Scientific Data

https://doi.org/10.1038/s41597-026-06773-8

Article in Press

Chromosome-level genome assembly and annotation of the critically endangered Siberian crane (*Leucogeranus leucogeranus*)

Received: 19 March 2025

Accepted: 31 January 2026

Cite this article as: Chen, Q., Zheng, C., Huang, P. *et al.* Chromosome-level genome assembly and annotation of the critically endangered Siberian crane (*Leucogeranus leucogeranus*). *Sci Data* (2026). https://doi.org/10.1038/s41597-026-06773-8

Qing Chen, Chenqing Zheng, Peng Huang, Nianhua Dai, Marria Vladimirtseva, Wenjuan Wang & Yang Liu

Chromosome-level genome assembly and annotation of the critically endangered Siberian crane (*Leucogeranus leucogeranus*)

Qing Chen[a,b], Chenqing Zheng[c], Peng Huang[d], Nianhua Dai[e], Marria Vladimirtseva[f], Wenjuan Wang[g, h*], Yang Liu[a*]

a State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Shenzhen, 518000, China
b Key Laboratory of Poyang Lake Environment and Resource Utilization, Ministry of Education, Center for Watershed Ecology, School of Life Science, Nanchang University, Nanchang, 330031, China
c Health data science center, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, 518020, China
d Wildlife Conservation Center of Jiangxi Province, Nanchang, 330038, China
e Institute of Biological Resources, Jiangxi Academy of Sciences, Nanchang, Jiangxi, 330095, China
f Institute for Biological Problems of Cryolitozone, Siberian Branch of Russian Academy for Science; Federal State Budgetary Institution National Park Lena Pillars, Sakha Republic, 630090, Russia
g School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, 100083, China
h Center for East Asian–Australasian Flyway Studies, Beijing Forestry University, Beijing, 100083, China

Abstract

The Siberian crane (*Leucogeranus leucogeranus*) is classified as Critically Endangered by the IUCN. Its current estimated population is over 6,900 individuals in East Asia, whereas the Western/Central Asian population is nearly extinct, with no recent records of its presence in the wild. Here, we present a high-quality, chromosome-level genome assembly of the Siberian crane generated by integrating Nanopore long-read data, MGISEQ-2000 short-read data, and Hi-C technology data. The assembled genome spans 1.31 Gb, with a scaffold N50 of 83.45 Mb, comprising 33 chromosomes and additional unplaced scaffolds. BUSCO assessment indicated that 97.3 percent of genes in the genome assembly are complete. We identified 10.9 percent repetitive sequences and 21,678 protein-coding genes, of which 88 percent were successfully assigned functional annotations. This high-quality genome assembly and annotation provide a valuable genomic resource for comparative genomic research aimed at understanding the ecology, evolutionary adaptations, and development of Gruidae birds.

Background & Summary

The Siberian crane (*Leucogeranus leucogeranus*) is one of the most endangered birds among the 15 crane species in the family Gruidae, order Gruiformes, and serves as a flagship species of the wetland ecosystem. This species has experienced a significant decline over the past century due to habitat loss and degradation, human disturbance, and hunting. It has been classified as Critically Endangered by the International Union for Conservation of Nature (IUCN) since 2000[1]. Two geographically isolated populations of the Siberian crane exist. The

Western/Central Asian population breeds in the central part and lower sections of the Ob River in Western Siberia, Russia, and migrates to wintering grounds at the Caspian Sea in Iran and Keoladeo National Park in India. This population has no recent confirmed reports of its presence in the wild[2]. The East Asian population primarily breeds in the northeastern Siberian tundra of Russia and winters at Poyang Lake in China. This population has shown gradual recovery due to widespread public concern and the implementation of conservation actions over the past 50 years[2]. According to recent population censuses in 2024, the East Asian population has reached more than 6,900 individuals (https://i.ifeng.com/c/8fYQVTguDew).

Eleven out of the 15 existing crane species are designated as threatened. These species exhibit distinctive plumage characteristics and morphological variations, with their distribution ranges showing significant diversity[2]. Genome resources provide a key foundation for exploring adaptations and evolutionary history and for serving as scientific indicators for conservation efforts[3,4]. A chromosome-level genome offers complete and accurate information of an organism's genetic material, enabling research to address questions such as: 1) How do chromosomal structural variations drive speciation and adaptation? 2) What genetic mechanisms determine characteristics evolution among members in the family Gruidae (e.g., the red bare crown skin of the red-crowned crane (*Grus japonensis*) or the crown feathers of the gray-crowned crane (*Balearica regulorum*))? 3) How did the sex chromosomes of Gruidae evolve? For example, by comparing the chromosome-level genomes of crane species, we can investigate the potential genetic basis of cranes' unique phenotypes through gene family and selective sweep analyses.

Furthermore, comparing the genomes of Gruidae species with those of non-Gruidae species within the order Gruiformes allows exploration of the genetic mechanisms underlying distinct morphological differences, such as variations in body size and limb structure. The karyotype of Gruiformes is $2n = 80$[5]. Currently, whole-genome assemblies are publicly available for eight crane species, but only three have been assembled to the chromosomal level. The first contig-level genome of the Siberian crane contained 1740 contigs with a contig N50 of 21.54 Mb, based on long-read and short-read sequencing[6]. However, the completeness of this genome assembly can be improved. By integrating Hi-C technology to reconstruct the sequence of an entire chromosome or chromosome arm, the contiguity and completeness of assembly fragments can be significantly enhanced[7].

In this study, we used chromosome conformation capture (Hi-C) technology combined with MGISEQ-2000 sequencing (short reads) and Oxford Nanopore sequencing (long reads) to build a chromosome-level reference genome of the Siberian crane. This approach significantly improves the consistency and completeness of the genome assembly. The final assembly is 1.31 Gb and consists of 32 autosomes, one Z chromosome, and 1403 unmounted scaffolds. Thirty-three chromosomes account for 1.19 Gb and 91.14% of the assembled sequences. The contig N50 is 21.54 Mb, with 94.7% completeness according to the Benchmarking Universal Single-Copy Ortholog (BUSCO) assessment. The scaffold N50 is 83.45 Mb, with 97.6% completeness. Additionally, we identified 21,678 protein-coding genes based on ab initio and homology predictions. The BUSCO assessment indicated 97.3% completeness for the annotated genes. Our study provides an essential genomic resource for future conservation efforts, evolutionary adaptation studies, and comparative analyses within the family Gruidae.

Methods

Sampling and sequencing

We obtained a fresh blood sample from a rescued male Siberian crane and fresh muscle tissue from a deceased male individual in Poyang Lake. Genomic DNA from the blood sample was obtained using a standard chloroform extraction for the long-read Oxford Nanopore PromethION platform. The blood sample was treated with lysis buffer, SDS, and proteinase K. After centrifugation, the supernatant was extracted twice with chloroform. DNA was precipitated overnight at −20°C using prechilled isopropanol and sodium acetate. The precipitate was centrifuged, washed with ethanol, dried, and dissolved in EB buffer. The long DNA segments were quantified and assessed for integrity. We also used the protocol of the DNeasy Blood & Tissue Kit (Qiagen, USA) to extract DNA for MGISEQ-2000 sequencing.

A nanopore library was prepared using a ligation sequencing kit (Oxford Nanopore Technologies, New York, NY, USA; SQK-LSK109) following the manufacturer's instructions. Before constructing libraries for MGISEQ-2000 sequencing, we assessed DNA quality using Qubit and gel electrophoresis and fragmented the DNA using a Covaris machine. We constructed paired-end small-fragment libraries with a 300–500 bp insert length using a PCR reaction system. A total of 121.08 Gb of clean long reads with 101× coverage and 123.08 Gb of clean short reads with 102× coverage were generated.

To generate a chromosome-scale genome assembly, we used a muscle sample for Hi-C sequencing. The tissue was first cross-linked for 10 min with 2% formaldehyde, and then glycine was used to terminate the cross-linking reaction. The purified DNA was cut with a restriction enzyme, followed by end repair using DNA polymerase and labeling with biotin-14-dCTP. The ligated DNA was sheared into ~300 bp fragments and purified by biotin-mediated pulldown. Finally, the Hi-C library was sequenced on the MGISEQ-2000 platform. We obtained a total of 133.1 Gb of clean data with 111× coverage after filtering low-quality reads and adaptor sequences using SOAPnuke v1.5.6.[8] (Table 1).

Genome assembly

We estimated genome size and heterozygosity using Jellyfish[9] with K-mer analysis (k = 17) and GenomeScope[10] with MGISEQ-2000 short reads. The estimated genome size and heterozygosity were 1.26 Gb and 0.22%, respectively (Fig. 1).

We assembled a chromosome-level genome by integrating long reads, short reads, and Hi-C reads. First, we assembled initial contigs using Canu v1.9[11] with the long reads, then corrected the assembly twice using Racon v1.4.13[12] and Medaka v1.6.0. The contig assemblies were further polished in one round based on the short reads using Pilon v1.4[13]. The draft assembly contained 1,740 contigs with an N50 of 21.54 Mb.

Hi-C reads were mapped to the draft genome using the Burrows-Wheeler Aligner (BWA) v0.7.15[14], and the mapping results were filtered using HiC-Pro v2.8.0[15] with default parameters. The 3D-DNA pipeline[16] was applied to correct and cluster the initially assembled contigs into scaffolds. Contig orientation was validated using JUICEBOX v1.8.8[17], and ambiguous fragments were manually corrected when significant contact frequency bands were interrupted by a green square. We obtained 1,436 scaffolds with an N50 of 83.4 Mb (Fig.

2; Table 2).

A total of 91.14% of the assembly was assigned to 33 chromosomes based on the Hi-C results (Fig. 3). For the final chromosome-level genome assembly, BUSCO[18] was used to assess completeness with aves_odb10. The results showed 98.9% complete BUSCO genes, 0.43% fragmented genes, and 1.06% missing genes (Fig. 2). The Quality Value (QV) was 37.6 estimated by mapping short reads to assembly using merqury[19].

We performed whole-genome synteny analysis among the common crane *Grus grus* (GCA_964106855.1), the whooping crane *Grus americana* (GCF_028858705.1), the East African grey crowned-crane *B. r. gibbericeps* (GCA_011004875.1), and the Siberian crane using NGenomeSyn v1.41[20] with mapping by Minimap2[21]. We removed the fragment with less than 50 kb. The sex chromosome (Z) was identified based on chromosome synteny. The analysis showed high consistency among the four crane genomes (Fig. 4; Table 3).

Annotation

Repetitive elements in the assembled genome were identified using two different strategies: homology-based and ab initio prediction. RepeatMasker v4.0.9[22] was applied to detect homologous repeat sequences by searching the Repbase library. The EDTA[23] annotation pipeline was used for TE prediction. The EDTA pipeline incorporates LTRharvest, the parallel version of LTR_FINDER, LTR_retriever, GRF, TIR-Learner, HelitronScanner, and RepeatModeler, as well as customized filtering scripts (https://github.com/oushujun/EDTA). A total of 142 Mb of repetitive sequences were identified, accounting for 10.90% of the assembled genome. DNA transposons, short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and long terminal repeat (LTR) retrotransposons accounted for 2.03%, 0.08%, 5.94%, and 2.88% of the assembled genome, respectively (Table 4).

Ab initio and homology-based approaches were used to detect gene structure from the repeat-masked genome generated during repetitive element prediction. For ab initio annotation, Fgenesh v1.6[24] and Augustus v3.3.3[25] in Braker were used to predict coding genes. For homology-based annotation, protein sequences from chicken (*Gallus gallus*), Magellanic penguin (*Spheniscus magellanicus*), hoatzin (*Opisthocomus hoazin*), crested ibis (*Nipponia nippon*), black-necked crane (*Grus nigricollis*), common cuckoo (*Cuculus canorus*), killdeer (*Charadrius vociferus*), and limpkin (*Aramus guarauna*) were aligned with the genome using Miniport v1.1[26].

The predicted data sets were combined into a nonredundant gene set using Maker v3.01.03[27], and completeness was assessed using BUSCO v5.6.1[28] with aves_odb10. A total of 21,678 protein-coding genes were detected. BUSCO assessment showed 97.3% complete BUSCO genes, 0.9% fragmented genes, and 1.2% missing genes. The predicted genes had an average length of 28,389.5 bp. Additionally, 218,615 exons were predicted, with an average length of 168.68 bp (Table 4). Compared with the three closely related species already published[27], our annotation results were more complete (Table 5).

Functional annotation of the predicted genes was performed by aligning them to the KEGG (Kyoto Encyclopedia of Genes and Genomes)[29], SwissProt, and UniProt databases[30] using BLAST+ v2.12.0[31]. Protein families, motifs, and domains were annotated using InterProScan v5.56. Respectively, 87% of genes matched UniProt[32], 81%

matched SwissProt, 75% matched KEGG, 80% matched terms from InterProScan, and 76% matched Gene Ontology (GO)[33] terms (Table 6). A total of 21,678 genes (88%) were successfully functionally annotated[39].

We also predicted noncoding RNA genes. Transfer RNAs (tRNAs) were predicted using tRNAscan-SE v2.0.11[34]. Ribosomal RNAs (rRNAs) were predicted using Barrnap v0.9 with default parameters (https://github.com/tseemann/barrnap). Small nuclear RNAs (snRNAs) and microRNAs (miRNAs) were identified by alignment to the Rfam database[35] and annotated using Infernal v1.1.4[36]. We identified 248 miRNAs (0.0015% of the genome), 1,643 tRNAs (0.0095%), 165 rRNAs (0.0045%), 648 snRNAs (0.0058%), and 24 lncRNAs (0.0004%) (Table 7).

Data Records

The Hic data are deposited into NCBI Sequence Read Archive database with accession number SRR35316027[37]. The sequencing data obtained from the MGISEQ-2000 platforms are deposited into NCBI Sequence Read Archive database with accession number SRR35316036-42[37]. The genome assembly is deposited into the DDBJ/ENA/GenBank with accession number JBQWBR000000000[38]. The annotation files are available from the Figshare repository[39]. The repository includes the assembled genome (FASTA format), gene annotation files (GFF3 format), repeat annotation data, and sequencing reads (FASTQ format) used for the assembly and validation.

Technical Validation

The assembled genome length was 1.31 Gb, close to the estimated size from K-mer analysis (Fig. 1) and within the normal range of avian genomes (0.9–2.1 Gb). A total of 33 scaffolds (>1 Mb) with a scaffold N50 of 87.89 Mb were assembled. The Hi-C heatmap exhibits a clear interaction pattern in 33 chromosome pairs (Fig. 5). BUSCO evaluation supported the final assembled results with a high proportion of completeness (Fig. 2).

We also used the Burrows-Wheeler Aligner (BWA) v0.7.18[14] to map short reads[6] to the reference genome to assess assembly completeness. Bamdest was used to calculate alignment coverage and mapping rate, which were approximately 99.98% and 99.64% (Table 8). This result indicates that the reads were consistently aligned with the assembled genome. Compared with published crane genomes, our genome length and scaffold N50 were similar (Table 9).

Data Availability

The Hi-C data described in this study are available at in the NCBI Sequence Read Archive database with accession number SRR35316027 (https://www.ncbi.nlm.nih.gov/sra/SRP618574). The sequencing data obtained from the MGISEQ-2000 platforms are deposited into NCBI Sequence Read Archive database with accession number SRR35316036-42 (https://www.ncbi.nlm.nih.gov/sra/SRP618574). The genome assembly is deposited into the DDBJ/ENA/GenBank with accession number JBQWBR000000000 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_053455625.1). The annotation files are available from Figshare (https://doi.org/10.6084/m9.figshare.30017956.v1). All data are publicly available and includes raw sequencing reads, assembled genome, genome annotation files, functional annotation results. Metadata describing the sample information, sequencing platforms, and assembly statistics are also provided in the same

repository.

Code Availability

The assembly and annotation were performed following the manuals of the corresponding bioinformatics tools with default parameters. The code of the quality assessment and result visualization is available at https://github.com/ChenqCQ/Siberian_crane_Chromosome.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32160132, 32471732) and the Forestry Administration of Guangdong Province, China (DFGP Project of Fauna of Guangdong-202115; Science and Technology Planning Projects of Guangdong Province-2021B1212110002). We appreciate the technical support from the Beijing Genomics Institute (BGI) and EasyATCG Science and Technology Company for sequencing, assembly, and annotation. We appreciate Dr. Russell Doughty for his substantial contributions to improving the readability of this manuscript.

Contributions

Yang Liu and Wenjuan Wang conceived and designed the experiments. Peng Huang, Nianhua Dai, and Marria Vladimirtseva collected the samples. Qing Chen performed quality assessment and analyzed the data. Qing Chen wrote the manuscript. Chenqing Zheng, Yang Liu, and Wenjuan Wang reviewed the manuscript.

Author information

Authors and Affiliations

State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Shenzhen, 518000, China
Qing Chen & Yang Liu
Key Laboratory of Poyang Lake Environment and Resource Utilization, Ministry of Education, Center for Watershed Ecology, School of Life Science, Nanchang University, Nanchang, 330031, China
Qing Chen
Health data science center, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, 518020, China
Chenqing Zheng
Wildlife Conservation Center of Jiangxi Province, Nanchang, 330038, China
Peng Huang
Institute of Biological Resources, Jiangxi Academy of Sciences, Nanchang, Jiangxi, 330095, China
Nianhua Dai
Institute for Biological Problems of Cryolitozone, Siberian Branch of Russian Academy for Science; Federal State Budgetary Institution National Park Lena Pillars, Sakha Republic, 630090, Russia
Marria Vladimirtseva
School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, 100083, China
Wenjuan Wang
Center for East Asian–Australasian Flyway Studies, Beijing Forestry University, Beijing, 100083, China

Wenjuan Wang

Corresponding authors
Correspondence to Yang Liu or Wenjuan Wang.

State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Shenzhen, 518000, China
Yang Liu
Email: liuy353@mail.sysu.edu.cn
Tel: 18819800772

School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, 100083, China
Wenjuan Wang
Email: wangwj870120@163.com
Tel:17326808607

Ethics declarations

Competing interests

The authors declare no competing interests.

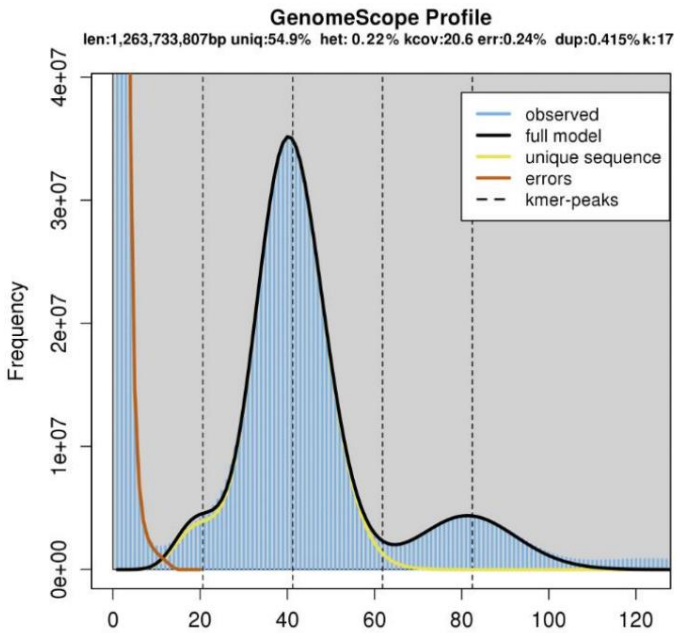

Fig.1 The genome size and heterozygosity evaluation with K-mer 17.

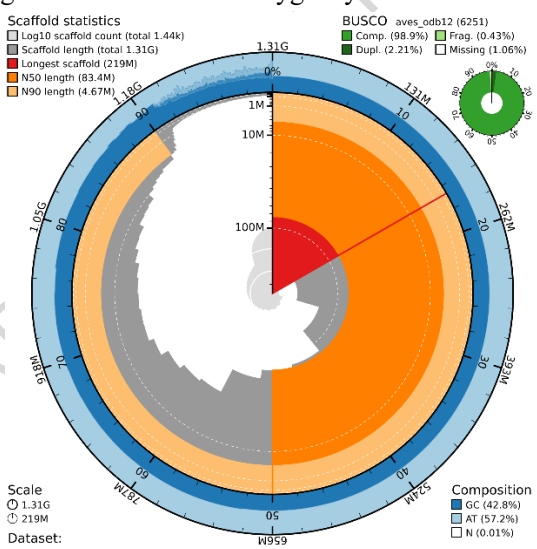

Fig.2 Assessment of the Siberian crane genome assembly. The scaffold statistics, genome scale, BUSCO assessment, and composition are exhibited.

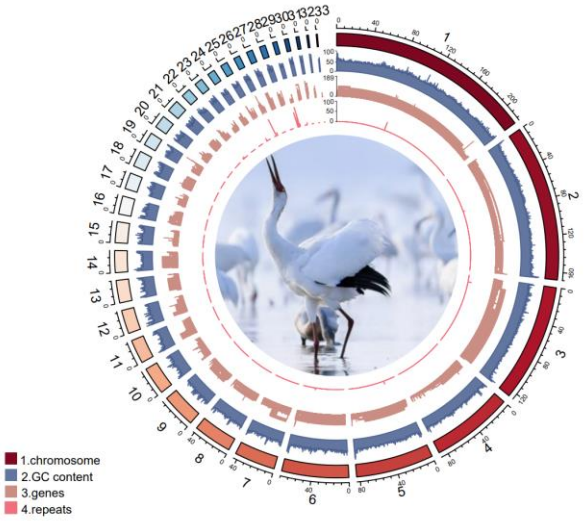

Fig. 3 Characterization of the Siberian crane genome assembly. The chromosome length is presented in Mb. The genes, repeats, and GC contents are shown with window size 100 kb. The Siberian crane image was taken by Haiyan Zhou.

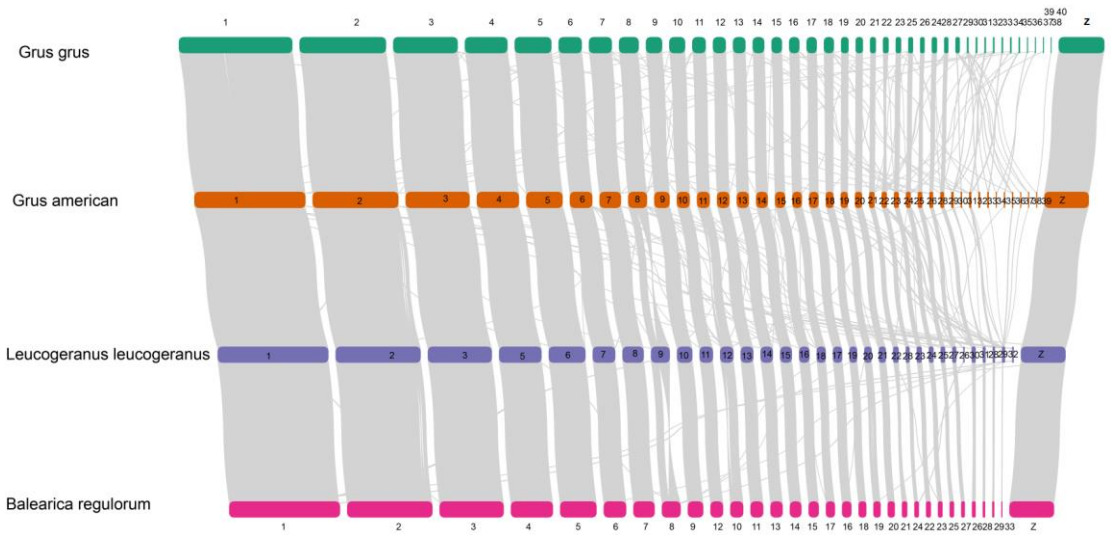

Fig. 4 The synteny blocks among the common crane, whooping crane, East African grey crowned-crane, and Siberian crane.

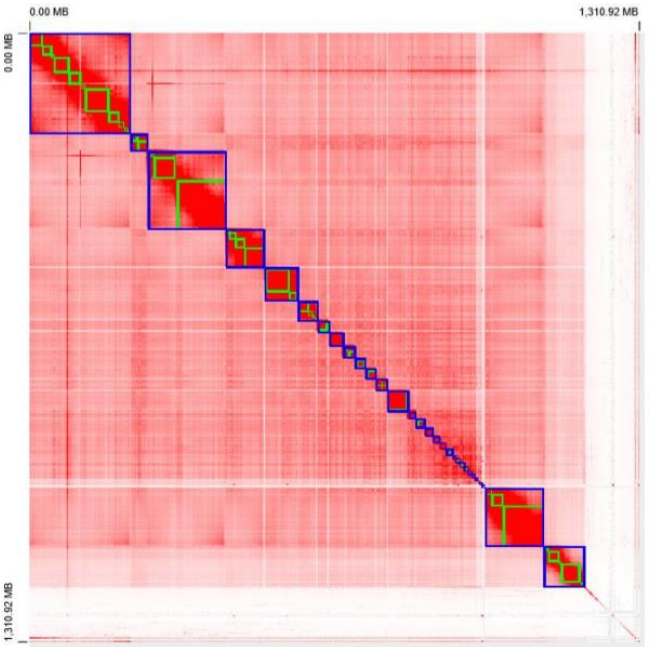

Fig.5 Hi-C interactions result for the assembled chromosomes. The coordinates show the length of the genome. The deeper colors on the heatmap indicate a stronger interaction between the respective genomic regions. The green rectangles represent the contig sequence, and the blue rectangles indicate the chromosome range defined by the clustering results based on the heat map.

Table 1 Sequencing statistics of the Siberian crane genome assembly.

Sequencing Strategy	Sequencing platform	Library size (bp)	Total data (Gb)	Sequence coverage (×)
Nanopore	PromethION	20000	121.08	101
Short reads	MGISEQ-2000	300-500	123.08	102
Hi-C	MGISEQ-2000	350	133.1	111
total	-	-	377.26	314

Table 2 Statistics of the Siberian crane genome assembly.

Assembly features	Size
Total size	1,311,077,122
Number of chromosomes	33
Number of scaffolds	1,436
Scaffold N50	83,445,140
Scaffold L50	5
Scaffold N90	4,672,211
Scaffold L90	29
Number of contigs	1,740
Contig N50	21,544,053
Contig L50	15

Contig N90	319,934
Contig L90	144
GC percent	42.41%

Table 3 Statistics of chromosome length of the Siberian crane genome.

Scaffold name	Chromosome	Length (bp)
HiC_scaffold_1	Chr1	219,277,396
HiC_scaffold_3	Chr2	168,079,306
HiC_scaffold_32	Chr3	126,409,554
HiC_scaffold_33	Chr4(Z)	87,888,569
HiC_scaffold_4	Chr5	83,445,140
HiC_scaffold_5	Chr6	72,065,773
HiC_scaffold_13	Chr7	43,918,989
HiC_scaffold_6	Chr8	41,356,376
HiC_scaffold_2	Chr9	36,456,921
HiC_scaffold_8	Chr10	29,585,496
HiC_scaffold_7	Chr11	25,634,983
HiC_scaffold_9	Chr12	25,388,104
HiC_scaffold_12	Chr13	24,246,241
HiC_scaffold_10	Chr14	23,588,028
HiC_scaffold_11	Chr15	22,454,793
HiC_scaffold_15	Chr16	19,611,771
HiC_scaffold_16	Chr17	17,464,057
HiC_scaffold_14	Chr18	17,015,953
HiC_scaffold_19	Chr19	14,194,114
HiC_scaffold_17	Chr20	13,666,851
HiC_scaffold_18	Chr21	12,996,071
HiC_scaffold_22	Chr22	9,965,744
HiC_scaffold_23	Chr23	8,921,811
HiC_scaffold_21	Chr24	8,090,433
HiC_scaffold_25	Chr25	7,729,171
HiC_scaffold_24	Chr26	6,752,316
HiC_scaffold_20	Chr27	6,661,960
HiC_scaffold_26	Chr28	6,437,487
HiC_scaffold_30	Chr29	4,672,211
HiC_scaffold_29	Chr30	4,364,670
HiC_scaffold_28	Chr31	3,298,599
HiC_scaffold_31	Chr32	1,975,651
HiC_scaffold_27	Chr33	1,327,317

Table 4 Statistics of the Siberian crane genome annotation.

Type	Length (bp)	Proportion (%)
Repeat elements		
DNA	26,598,152	2.03
SINEs	1,099,472	0.08
LINEs	72,004,847	5.49
LTR	37,745,429	2.88
other	221,302	0.02
Total	142,914,880	10.90

Gene structure	
Average gene length	28389.5 bp
Average exon length	168.68 bp
Exon number	218,615
Average intron length	2792.46 bp
Intron number	206,685

CDS: coding region sequences; LINE: long interspersed nuclear elements; SINE: short interspersed nuclear elements; LTR: long terminal repeat

Table 5 Statistics of the Siberian crane genome annotation compared with other related genomes.

	Leucogeranus leucogeranus	*Charadrius vociferus*	*Balearica regulorum*	*Chlamydotis macqueenii*
Repeat elements (% of the genome)				
LINE	5.49	4.53	3.35	3.97
SINE	0.08	0.13	0.14	0.17
LTR	2.88	1.12	1.51	1.40
DNA	2.03	0.20	0.24	0.23
Genome structure				
No. of predicted protein-coding gene	21,678	16,856	14,173	13,582
Average CDS length (kp)	28.4	19.1	13.8	12.9
Average exon length (bp)	168.68	161.8	162.7	162.9

Table 6 The functional annotation of protein-coding genes of the Siberian crane.

Database	Number	Percentage (%)
Total	21,678	100
UniProt (2024.1.15)	18,994	87
Swiss-Prot (2024.1.15)	17,625	81
KEGG (76)	16,467	75
InterProScan (5.54)	17,524	80
GO (5.54)	16,521	76
No assign	2,634	12

Table 7 The annotation of non-coding RNA genes of the Siberian crane.

Type	Copy number	Total length (bp)	Percentage (%)
miRNA	248	19,415	0.0015

tRNA		1,643	124,648	0.0095
rRNA	rRNA	83	21,003	0.0016
	18S	2	2,331	0.0002
	28S	6	10,182	0.0008
	5.8S	1	152	0.0000
	5S	73	8,337	0.0006
snRNA	snRNA	325	37,580	0.0029
	CD-box	172	15,585	0.0012
	HACA-box	80	11,285	0.0009
	scaRNA	15	2,801	0.0002
	splicing	56	7,495	0.0006
ribozyme		9	1,184	0.0001
lncRNA		24	4,727	0.0004

Table 8 The short-reads alignment for Siberian crane genome assembly of the Siberian crane

type	value
Mapping rate (%)	99.64
Average sequencing depth (X)	24.62
Coverage of genome >= 0X (%)	99.99
Coverage of genome >= 4X (%)	98.39
Coverage of genome >= 10X (%)	94.98

Table 9 Genome characters statistic of four crane species.

Species	Genome Length (Gb)	Number of Chromosome	Number of Scaffold	Scaffold N50 (Mb)	GC content (%)	Number of gene	accession number
Siberian crane	1.31	33	1,436	83.45	42.42	21,678	CNA0148167
Common crane	1.4	40	753	83.7	44	-	GCA_964106855.1
Whooping crane	1.3	40	929	82.9	43	20,835	GCF_028858705.1
Grey crowned-crane	1.2	37	104	82.6	42.5	-	GCA_011004875.1

References

1 BirdLife International. Species factsheet: *Leucogeranus leucogeranus*. http://www.birdlife.org (2025).

2 Mirande, C. M. & Harris, J. T. in *Crane Conservation Strategy* (Baraboo, Wisconsin, USA: International Crane Foundation Press, (2019).

3 Dussex, N. Comparative Population Genomics Reveal the Determinants of Genome Erosion in Two Sympatric Neotropical Falcons. *Mol. Ecol.* **34**, e17686, https://doi.org/10.1111/mec.17686 (2025).

4 Theissinger, K. *et al.* How genomics can help biodiversity conservation. *Trends Genet.* **39**, 545-559, https://doi.org/10.1016/j.tig.2023.01.005 (2023).

5 Kaewmad, P. et al. First Karyological Analysis of Black the Crowned Crane (Balearica pavonina) and the Scaly-Breasted Munia (Lonchura punctulata). Cytologia 78, 205-211, https://doi.org/10.1508/CYTOLOGIA.78.205 (2013).

6 Chen, Q. *et al.* Understanding the Past to Preserve the Future: Genomic Insights Into the Conservation Management of a Critically Endangered Waterbird. *Mol. Ecol.* **34**, e17606 https://doi.org/10.1111/mec.17606 (2025).

7 Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21,** https://doi.org/10.1186/s13059-020-1935-5 (2020).

8 Chen, Y. *et al.* SOAPnuke: a mapreduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6, https://doi.org/10.1093/gigascience/gix120 (2018).

9 Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770, https://doi.org/10.1093/bioinformatics/btr011 (2011).

10 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204, https://doi.org/10.1093/bioinformatics/btx153 (2017).

11 Koren, S., Walenz, B. P., Berlin, K., Miller, J. R. & Phillippy, A. M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722-736, https://doi.org/10.1101/gr.215087.116 (2016).

12 Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* **27**, 737-746, https://doi.org/10.1101/gr.214270.116 (2017).

13 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, https://doi.org/10.1371/journal.pone.0112963 (2014).

14 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).

15 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259, https://doi.org/10.1186/s13059-015-0831-x (2015).

16 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. **356**, 92-95, https://doi.org/10.1126/science.aal3327 (2017).

17 Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99-101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).

18 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212,

https://doi.org/10.1093/bioinformatics/btv351 (2015).

19 Rhie A., Walenz B.P., Koren S., & Phillippy A.M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 21, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).

20 He, W. *et al.* NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. *Bioinformatics* **39**, btad121, https://doi.org/10.1093/bioinformatics/btad121 (2023).

21 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, https://doi.org/10.1093/bioinformatics/bty191 %J Bioinformatics (2018).

22 Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc.* **25**, 1-14, https://doi.org/10.1002/0471250953.bi0410s05 (2004).

23 Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275, https://doi.org/10.1186/s13059-019-1905-y (2019).

24 Salamov, A. A. & Solovyev, V. V. Ab initio gene finding in Drosophila genomic DNA. *Genome Res.* **10**, 516-522, https://doi.org/10.1101/gr.10.4.516 (2000).

25 Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637-644, https://doi.org/10.1093/bioinformatics/btn013 (2008).

26 Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014, https://doi.org/10.1093/bioinformatics/btad014 (2023).

27 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491-504, https://doi.org/10.1186/1471-2105-12-491 (2011).

28 Zhang, G. *et al*. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**,1311-1320, https://doi.org/10.1126/science.1251385 (2014).

29 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726-731, https://doi.org/10.1016/j.jmb.2015.11.006 (2016).

30 The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158-D169, https://doi.org/10.1093/nar/gkw1099 (2016).

31 Altschul, S. F, Gish W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410, https://doi.org/10.1006/jmbi.1990.9999 (1990).

32 Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351-D360, https://doi.org/10.1093/nar/gky1100 (2018).

33 The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330-D338, https://doi.org/10.1093/nar/gky1055 (2018).

34 Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **25**, 955-964, https://doi.org/10.1093/nar/25.5.955 (1997).

35 Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121-D124, https://doi.org/10.1093/nar/gki081 (2005).

36 Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**,

2933-2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).

37 Chen, Q. Chromosome-level genome assembly and annotation of the critically endangered Siberian crane (Leucogeranus leucogeranus). NCBI Sequence Read Archive. https://identifiers.org/ncbi/insdc.sra:SRP618574 (2025).

38 Chen, Q. Chromosome-level genome assembly and annotation of the critically endangered Siberian crane (Leucogeranus leucogeranus). NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_053455625.1 (2025).

39 Chen, Q. Chromosome-level genome assembly and annotation of the critically endangered Siberian crane (Leucogeranus leucogeranus). figshare. Dataset. https://doi.org/10.6084/m9.figshare.30017956.v1 (2025).