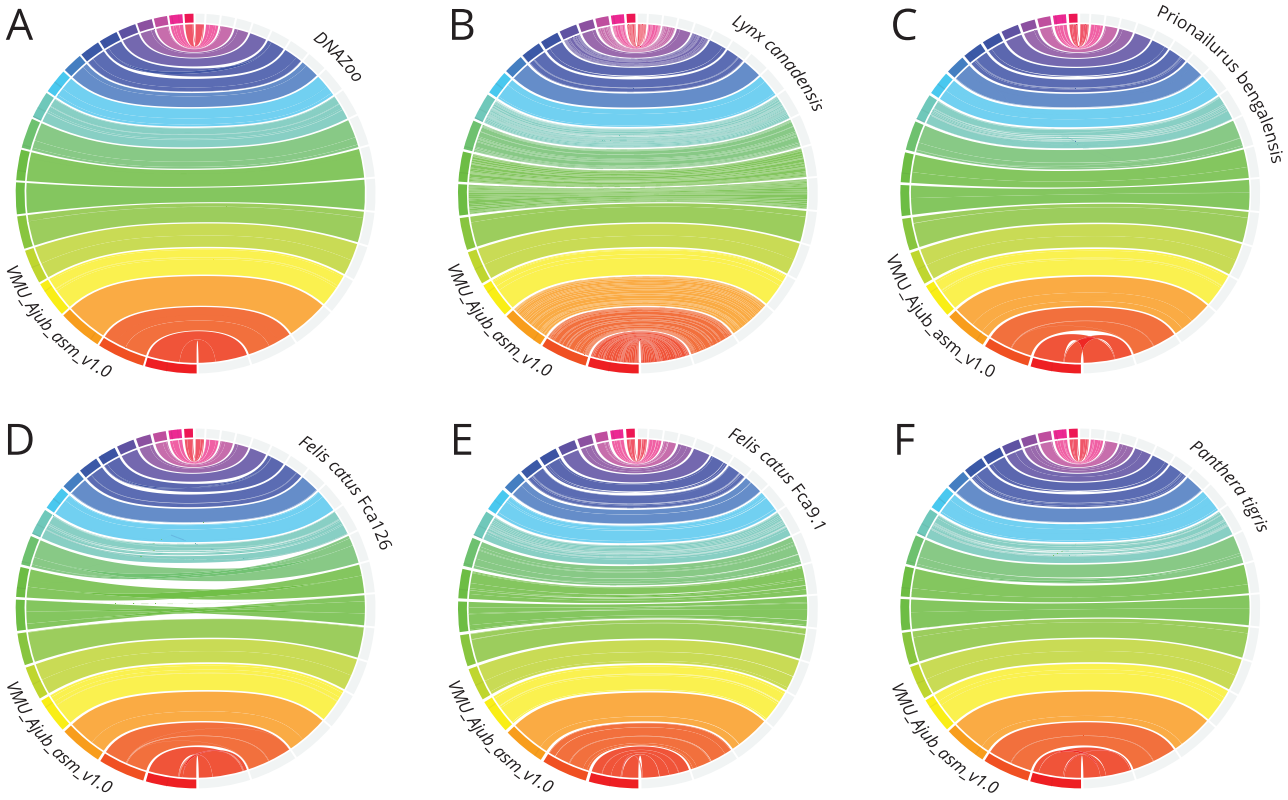

Fig. 2. Synteny between the cheetah *Acinonyx jubatus* and other felid species. Circos plots generated with JupiterPlot comparing the synteny of the chromosome-scale cheetah genome assembly *VMU_Ajub_asm_v1.0* (A-F, left) with six available chromosome-scale assemblies of other felids (right): A) A previous cheetah assembly from DNAZoo, B) the Canada lynx *Lynx canadensis*, C) the leopard cat *Prionailurus bengalensis*, D) & E) the domestic cat *Felis catus* (Fca126, Fca9.1), and F) the tiger *Panthera tigris*. Ribbons between scaffolds indicate syntenic regions. Chromosome-scale scaffolds are sorted by size from the largest (bottom) to the smallest (top).

structural differences are evident between *VMU_Ajub_asm_v1.0* and the most recent domestic cat assembly (Fca126, GCF_018350175.1, Fig. 2D). Yet, when compared with the previous cat assembly (Fca9.1, GCA_000181335.5, Fig. 2E), only very small rearrangements were identified, suggesting potential assembly errors in the latest cat assembly.

Conclusion

Highly contiguous annotated chromosome-scale genome assemblies are valuable references for evolutionary or conservation genomic analyses and enable in-depth studies on structural variation or the diversity and function of certain genes (e.g., immune response genes). However, genome assemblies from nonmodel organisms of this quality are still relatively rare. The presented new cheetah assembly *VMU_Ajub_asm_v1.0*, which is the first long-read-based assembly for this species, has a much-improved contiguity and will thus enable more in-depth genomic analyses for this threatened species. This genome resource provides a solid foundation to address key biological questions like understanding the process of natural selection and adaptation.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

This study was funded by the Central European Science Partnership (CEUS) project Austrian Science Fund (FWF) I5081-B/ GACRCzech Republic 21-28637L (to P.H. and P.B.).

Acknowledgments

We thank the Genome Technology Center (RGTC) at Radboudumc for the use of the Sequencing Core Facility (Nijmegen, The Netherlands), which provided the PacBio SMRT sequencing service on the Sequel IIe platform. We also thank Rui Bernardino from the zoo of Lisbon for the cheetah sample.

Conflict of Interest

None declared.

Data Availability

All underlying read data and the assembly are available at GenBank under BioProject PRJNA854353. A detailed list of commands used to generate the presented assembly and related analyses are available as Supplementary Material 2. The annotation, assembly, repeat masked assemblies, and all commands are also available at Dryad (https://doi.org/10.5061/dryad.xksn02vkr).

References

BaoW, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 2015:6(1):11. doi:10.1186/s13100-015-0041-9.

Belbachir F. Acinonyx jubatus ssp. Hecki. The IUCN Red List of Threatened Species 2008: E.T221A13035738. IUCN; 2008. http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T221A13035738.en.

Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. *Genes* 2019:10(11):11. doi:10.3390/genes10110846.

Bredemeyer KR, Harris AJ, Li G, Zhao L, Foley NM, Roelke-Parker M, O'Brien SJ, Lyons LA, Warren WC, Murphy WJ. Ultracontinuous single haplotype genome assemblies for the domestic cat (Felis catus) and Asian Leopard Cat (Prionailurus bengalensis). *J Hered*. 2021:112(2):165–173. doi:10.1093/jhered/esaa057.

Broad Institute. Picard toolkit. Broad Institute; 2019. [accessed 2022 Aug 10]. http://broadinstitute.github.io/picard/.

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018:34(17):i884–i890. doi:10.1093/bioinformatics/bty560.

Chu, J. Jupiter Plot: a Circos-based tool to visualize genome assembly consistency. 2018. doi:10.5281/zenodo.1241235.

Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, Kliver S, Schmidt-Küntzel A, Koepfli K-P, Johnson W, et al. Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biol*. 2015:16(1):277. doi:10.1186/s13059-015-0837-4.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017:356(6333):92–95. doi:10.1126/science.aal3327.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*. 2016:3(1):95–98. doi:10.1016/j.cels.2016.07.002.

Durant SM, Mitchell N, Groom R, Pettorelli N, Ipavec A, Jacobson AP, Woodroffe R, Böhm M, Hunter LTB, Becker MS, et al. The global decline of cheetah Acinonyx jubatus and what it means for conservation. *Proc Natl Acad Sci USA*. 2017:114(3):528–533. doi:10.1073/pnas.1611122114.

Durant SM, Groom R, Ipavec A, Mitchell N, Khalatbari L. *IUCN red list of threatened species: acinonyx jubatus. IUCN Red List of Threatened Species*. IUCN; 2021. doi:10.2305/IUCN.UK.2022-1.RLTS.T219A124366642.en.

Farhadinia MS, Hunter LTB, Jourabchian A, Hosseini-Zavarei F, Akbari H, Ziaie H, Schaller GB, Jowkar H. The critically endangered Asiatic cheetah Acinonyx jubatus venaticus in Iran: a review of recent distribution, and conservation status. *Biodivers Conserv*. 2017:26(5):1027–1046. doi:10.1007/s10531-017-1298-8.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020:117(17):9451–9457. doi:10.1073/pnas.1921046117.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013:29(8):1072–1075. doi:10.1093/bioinformatics/btt086.

Humble E, Stoffel MA, Dicks K, Ball AD, Gooley RM, Chuven J, Pusey R, Remeithi M, Koepfli K-P, Pukazhenthi B, et al. *Conservation management strategy impacts inbreeding and genetic load in scimitar-horned oryx* (p. 2022.06.19.496717). *bioRxiv*. 2022: doi:10.1101/2022.06.19.496717.

IUCN Cat Specialist Group. Conservation of the Cheetah Acinonyx Jubatus in Asia and North-Eastern Africa. 5th Meeting of the Sessional Committee of the CMS Scientific Council (ScC-SC5). (2021). https://www.cms.int/dugong/sites/default/files/document/cms_scc-sc5_inf.8_conservation-of%20the-cheetah-in-asia-north-eastern-africa_e.pdf.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020:21(1):241. doi:10.1186/s13059-020-02154-5.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014:30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res*. 2016:44(9):e89–e89. doi:10.1093/nar/gkw092.

Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf*. 2018:19(1). doi:10.1186/s12859-018-2203-5.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019:37(5):540–546. doi:10.1038/s41587-019-0072-8.

Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Research*. 2017:6:1287. doi:10.12688/f1000research.12232.1.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv:1303.3997*. 2013 doi:10.48550/arXiv.1303.3997.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018a:34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

Li, H. seqtk: Toolkit for processing sequences in FASTA/Q formats. 2018b. https://github.com/lh3/seqtk.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009:25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

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. *Mol Biol Evol*.2021:38(10):4647–4654. doi:10.1093/molbev/msab199.

O'Brien SJ, Menninger JC., Nash WG. *Atlas of mammalian chromosomes*. John Wiley &Sons; 2006.

Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016:32(2):292–294. doi:10.1093/bioinformatics/btv566.

Prost S, MachadoAP, Zumbroich J, Preier L, Mahtani-Williams S, Meissner R, Guschanski K, Brealey JC, Fernandes CR, Vercammen P, et al. Genomic analyses show extremely perilous conservation status of African and Asiatic cheetahs (Acinonyx jubatus). *Mol Ecol*. 2022:31(16):4208–4223. doi:10.1111/mec.16577.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: Protein domains identifier. *Nucleic Acids Res*. 2005:33(suppl_2):W116–W120. doi:10.1093/nar/gki442.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020:21(1):245. doi:10.1186/s13059-020-02134-9.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021:592(7856):737–746. doi:10.1038/s41586-021-03451-0.

Sharp NCC. Timed running speed of a cheetah (Acinonyx jubatus). *J Zool*. 1997:241(3):493–494. doi:10.1111/j.1469-7998.1997.tb04840.x.

Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017:35(11):1026–1028. doi:10.1038/nbt.3988.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and

genome assembly improvement. *PLoS One*. 2014:9(11):e112963. doi:10.1371/journal.pone.0112963.

Wold J, Koepfli KP, Galla SJ, Eccles D, Hogg CJ, Le Lec MF, Guhlin J, Santure AW, Steeves TE. Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern. *Mol Ecol*. 2021:30(23):5949–5965. doi:10.1111/mec.16141.

Wurster-Hill DH, Gray CW. Giemsa banding patterns in the chromosomes of twelve species of cats (Felidae). *Cytogenet Genome Res*. 1973:12(6):377–397. doi:10.1159/000130481.

Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, Fan G, Liu X, Xu X, Deng L, et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience*. 2020:9(giaa094). doi:10.1093/gigascience/giaa094.

Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol*. 2000:7(1–2):203–214. doi:10.1089/10665270050081478.

Zhou C, McCarthy SA, Durbin R. YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics*. 2022:39(1):btac808. doi:10.1093/bioinformatics/btac808.