

DATA NOTE

Open Access



Genomic insights into the endangered white-eared night heron (*Gorsachius magnificus*)

Haoran Luo¹, Qingxian Lin^{1*}, Wenzhen Fang¹, Xiaolin Chen¹ and Xiaoping Zhou^{1*}

Abstract

Objectives A genome sequence of a threatened species can provide valuable genetic information that is important for improving the conservation strategies. The white-eared night heron (*Gorsachius magnificus*) is an endangered and poorly known ardeid bird. In order to support future studies on conservation genetics and evolutionary adaptation of this species, we have reported a *de novo* assembled and annotated whole-genome sequence of the *G. magnificus*.

Data description The final draft genome assembly of the *G. magnificus* was 1.19 Gb in size, with a contig N50 of 187.69 kb and a scaffold N50 of 7,338.28 kb. According to BUSCO analysis, the genome assembly contained 97.49% of the 8,338 genes in the Aves (odb10) dataset. Approximately 10.52% of the genome assembly was composed of repetitive sequences. A total of 14,613 protein-coding genes were predicted in the genome assembly, with functional annotations available for 14,611 genes. The genome assembly exhibited a heterozygosity rate of 0.49 heterozygosity per kilobase pair. This draft genome of *G. magnificus* provides valuable genomic resources for future studies on conservation and evolution.

Keywords The white-eared night heron, Endangered bird, Genome assembly, Genome annotation, Heterozygosity

Objectives

The white-eared night heron (*Gorsachius magnificus*) is a medium-sized ardeid bird that distributes in tropical and subtropical moist lowland forests in southern and southwestern China [1–3], northern Vietnam [4], and northeastern India [5]. Due to its small and fragmented population, the *G. magnificus* is currently listed as an Endangered (EN) species by the International Union for the Conservation of Nature (IUCN) [6] and is listed in

the first order of the National Key Protected Wild Animal List in China [7].

The *G. magnificus* is not well understood due to its solitary, nocturnal, and cryptic behaviour [4, 8]. Previous studies on the *G. magnificus* primarily focused on documenting its distribution in new areas [2, 4, 9]. These findings expanded our knowledge of the *G. magnificus*' range and resulted in its threat category being changed from Critically Endangered to Endangered by IUCN in 2000 [10]. Some researchers have suggested further downgrading its threat status because the *G. magnificus* has been observed in almost the entire southern region of China [11]. However, ecological niche modelling predicted that suitable habitats for the *G. magnificus* are limited and scattered within the mountain chains of southern China [12]. It also suggested that future climate change could alter its distribution and pose a threat to its population. Therefore, the authors recommended exercising caution

*Correspondence:

Qingxian Lin

lqx@xmu.edu.cn

Xiaoping Zhou

xpzhou@xmu.edu.cn

¹Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, 361102 Xiamen, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

when considering any downgrading of the threat status for the *G. magnificus*.

The genetic information of a threatened species is valuable for various aspects of their conservation biology, including estimating effective population size, inbreeding level, genetic diversity, and population structure [13]. However, up until now, very little genetic information has been available for *G. magnificus*, except for the sequence of complete mitochondrial DNA [14]. In this study, we sequenced the genome of the *G. magnificus* and evaluated genetic diversity using heterozygous single nucleotide polymorphisms (SNPs) at the whole genome level. The genetic diversity of the *G. magnificus* can help to correctly assess its conservation status. Additionally, the draft genome sequence can facilitate future conservation biology studies to help protect this endangered species.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types and extension	Data repository and identifier (DOI or accession number)
Data file 1	Statistics of sequencing depths,	Spreadsheet (.xlsx)	Figshare: https://doi.org/10.6084/m9.figshare.24311935 [28]
Data file 2	contigs, scaffolds and GC content of the <i>G. magnificus</i> genome assembly	Text raw output (.txt)	Figshare: https://doi.org/10.6084/m9.figshare.24311935 [28]
Data set 1		Fastq files (.fq, gz)	NCBI SRA Database: SRP364708 https://identifiers.org/insdc.sra:SRP364708 [29]
Data set 2		Fasta file (.fa)	NCBI GenBank Database: JALHKT0000000000 https://identifiers.org/nucleotide:JALHKT0000000000 [30]
Data file 3	Statistics of complete, fragmented and missing BUSCOs of the <i>G. magnificus</i> genome assembly	Portable document format (.pdf)	Figshare: https://doi.org/10.6084/m9.figshare.24311935 , [28]
Data file 4	Sequencing reads of <i>G. magnificus</i> genomic DNA	Gff file (.gff)	Figshare: https://doi.org/10.6084/m9.figshare.24311935 , [28]
Data file 5	Genome assembly of <i>G. magnificus</i> Comparison of the heterozygosity of <i>G. magnificus</i> and its close related species Structure annotation gene model file Gene model functional annotation files	Zipped (compressed) file (.zip)	Figshare: https://doi.org/10.6084/m9.figshare.24311935 , [28]

Data description

A muscle sample of a dead *G. magnificus* was provided by the Jiulingshan National Reserve, which was found and confiscated by the reserve personnel from a local farmers' market in Jing'an Town (115.11' 25" E, 28.69' 21"N), Yichun City, Jiangxi province, China in 2007. The muscle tissue was stored at -80 °C after collection. This research was approved by the Ethics Committee for Animal Experimentation of the Xiamen University. Genomic DNA was extracted using the QIAGEN® Puregene Tissue Core Kit A (Qiagen, Beijing, China), following the manufacturer's instructions. Two short insert libraries (230 and 500 bp) were prepared using the Illumina TruSeq DNA Library Preparation Kit (Illumina, San Diego, USA), while three mate pair libraries (2, 5, and 10 kb) were constructed using the Nextera Mate Pair Sample Preparation Kit (Illumina, San Diego, USA). The libraries were sequenced on the Illumina HiSeq 2500 sequencing platform at Novogene (Beijing). The sequencing depths were 30.30× and 26.10× for the two short insert libraries, and 16.05×, 11.26× and 9.90× for the three mate pair libraries, respectively. Raw reads were filtered using Cutadapt [15] and Trimmomatic [16] to remove adapters and low-quality reads (quality score < 20), respectively. The resulting clean reads were assembled using SOAPdenovo2 following the official guidelines [17]. The estimated genome size was determined using 17 k-mers with Jellyfish [18]. A total of 123.14 Gb of clean sequence data was obtained, and the estimated genome size was approximately 1.32 Gb based on the 17 K-mer distribution. The final genome assembly had a length of 1.19 Gb, with contig and scaffold N50 sizes of 187.69 kb and 7,338.28 kb, respectively (Table 1, Data file 1). The longest scaffold was 29,902.37 kb. The completeness of the genome assembly was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.3 [19] with the Aves (odb10) dataset. The analysis showed that 97.49% of the BUSCO genes (8103 single-copy and 25 duplicated) were complete, 0.56% were fragmented (47 genes), and 1.95% were missing (163 genes) (Table 1, Data file 2).

RepeatModeler and RepeatMasker (<http://www.repeat-masker.org>) with the Repbase2 repeat database were used to identify interspersed repeats using a de novo and homology-based approach with default parameters, respectively. TRF [20] was used to identify tandem repeats. The results showed that approximately 10.52% of the assembled genome consisted of repetitive sequences. This included 0.72% tandem repeat sequences, 0.62% DNA repeat elements, 6.07% long interspersed nuclear elements (LINE), 0.11% short interspersed nuclear elements (SINE), 1.67% long terminal repeat elements (LTR), and 1.33% unknown repetitive sequences. A total of 14,613 protein-coding genes were predicted using Genewise2 [21], with an average coding sequence (CDS)

length of 1417.22 bp. Out of these, 14,611 (99.99%) protein-coding genes were functionally annotated by searching against EggNog-mapper V2.1.2 [22], KEGG [23], GO [24], SwissProt [25], and TrEMBL [26] databases. Additionally, the assembled genome also contained 368 candidate microRNA genes (31,150 bp, 0.0026% of the genome) and 230 candidate tRNAs (17,191 bp, 0.0014% of the genome). The genetic diversity of the *G. magnificus* was assessed using individual heterozygosity from SNPs' genotypes, which indicated a genetic diversity of 0.49 heterozygosity per kilobase pair (578,569 heterozygosity loci) (Table 1, Data file 3). This heterozygosity was extremely low compared to other ardeid birds. For example, the heterozygosity of the Japanese night heron (*G. goisagi*, VU), the little egret (*Egretta garzetta*), the boat-billed heron (*Cochlearius cochlearius*, LC) and the black crown night heron (*Nycticorax nycticorax*) were 0.83, 2.51, 4.88 and 6.25 per kilobase pair, respectively. Additionally, the heterozygosity of the *G. magnificus* was comparable to the Crested ibis *Nippon nippon*, a species that faced near-extinction and rebirth, with a genetic diversity of 0.43 heterozygosity per kilobase pair [27]. The low genome-wide heterozygosity observed in other endangered bird species has been shown to be associated with inbreeding depression and the accumulation of harmful mutations [27]. Therefore, we suggested that the endangered conservation status of *G. magnificus* be maintained, and that greater emphasis be placed on restoring its genetic diversity in future conservation efforts.

In conclusion, the draft genome sequence of *G. magnificus* can serve as a valuable resource for future studies on the genetic mechanisms underlying adaptations and the estimation of important parameters relevant to conservation efforts.

Limitations

Due to the difficulty of obtaining fresh tissues, this draft genome assembly was generated using short-read shotgun sequencing. Consequently, the final genome assembly size was smaller than the K-mer estimation, suggesting that the assembly contained a degree of fragments and gaps. In the future, if fresh tissues could be collected, long-read sequencing technologies, such as Pacific BioSciences (PacBio) or Oxford Nanopore sequencing, would help to improve the completeness and accuracy of the genome assembly.

Abbreviations

SNPs	Single nucleotide polymorphisms
BUSCO	Benchmarking Universal Single-Copy Orthologs
IUCN	International Union for the Conservation of Nature
CDS	Coding sequence
LINE	Long interspersed nuclear elements
SINE	Short interspersed nuclear elements
LTR	Long terminal repeat elements
PacBio	Pacific BioSciences

Acknowledgements

We thank the Jiulingshan National Reserve in Jiangxi for help in providing the tissue samples for this study.

Author contributions

HL designed and performed the experiments, completed the data analyses and wrote the original draft manuscript. QL and XZ conceived, directed, and coordinated this study, helped with data analyses and writing, and revised the manuscript. WF and XC collected the samples and assisted with the project designment. All authors read and approved the final manuscript.

Funding

This work was supported by the Fujian Natural Science Foundation of China (no. 2022J01054).

Data availability

The draft genome assembly data are available at GenBank with the accession number: JALHKT000000000 (<https://identifiers.org/nucleotide:JALHKT000000000>) [29]. The associated BioProject number is PRJNA816834 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA816834/>) [30]. The genomic annotation files for genomic analyses can be found at Figshare repository (<https://doi.org/10.6084/m9.figshare.24311935>) [28].

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee for Animal Experimentation of the Xiamen University. The experiments were conducted in accordance with ethical guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 October 2023 / Accepted: 18 January 2024

Published online: 30 January 2024

References

- del Hoyo J, Elliot A, Sargatal J. Handbook of the birds of the World. Volume II. Barcelona: Lynx editions; New World Vultures to Guinea-fowls; 1994.
- Fellowes JR, Fang Z, Shing LK, Hau BCH, Lau MWN, Lam VWY, Young L, Hafner H. Status update on White-eared night Heron *Gorsachius magnificus* in South China. *Bird Conserv Int*. 2001;11(2):101–11.
- He FQ, Fellowes JR, Chan BPL, Lau MWN, Lin JS, Shing LK. An update on the distribution of the 'endangered' white-eared night Heron *Gorsachius magnificus* in China. *Bird Conserv Int*. 2007;17(1):93–101.
- Pilgrim JD, Walsh DF, Tran TT, Nguyen DT, Eames JC, Le MH. The endangered white-eared night Heron *Gorsachius magnificus* in Vietnam: status, distribution, ecology and threats. *Forktail* 2009, 25:142–6.
- Hossain S, Sharma P, Chowdhury S, Das DK, Goldberg N, Beck J, Eaton J, Gallardy R, Robi R, Spencer A, et al. White-eared night Heron *Gorsachius magnificus* records in the Bangladesh sundarbans: a new species for the country. *Indian BIRDS*. 2023;19:50.
- BirdLife I. (2023) Species factsheet: *Oroanassa magnifica*. Downloaded from <http://datazone.birdlife.org/species/factsheet/white-eared-night-heron-orooanassa-magnifica> on 08/10/2023.
- National-Forestry-and-Grassland-Administration: 2021 List of Key Protected Wild Animals in China. National Forestry and Grassland Administration. 2021, China.
- Li BC, Jiang PP, Ding P. First Breeding Observations and a New Locality Record of White-eared night-heron *Gorsachius magnificus* in Southeast China. *Waterbirds*. 2007;30(2):301–4.
- Jiang A, Tan L, Feng H. Breeding observations of White-Eared Night-herons (*Gorsachius magnificus*) in Artificial forests of Southern China. *Waterbirds*. 2017;40(2):173–9.
- BirdLife-International: *Gorsachius magnificus* (amended version of 2016 assessment). *The IUCN Red List of Threatened Species* 2017 2017.

11. He F, Yang X, Deng X, Zhu K, Li L, Lin J, Jiang H, Zhi L. The White-eared night Heron (*Gorsachius magnificus*): from behind the bamboo curtain to the front stage. *Chin Birds*. 2012;2(4):163–6.
12. Hu J, Liu Y. Unveiling the conservation biogeography of a data-deficient endangered bird species under climate change. *PLoS ONE*. 2014;9(1):e84529.
13. Khan S, Nabi G, Ullah MW, Yousaf M, Manan S, Siddique R, Hou H. Overview on the Role of Advance Genomics in Conservation Biology of Endangered Species. *Int J Genomics* 2016, 2016:3460416.
14. Zhou X, Yao C, Lin Q, Fang W, Chen X. Complete mitochondrial genomes render the night Heron Genus *Gorsachius* non-monophyletic. *J Ornithol*. 2015;157(2):505–13.
15. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 2011, 17(1):10.
16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
17. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010;20(2):265–72.
18. Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764–70.
19. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
20. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
21. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res*. 2004;14(5):988–95.
22. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, Orthology assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*. 2021;38(12):5825–9.
23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–29.
25. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. *Methods Mol Biol*. 2007;406:89–112.
26. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000;28(1):45–8.
27. Li S, Li B, Cheng C, Xiong Z, Liu Q, Lai J, Carey HV, Zhang Q, Zheng H, Wei S, et al. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biol*. 2014;15(12):557.
28. Luo H, Lin Q, Fang W, Chen X, Zhou X. Datasets of genome of the white-eared night heron (*Gorsachius magnificus*). Figshare, <https://doi.org/10.6084/m9.figshare.24311935>.
29. Luo H, Lin Q, Fang W, Chen X, Zhou X. NCBI SRA database of genome of the white-eared night heron (*Gorsachius magnificus*). NCBI, <https://identifiers.org/insdc.sra:SRP364708>.
30. Luo H, Lin Q, Fang W, Chen X, Zhou X. NCBI genome assembly of the white-eared night heron (*Gorsachius magnificus*). NCBI, <https://identifiers.org/nucleotide:JALHKT000000000>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.