



OPEN

DATA DESCRIPTOR

# A chromosome-level genome assembly of an avivorous bat species (*Nyctalus aviator*)

Yang Geng<sup>1,5</sup>, Yingying Liu<sup>1,5</sup>, Yu Zhang<sup>2,5</sup>, Lixin Gong<sup>1</sup>, Yu Han<sup>2</sup>, Zhenglanyi Huang<sup>1</sup>, Can Ke<sup>1</sup>, Hui Wu<sup>3</sup>, Aiqing Lin<sup>1,4</sup>, Jiang Feng<sup>1,3,4</sup>✉ & Tinglei Jiang<sup>1,4</sup> ✉

Currently, three carnivorous bat species, namely *Ia io*, *Nyctalus lasiopterus*, and *Nyctalus aviator*, are known to actively prey on seasonal migratory birds (hereinafter referred to as “avivorous bats”). However, the absence of reference genomes impedes a thorough comprehension of the molecular adaptations of avivorous bat species. Herein, we present the high-quality chromosome-scale reference genome of *N. aviator* based on PacBio subreads, DNBSAQ short-reads and Hi-C sequencing data. The genome assembly size of *N. aviator* is 1.77 Gb, with a scaffold N50 of 102 Mb, of which 99.8% assembly was anchored into 21 pseudo-chromosomes. After masking 635.1 Mb repetitive sequences, a total of 19,412 protein-coding genes were identified, of which 99.3% were functionally annotated. The genome assembly and gene prediction reached 96.1% and 96.1% completeness of Benchmarking Universal Single-Copy Orthologs (BUSCO), respectively. This chromosome-level reference genome of *N. aviator* fills a gap in the existing information on the genomes of carnivorous bats, especially avivorous ones, and will be valuable for mechanism of adaptations to dietary niche expansion in bat species.

## Background & Summary

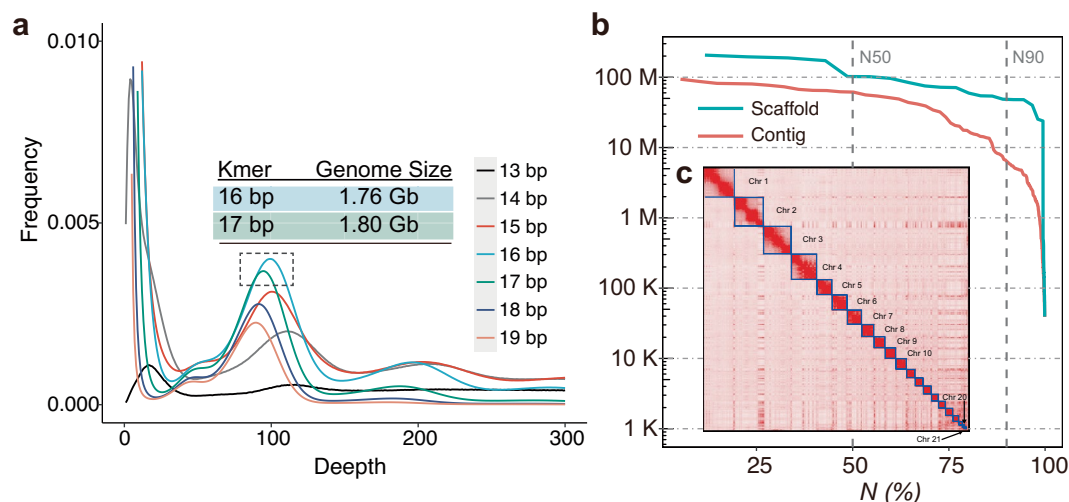
As an important component of the ecological niche, the dietary niche of animals reflects variations in their food intake, which influences their survival and reproduction<sup>1</sup>. Changes to the diet of animals may induce phenotypic variations to open new ecological opportunities<sup>2</sup>, such as physiological (i.e., nutrient assimilation and energy metabolism), morphological, and behavioral variations<sup>3</sup>. Consequently, studies on the genomic adaptations of species with dietary niche variations (i.e., niche expansion) could provide insight into the genetic mechanisms responsible for the ecological niche breadth evolution. Chiroptera (bat) species, serve as an excellent subjects for studying the evolutionary mechanisms of dietary niches due to their diverse diets, which include insectivory, carnivory, piscivory, frugivory, nectarivory, and sanguivory<sup>4</sup>.

Currently, three carnivorous bat species, namely *Ia io*, *Nyctalus lasiopterus*, and *Nyctalus aviator*, are known to actively prey on seasonal migratory birds<sup>5–7</sup>. They usually consume insects in summer and prey on nocturnal migratory birds through an aerial-hawking strategy during spring and autumn. Comparing to closely related insectivorous bat species, the dietary niches of avivorous bat species have expanded from insects to birds<sup>8,9</sup>. Previous studies have identified similarities in the morphology and behavior of three avivorous bat species. However, there remains a lack of understanding of the molecular mechanisms that drive the evolution of this specific feeding habit. For example, previous research has identified physiological adaptations related to avivorous diet by comparing the genomes of *I. io* against other bat species<sup>8</sup>. However, it remains unknown whether these adaptations are also present in *N. aviator* and *N. lasiopterus*, which are distantly related. Additionally, the direct interactions between avivorous bat species and birds contribute to the transmission of viruses<sup>10,11</sup>. For instance, the typical influenza A virus (IAV) is capable of infecting bat cells, and H9 IAV has been identified in bats<sup>12</sup>. Recently, the hemagglutinin (HA) gene of H19 IAV, which was isolated from a wild duck, has exhibited characteristics of both avian and bat influenza viruses<sup>13</sup>. However, little is known about the adaptation of

<sup>1</sup>Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, Changchun, 130117, China. <sup>2</sup>Key Laboratory of Molecular Epigenetics of Ministry of Education, School of Life Sciences, Northeast Normal University, Changchun, 130024, China. <sup>3</sup>College of Life Science, Jilin Agricultural University, Changchun, 130118, China. <sup>4</sup>Key Laboratory of Vegetation Ecology of Education Ministry, Institute of Grassland Science, Northeast Normal University, Changchun, 130024, China. <sup>5</sup>These authors contributed equally: Yang Geng, Yingying Liu, Yu Zhang. ✉e-mail: [fengj@nenu.edu.cn](mailto:fengj@nenu.edu.cn); [jiangtl730@nenu.edu.cn](mailto:jiangtl730@nenu.edu.cn)

Library	Reads number	data (Gb)	Max length (bp)
PacBio	15,539,399	298.86	368,564
DNBSEQ	667,487,037	200.25	/
Hi-C	667,262,960	200.18	/

**Table 1.** Statistics on the genome sequencing data of *N. aviator*.



**Fig. 1** The results of genome assembly for *N. aviator*. **(a)** Genome size estimation by different kmers. The estimated genome size of *N. aviator*, based on 16 bp and 17 bp kmers, produced consistent outcomes, suggesting a genome size of around 1.8 Gb. **(b)** Length distribution of genome assembly at contig- (red) and scaffold-level (green). It indicates the percentage (x%) of the assembly that consists of contigs and scaffolds of at least a certain size. **(c)** Hi-C Map for *N. aviator*. The chromosomes have been ordered by size.

immunity in avivorous bat species. These issues merit further investigation to achieve a more comprehensive understanding of the genetic basis underlying the dietary niche evolution, particularly in distantly related taxa with similar expansion in dietary niche. Nonetheless, the absence of genomes of high quality impedes the possibility of conducting comprehensive research.

Here, we presented a high-quality chromosome-level genome assembly of *N. aviator* using a combination of PacBio subreads (299.16 Gb), DNBSEQ short reads (200.25 Gb), and high-throughput chromatin conformation capture (Hi-C) sequencing data (200.17 Gb) (Table 1). The genome survey revealed an estimated genome size of approximately 1.8 Gb for *N. aviator* Fig. 1a. Finally, we generate a 1.77 Gb genome assembly of *N. aviator* with contig N50 and scaffold N50 of 61.24 Mb and 101.86 Mb, respectively (Fig. 1b, Table 2). Approximately 99.8% of genome sequences were mounted to 21 chromosome-level (20 autosomes and X chromosome) scaffolds (Figs. 1c, 2a), which is consistent with the diploid chromosome number of *N. aviator* ( $2n = 42$ )<sup>14</sup>. The whole-genome synteny analysis showed a strong synteny (>93%) among *N. aviator* and closely related species *Pipistrellus kuhlii* (GCF\_014108245.1<sup>15</sup>, Fig. 2b). The synteny analysis using the chromosome-level genome of *Myotis daubentonii* (GCF\_963259705.1<sup>16</sup>) revealed that the genome assembly of *N. aviator* has attained chromosome-level resolution and successfully characterized the X chromosome (Fig. 2c). The genome assembly consisted of 635.1 Mb (35.75%) repetitive sequences (Fig. 3a, Table 3). After masking repetitive sequences, a total of 19,412 protein-coding genes were predicted, and 99.07% of them were functionally annotated (Fig. 3b, Tables 4, 5). The assessment using Benchmarking Universal Single-Copy Orthologs (BUSCO) revealed 96.1% completion rates for both genome assembly and annotation, as shown in Fig. 3c. This indicating a high-quality assembly and annotation of the genome. In summary, the genome assembly of *N. aviator* establishes a foundation for comprehending the genetic adaptation of bat species with diverse diets and serves as a valuable resource for conducting further studies on the evolutionary mechanisms of dietary niche expansion.

## Methods

**Sample collection and sequencing.** In this research, a female healthy *N. aviator* individual was randomly captured on September 15, 2021, in Congjiang county, Xingyi city, Guizhou province, China. This individual was anesthetized with ether prior to the euthanasia procedure (cervical dislocation). The nine tissues (muscle, brain, lung, liver, heart, ovary, spleen, kidney, and stomach) were sampled for DNA and RNA extraction. All tissues were frozen immediately using liquid nitrogen and then were stored in a  $-80^{\circ}\text{C}$  freezer. For genome sequencing, genomic DNA was extracted from muscle tissue for three sequencing libraries construction, PacBio CLR library (20–40 kb), DNBSEQ library (paired-end 150 bp), and Hi-C library (paired-end 150 bp). The subreads were sequenced using the PacBio Sequel II platform, the short reads and the Hi-C reads were sequenced using DNBSEQ platform. The raw data of DNBSEQ short reads and Hi-C sequencing data were filtered using

Genome assembly statistics	
size (bp)	1,776,615,751
contig number	337
contig N50	61 Mb
scaffold number	195
scaffold N50	101 Mb
GC content (%)	42.7
Hi-C loading rate (%)	99.45
Percent gaps (%)	0.002
Genome assembly consistency	
PacBio subreads mapping rate (%)	99.96
DNBSEQ short reads mapping rate (%)	99.29
DNASEQ short reads coverage > 4X (%)	99.71
DNASEQ short reads coverage > 10X (%)	99.55
DNASEQ short reads coverage > 20X (%)	99.27
DNASEQ short reads coverage > 40X (%)	98.23
BUSCO completeness of genome assembly (a total of 9,926 orthologes)	
Complete BUSCOs	8,862
Complete and single-copy BUSCOs	8,717
Complete and duplicated BUSCOs	145
Fragmented BUSCOs	52
Missing BUSCOs	312

**Table 2.** Statistics on genome assembly and the genome evaluation for *N. aviator*.

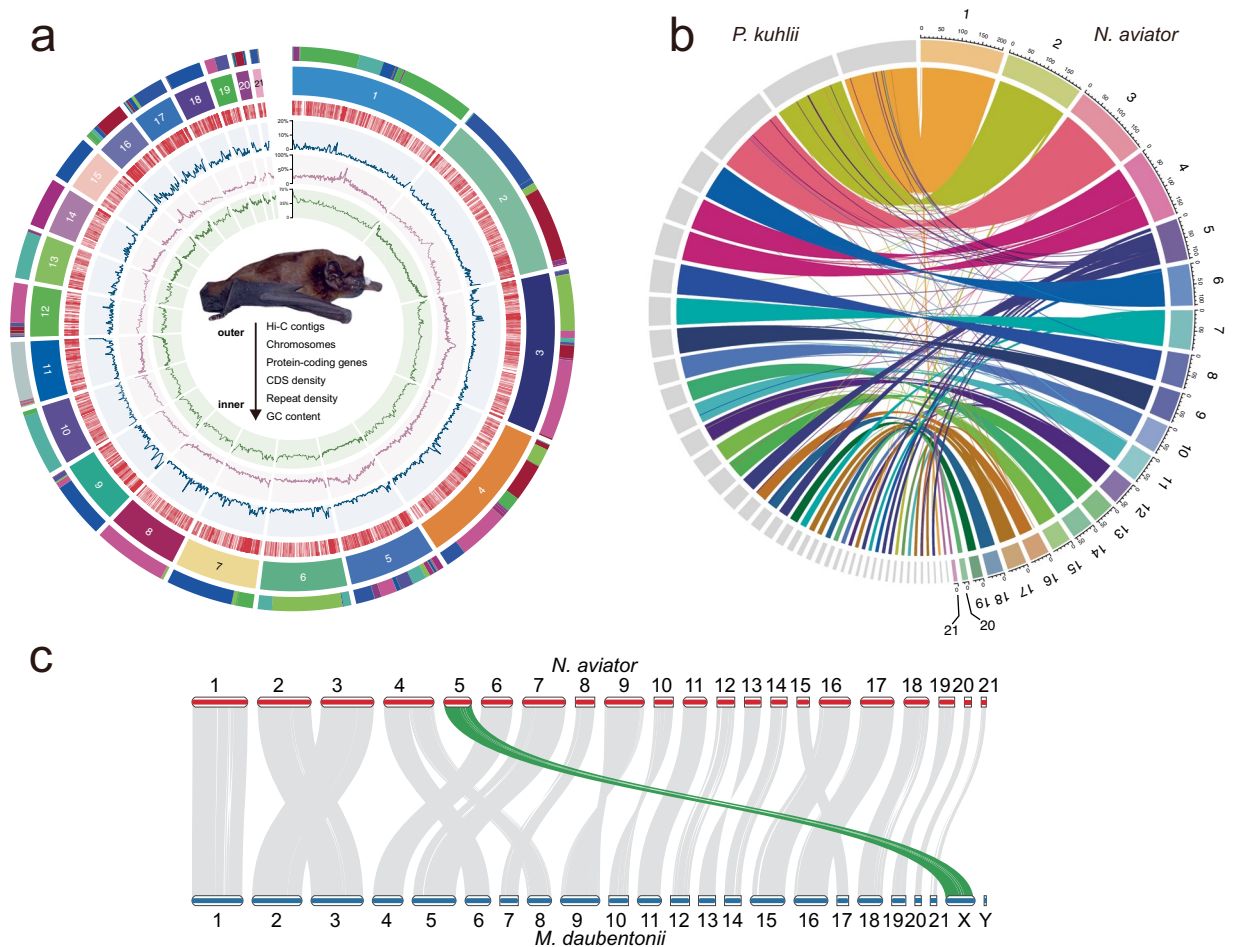
SOAPnuke version 2.1.5<sup>17</sup>. For transcriptome sequencing, total RNA was extracted from each tissue and used for constructing sequencing libraries (paired-end 150 bp), respectively. A total of nine libraries were sequenced on BGISEQ platform. All transcriptome sequencing data was also filtered using SOAPnuke. To ensure the quality of sequencing data, all fastq files were filtered using fastp version 0.23.4 (with parameter: -q 20)<sup>18</sup> to exclude sequences with a phread score below 20, except for PacBio subreads. All procedures involving the capture of bats and experimental procedures were approved by the Science and Technology Ethics Committee of Northeast Normal University, China (permit ID: NENU-202302001).

**Genome assembly.** *Genome survey.* The GCE (genomic character estimator) version 1.02<sup>19</sup> was used to assess the genome size of *N. aviator* based on 200.25 Gb clean short reads before genome assembly. A range of kmers (13–19 bp) lengths were used to estimate genome size of *N. aviator*. The genome size of *N. aviator* was estimated to be approximately 1.8 Gb based on the assessment results when using kmer lengths of both 16 and 17 bp. The subsequent assembly of the genome was guided by the genome size of 1.8 Gb (Fig. 1a).

*Genome assembly.* A total of 299.16 Gb PacBio subreads were corrected using NextDenovo version 2.5.0 (<https://github.com/Nextomics/NextDenovo>). Subsequently, the corrected subreads were pairwise aligned with each other using kbm2 (with parameters: -t 10 -c 2) from the WTDBG2 version 2.5<sup>20</sup>. Several rounds of parameter optimization (with parameters: -A --node-drop 0.25 --node-len [1536, 2048, 2304, 2560] --node-max 400 -s [0.05, 0.07] -e 3 --rescue-low-cov-edges --no-read-length-sort --aln-dovetail [4608, 9216, -1]) were conducted to attain optimal assembly results. The results of parameter optimizations were sorted based on the contig N50 of the assembly, and the longest one was retained. The consensus sequence of the best assembly result is obtained by using wtpoa-cns from WTDBG2. NextPolish version 1.4.0<sup>21</sup> was utilized to correct the assembly results of WTDBG2, aiming to reduce the assembly error rate. Both PacBio subreads and DNBSEQ short reads were employed for this correction. The error-corrected results served as the final contig-level assembly of *N. aviator* for subsequent analysis.

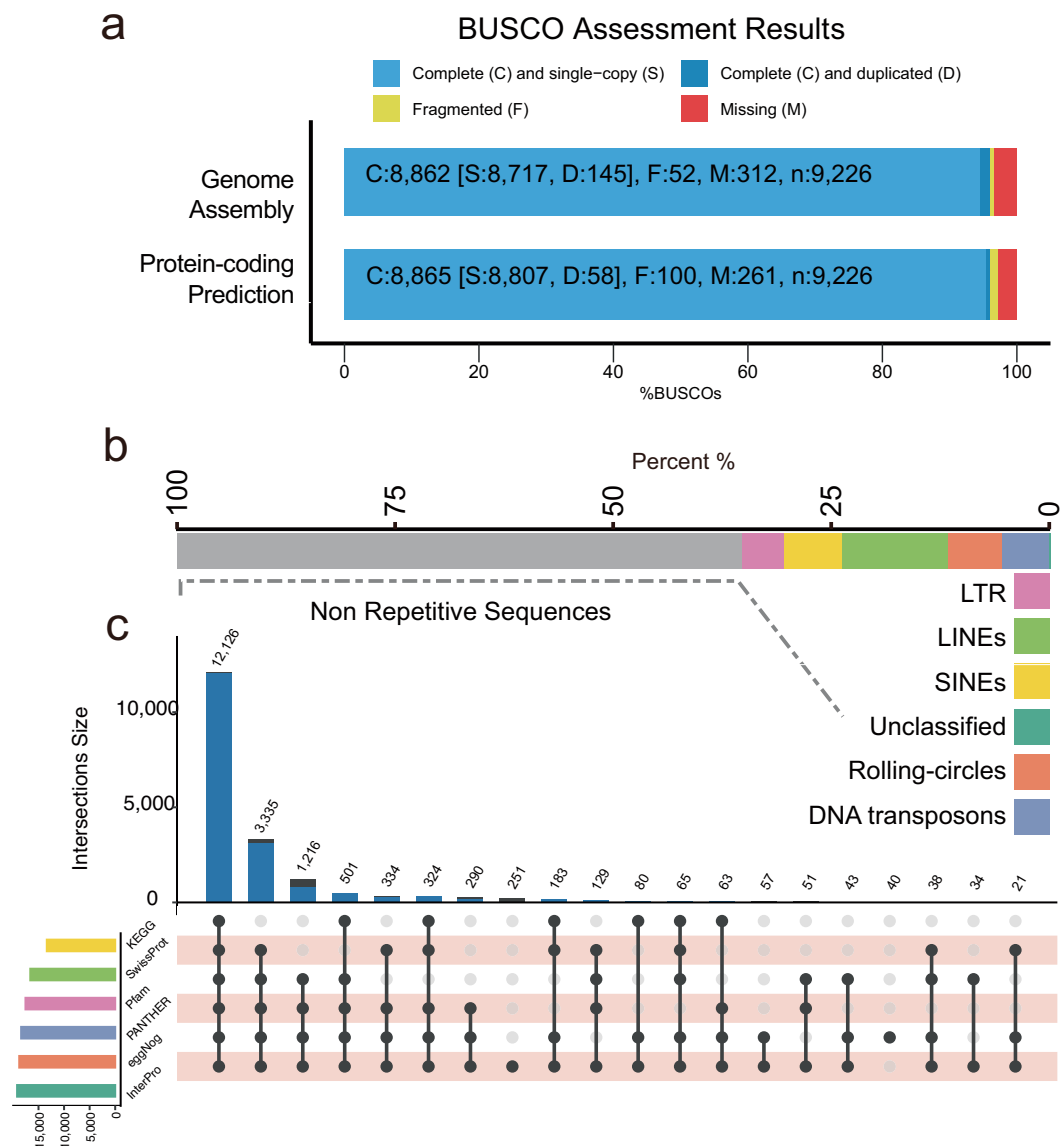
*Hi-C scaffolding.* The Hi-C sequencing data was employed to extend the contig-level assembly, expanding the contig into a chromosome-level scaffold. A total of 200.17 Gb Hi-C sequencing data was filtered using HiC-Pro version 3.1.0<sup>22</sup>. The filtered valid pairs were aligned to the contig assembly using chromap version 0.2.4-r467<sup>23</sup>. Subsequently, the chromosome-level scaffold assembly was performed using YaHS version 1.2a.1<sup>24</sup>. The scaffold assembly was visualized using Juicebox Assembly Tools version 2.20<sup>25</sup> and corrected manually. The final assembly was generated using YaHS based on the reviewed assembly file mentioned above. The completeness of the genome assembly was assessed using BUSCO version 5.2.2<sup>26</sup> with the mammal database (with mammalia\_odb10).

**Genome annotation.** The EDTA version 2.0.0 (with parameters: --sensitive 1 --anno 1 --evaluate 1) was used to annotate repeat elements of *N. aviator* genome assembly<sup>27</sup>. The CDS sequences of *P. kuhlii* were used as input of EDTA to improve the accuracy. The high-quality repeat sequence data of six bats described in this article<sup>28</sup> was download and used as curated library in EDTA. The genomic region containing repetitive sequences



**Fig. 2** The circos plots depict the genomic structure and genome syntenic blocks of *N. aviator*. (a) The tracks, arranged from outer to inner, represent the contigs that make up the scaffolds (adjacent contigs are shown in different colors), 21 chromosome-level scaffolds (The scaffolds were sorted by length, with 1 representing the longest and 21 the shortest), positions of protein-coding genes, density of CDS, density of repetitive sequences, and GC content. CDS density, repetitive sequence density, and GC content are calculated based on 1 Mb windows. (b) The circos plot illustrates the syntenic blocks shared between *N. aviator* and *P. kuhlii*. Only scaffolds over 4 Mb and syntenic blocks larger than 5 kb are depicted. (c) The pairwise synteny among *N. aviator* and *M. daubentonii*. The identified X chromosome of *N. aviator* was highlighted in green.

was masked and utilized for subsequent analyses. The protein-coding genes were predicted with three different strategies: (1) de novo prediction; (2) homology-based prediction; (3) transcriptome-based prediction. All cleaned transcriptome sequences of all tissues of *N. aviator* were mapped to genome using HISAT2 version 2.2.1<sup>29</sup> and were assembled using StringTie version 2.2.1<sup>30</sup>. The transcriptome assembly identified coding regions by utilizing TransDecoder version 5.5.0 (<https://github.com/TransDecoder/TransDecoder>) and constructing the transcriptome database with PASA version 2.5.2<sup>31</sup>. For homology-based prediction, the proteins sequences of bat species were extracted from OrthoDB1<sup>32</sup>, and alignment against the genome assembly of *N. aviator* using miniprot version 0.11<sup>33</sup>. For de novo prediction, the protein sequences and transcriptome alignments mentioned above were used to generate gene prediction by using Braker3 version 3.0.6<sup>34</sup>. In order to enhance the annotation results, we utilized the transcriptome evidence classified as 'I', 'PI', and 'UL' by TOGA version 1.1.6<sup>35</sup> as additional evidence. With humans genome as a reference, the genome assembly of *N. aviator* was aligned to reference by using make\_lastz\_chain ([https://github.com/hillerlab/make\\_lastz\\_chains](https://github.com/hillerlab/make_lastz_chains)) to create a pairwise genome alignment, serving as input for TOGA. The evidence of gene prediction mentioned above was integrated by EvidenceModeler (referred to as EVM in Table 4) version 2.1.1<sup>36</sup> with (1) evidence of Braker3 set to weight 1; (2) the evidence of miniprot set to weight 3; (3) the evidence of PASA set to weight 10; (4) the evidence of TransDecoder and TOGA set to weight 8. Then, two rounds of PASA were conducted to update the integrated gene predictions. We extracted protein-coding sequences from annotation results, and translated them into protein. The short protein sequences (< 50 aa) were removed. Filtered annotation results were aligned to proteins of mammalian database of RefSeq non-redundant protein sequence database (referred to as NR in Table 5) using DIAMOND version 2.0.14<sup>37</sup>. Potential noncoding (e-value of hits < 1e-5) sequences were removed. In total, 19,412 protein-coding genes were predicted in *N. aviator* genome with an average transcript and coding sequences (CDS) length of 36,782.05 bp and 174.47 bp, respectively (Table 4). The proteins coded by genes were search against the SwissProt



**Fig. 3** The functional annotation and completeness of the genome of *N. aviator*. **(a)** The genome assembly and protein-coding gene prediction were assessed using BUSCO. Both the completeness of the assembly and the prediction are over 96%, indicative of high quality. **(b)** The stacked histogram depicts the proportion of repeated sequences in the genome assembly. Each color represents a different type of repetitive sequence. **(c)** The upshot plot displays the functional annotation of predicted protein-coding genes.

Repeat Classes	Count	Length (bp)	Percent (%)
DNA transposons	804,159	96,517,425	5.43
SINEs	637,769	117,935,009	6.64
LINEs	496,985	215,002,783	12.1
LTR elements	377,645	87,159,734	4.91
Total interspersed repeats		520,421,438	29.29
Rolling-circles	644,248	108,683,329	6.12
Unclassified	27,474	3,806,487	0.21
Satellites	17,899	1,497,615	0.08
Simple repeats	43,125	5,687,426	0.32
Low complexity	1,814	327,949	0.02
Total masked		635,129,906	35.75

**Table 3.** Summary of repetitive elements in the genome assembly of *N. aviator*.

Gene set	gene number	transcript number	Average transcript length(bp)	Average CDS length(bp)	Average exons per gene	Average exon length(bp)
Braker3	38,290	44,555	38,006.71	171.79	8.97	171.70
miniprot	/	58,941	48,326.22	166.43	/	/
TOGA	53,428	53,428	31,077.71	165	8.57	165
StringTie	/	61,318	36,801.89	/	8.37	322.73
TransDecoder	20,742	57,593	46,048.91	169.73	29.51	337.68
PASA	/	315,642	8,064.39	/	2.71	409.34
EVM	38,879	38,879	22,535.98	182.2	5.97	182.2
PASA-update	38,812	50,258	28,936.69	173.44	9.96	273.78
Final set*	19,412	19,412	36,782.05	174.47	9.74	259.47

**Table 4.** Statistics on protein-coding genes prediction with three strategies for *N. aviator*, ‘\*’ indicate that the final dataset excluded potential non-coding and short coding genes (coding protein sequence < 50 aa), while preserving the longest transcript for each gene. ‘/’ indicate that the software is unable to predict the corresponding evidence.

Database	Number	Percent (%)
eggNOG	18,769	96.69
SwissProt	16,607	85.55
InterPro	19,188	98.52
All	19,277	99.30

**Table 5.** Summary of gene functional annotation of genome assembly for *N. aviator*.

mammalian database using DIAMOND version 2.0.14 (with parameter: blastp -e 1e-5), the eggNOG database using eggNOG-mapper version 2.1.7<sup>38</sup>, and InterPro database using InterProScan version 5.65-97.0<sup>39</sup> (Table 5).

**Identification of X chromosome.** Based on the karyological studies of *N. aviator*, the karyotype of *Nyctalus* species is strikingly similar to that of *Myotis* species. We select *M. daubentonii* as reference. The protein-coding genes and annotations of *M. daubentonii* were downloaded from NCBI RedSeq database (accession: GCF\_963259705.1<sup>16</sup>) whose X chromosome had been identified. The MCscan (python version)<sup>40</sup> was used to identify synteny between *M. daubentonii* and *N. aviator*. The X chromosome of *N. aviator* was identified based on the syntenic blocks.

### Data Records

The final genome assembly of *N. aviator* has been submitted to the GeneBank database under the accession number GCA\_036971965.1<sup>41</sup> and the Genome Warehouse in National Genomics Data Center under accession number GWHSEW00000000. The raw genome (PacBio, DNBSEQ short reads, Hi-C) and transcriptome sequencing data have been submitted to the Sequence Read Archive at NCBI under accession numbers SRP485754<sup>42</sup>.

### Technical Validation

The mapping rates of DNBSEQ short reads and PacBio subreads were 99.26% and 99.96%, respectively, of which, over 98% of the genome assembly with >40 × coverage. This suggests a significant level of consistency in the assembly of the genome. We employed the genome of *P. kuhlii* (GCF\_014108245.1<sup>15</sup>) as a reference genome and utilized lastz version 1.04.00<sup>43</sup> to align the genome of the *N. aviator* against the reference genome. Genome synteny analysis of *N. aviator* and *P. kuhlii* revealed that more than 93% of the genome assembly consists of syntenic blocks. The X chromosome of *N. aviator* was also successfully identified through pairwise synteny analysis between *N. aviator* and *M. daubentonii*. The BUSCO assessment revealed that the genome assembly of *N. aviator* contained 96.1% of orthologs from the mammalia\_odb10 dataset, comprising 8717 single-copy, 145 duplicated, 52 fragmented, and 312 missing BUSCOs. Furthermore, the final gene annotation of the assembly annotated 96.1% of the orthologs from BUSCO, consisting of 8807 single-copy, 28 duplicated, 100 fragmented, and 261 missing BUSCOs.

### Code availability

In this study, all analyses were conducted following the manuals and tutorials of software and pipeline. The detailed software versions are specified in the methods section. Unless specified otherwise, default or author-recommended parameters were used for software and analysis pipeline. Detailed information about the parameters and custom scripts utilized in this research can be obtained by downloading them from <https://github.com/life404/genome-NycAvi.git>.

Received: 1 February 2024; Accepted: 29 April 2024;  
Published online: 10 May 2024

## References

- Machovsky-Capuska, G. E., Senior, A. M., Simpson, S. J. & Raubenheimer, D. The Multidimensional Nutritional Niche. *Trends in Ecology & Evolution* **31**, 355–365, <https://doi.org/10.1016/j.tree.2016.02.009> (2016).
- Yoder, J. B. *et al.* Ecological opportunity and the origin of adaptive radiations: Ecological opportunity and origin of adaptive radiations. *Journal of Evolutionary Biology* **23**, 1581–1596, <https://doi.org/10.1111/j.1420-9101.2010.02029.x> (2010).
- Palm, W. & Thompson, C. B. Nutrient acquisition strategies of mammalian cells. *Nature* **546**, 234–242, <https://doi.org/10.1038/nature22379> (2017).
- Altringham, J. D. *Bats: From Evolution to Conservation*. <https://doi.org/10.1093/acprof:osobl/9780199207114.001.0001> (Oxford University Press, 2011).
- Gong, L., Shi, B., Wu, H., Feng, J. & Jiang, T. Who's for dinner? Bird prey diversity and choice in the great evening bat, *Ia io*. *Ecol Evol* **11**, 8400–8409, <https://doi.org/10.1002/ece3.7667> (2021).
- Ibáñez, C., Juste, J., García-Mudarra, J. L. & Agirre-Mendi, P. T. Bat predation on nocturnally migrating birds. *Proceedings of the National Academy of Sciences* **98**, 9700–9702, <https://doi.org/10.1073/pnas.171140598> (2001).
- Thabah, A. *et al.* Diet, Echolocation Calls, and Phylogenetic Affinities of the Great Evening Bat (*Ia io*; Vespertilionidae): Another Carnivorous Bat. *Journal of Mammalogy* **88**, 728–735, <https://doi.org/10.1644/06-MAMM-A-167R1.1> (2007).
- Gong, L. *et al.* Behavioral innovation and genomic novelty are associated with the exploitation of a challenging dietary opportunity by an avivorous bat. *Science* 104973, <https://doi.org/10.1016/j.isci.2022.104973> (2022).
- Ibáñez, C. *et al.* Molecular identification of bird species in the diet of the bird-like noctule bat in Japan. *J Zool* **313**, 276–282, <https://doi.org/10.1111/jzo.12855> (2021).
- Abulreesh, H. H., Goulder, R. & Scott, G. W. Wild birds and human pathogens in the context of ringing and migration. *Ringling & Migration* **23**, 193–200, <https://doi.org/10.1080/03078698.2007.9674363> (2007).
- Mollentze, N. & Streicker, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences* **117**, 9423–9430, <https://doi.org/10.1073/pnas.1919176117> (2020).
- Kandeil, A. *et al.* Isolation and Characterization of a Distinct Influenza A Virus from Egyptian Bats. *J Virol* **93**, e01059-18, <https://doi.org/10.1128/JVI.01059-18> (2019).
- Karamendin, K., Kydyrmanov, A. & Fereidouni, S. Has avian influenza virus H9 originated from a bat source? *Front Vet Sci* **10**, 1332886, <https://doi.org/10.3389/fvets.2023.1332886> (2023).
- Harada, M., Uchida, T., Yosida, T. & Takada, S. Karyological studies of two Japanese noctule bats (chiroptera). *Caryologia* **35**, 1–9, <https://doi.org/10.1080/0008714.1982.10796917> (1982).
- NCBI GenBank [https://identifiers.org/refseq:gcf:GCF\\_014108245.1](https://identifiers.org/refseq:gcf:GCF_014108245.1) (2020).
- NCBI GenBank [https://identifiers.org/refseq:gcf:GCF\\_963259705.1](https://identifiers.org/refseq:gcf:GCF_963259705.1) (2023).
- 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).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, <https://doi.org/10.1093/bioinformatics/bty560> (2018).
- Liu, B. *et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv preprint* <https://doi.org/10.48550/ARXIV.1308.2012> (2013).
- Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nature Methods* **17**, 155–158, <https://doi.org/10.1038/s41592-019-0669-3> (2020).
- Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255, <https://doi.org/10.1093/bioinformatics/btz891> (2020).
- 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).
- Zhang, H. *et al.* Fast alignment and preprocessing of chromatin profiles with Chromap. *Nature communications* **12**, 6566, <https://doi.org/10.1038/s41467-021-26865-w> (2021).
- Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808, <https://doi.org/10.1093/bioinformatics/btac808> (2023).
- 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).
- Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**, 543–548, <https://doi.org/10.1093/molbev/msx319> (2018).
- 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).
- Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584, <https://doi.org/10.1038/s41586-020-2486-3> (2020).
- Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357–360, <https://doi.org/10.1038/nmeth.3317> (2015).
- Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290–295, <https://doi.org/10.1038/nbt.3122> (2015).
- Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666, <https://doi.org/10.1093/nar/gkg770> (2003).
- Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research* **51**, D445–D451, <https://doi.org/10.1093/nar/gkac998> (2023).
- Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014, <https://doi.org/10.1093/bioinformatics/btad014> (2023).
- Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* **3**, lqaa108, <https://doi.org/10.1093/nargab/lqaa108> (2021).
- Kirilenko, B. M. *et al.* Integrating gene annotation with orthology inference at scale. *Science* **380**, eabn3107, <https://doi.org/10.1126/science.abn3107> (2023).
- Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7, <https://doi.org/10.1186/gb-2008-9-1-r7> (2008).
- Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368, <https://doi.org/10.1038/s41592-021-01101-x> (2021).
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829, <https://doi.org/10.1093/molbev/msab293> (2021).
- Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240, <https://doi.org/10.1093/bioinformatics/btu031> (2014).
- Tang, H. *et al.* Synteny and Collinearity in Plant Genomes. *Science* **320**, 486–488, <https://doi.org/10.1126/science.1153917> (2008).
- NCBI GenBank [https://identifiers.org/ncbi/insdc.gca:GCA\\_036971965.1](https://identifiers.org/ncbi/insdc.gca:GCA_036971965.1) (2024).

42. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP485754> (2024).  
43. Harris, R. Improved pairwise alignment of genomic DNA. (The Pennsylvania State University, 2007).

### Acknowledgements

This study was supported by the National Natural Science Foundation of China (grant nos. 32371562, 32192424, 31922050, 32171489 and 32301289), and the Fund of the Jilin Province Science and Technology Development Project (20220101273JC), the Fundamental Research Funds for the Central Universities (2412023YQ002), and the Special Foundation for National Science and Technology Basic Research Program of China (2021FY100301).

### Author contributions

Designed research: Tinglei Jiang and Jiang Feng. Collection of samples: Lixin Gong, Zhenglanyi Huang, Can Ke. Genome assembly and data analysis: Yang Geng. DNA and RNA extraction: Yu Zhang, Yu Han. Manuscript writing: Yang Geng and Yingying Liu. Hui Wu, Aiqing Lin, Jiang Feng, and Tinglei Jiang provided suggestions for the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to J.F. or T.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 Author(s) 2024