



A reference genome for the eastern bettong (*Bettongia gaimardi*)

Silver, Luke W.; Edwards, Richard J.; Neaves, Linda; Manning, Adrian D.; Hogg, Carolyn J.; Banks, Sam

Published in:
F1000Research

DOI:
[10.12688/f1000research.157851.2](https://doi.org/10.12688/f1000research.157851.2)

Published: 01/01/2025

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Silver, L. W., Edwards, R. J., Neaves, L., Manning, A. D., Hogg, C. J., & Banks, S. (2025). A reference genome for the eastern bettong (*Bettongia gaimardi*). *F1000Research*, 13, 1-15. Article 1544. <https://doi.org/10.12688/f1000research.157851.2>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



GENOME NOTE

REVISED A reference genome for the eastern bettong (*Bettongia gaimardi*)

[version 2; peer review: 3 approved]

Luke W Silver ^{1,2}, Richard J Edwards^{3,4}, Linda Neaves⁵, Adrian D. Manning⁵, Carolyn J Hogg ^{1,2}, Sam Banks⁶¹Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, The University of Sydney, Camperdown, NSW, 2006, Australia²The University of Sydney School of Life and Environmental Sciences, Camperdown, New South Wales, 2006, Australia³Minderoo OceanOmics Centre at UWA, The University of Western Australia Oceans Institute, Crawley, Western Australia, 6009, Australia⁴Evolution and Ecology Research Centre, University of New South Wales School of Biotechnology and Biomolecular Sciences, Kensington, New South Wales, 2033, Australia⁵Australian National University Fenner School of Environment and Society, Acton, Australian Capital Territory, 2601, Australia⁶Charles Darwin University Research Institute for the Environment and Livelihoods, Casuarina, Northern Territory, 0909, Australia**v2** First published: 20 Dec 2024, 13:1544
<https://doi.org/10.12688/f1000research.157851.1>Latest published: 27 Jan 2025, 13:1544
<https://doi.org/10.12688/f1000research.157851.2>**Abstract****Abstract**

The eastern or Tasmanian bettong (*Bettongia gaimardi*) is one of four extant bettong species and is listed as 'Near Threatened' by the IUCN. We sequenced short read data on the 10x system to generate a reference genome 3.46Gb in size and contig N50 of 87.36Kb and scaffold N50 of 2.93Mb. Additionally, we used GeMoMa to provide and accompanying annotation for the reference genome. The generation of a reference genome for the eastern bettong provides a vital resource for the conservation of the species.

Keywords

Genome annotation, reference genome, bettong, marsupial



This article is included in the **Genomics and Genetics** gateway.

Open Peer Review**Approval Status**

	1	2	3
version 2 (revision) 27 Jan 2025			 view
version 1 20 Dec 2024	 view	 view	 view

1. **Naoki Osada** , Hokkaido University, Sapporo, Japan2. **Jonathan J Hughes** , University of California, Riverside, USA3. **Németh Attila** , University of Debrecen, Debrecen, Hungary

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Luke W Silver (luke.silver@sydney.edu.au)

Author roles: **Silver LW:** Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation; **Edwards RJ:** Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation; **Neaves L:** Formal Analysis, Resources; **Manning AD:** Conceptualization, Data Curation, Funding Acquisition, Supervision, Writing – Review & Editing; **Hogg CJ:** Supervision, Writing – Review & Editing; **Banks S:** Conceptualization, Data Curation, Funding Acquisition, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: LWS is supported by the Australian BioCommons which is enabled by NCRIS via Bioplatforms Australia and the University of Sydney. RJE was supported by the Australian Research Council (LP180100721). The eastern bettong translocation program was funded by Australian Research Council Linkage Projects LP 110100126 and LP140100209, including cash and in-kind support from the ACT Government. ADM was supported by an Australian Research Council Future Fellowship (FT100100358) during the first phase of the bettong translocation project.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Silver LW *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Silver LW, Edwards RJ, Neaves L *et al.* **A reference genome for the eastern bettong (*Bettongia gaimardi*)** [version 2; peer review: 3 approved] F1000Research 2025, 13:1544 <https://doi.org/10.12688/f1000research.157851.2>

First published: 20 Dec 2024, 13:1544 <https://doi.org/10.12688/f1000research.157851.1>

REVISED Amendments from Version 1

We have addressed concerns from all reviewers and updated the manuscript to reflect these changes. Including, providing additional details of the sequencing and assembly methodology reflected by providing specific information on software used and the amount of data generated. We have also provided reasoning for the sequencing methods used but acknowledge the increase in genome quality that could be achieved by using long read technologies and sequencing multiple individuals.

Any further responses from the reviewers can be found at the end of the article

Introduction

The eastern or Tasmanian bettong (*Bettongia gaimardi*) is a small nocturnal Australian marsupial in the potoroid family and is considered an important ecosystem engineer due to its habit of digging and feeding on fungi (Munro et al. 2019, Ross et al. 2019a, 2019b). Eastern bettongs were once widespread across south-eastern Australia but are reported to have gone extinct on mainland Australia around the 1920's due to predation from introduced carnivores and land clearing (Short 1998). Eastern bettongs are now confined to the eastern half of Tasmania where they are listed as 'Near Threatened' by the IUCN Red List (Burbidge et al. 2016).

Australia has experienced the highest extinction rate of mammals on any continent over the past 200 years, accounting for 28% of the world's mammal extinctions since the year 1600 (McKenzie et al. 2007). Nationwide, a number of reintroduction programs are being implemented for the conservation of locally-extinct mammals, typically in the 'critical weight range' of 35 – 5500g (Burbidge and McKenzie 1989). In the case of the eastern bettong, the species was reintroduced from the state of Tasmania to two fenced reserves in the Australian Capital Territory (ACT) between 2011 and 2012 (Batson et al. 2016).

The generation of a reference genome will provide a valuable resource in the management of the two reintroduced populations of eastern bettongs and contribute to the global effort to sequence all eukaryotic life on Earth (Lewin et al. 2022). To generate a reference genome, we sequenced DNA with 10x Genomics short reads and used GeMoMa to produce a genome annotation.

Methods**Sample collection and DNA/RNA extraction and sequencing**

We used muscle tissue from a deceased male pouch young (*B. gaimardi*) individual collected from Mulligan's Flat Woodland Sanctuary during population monitoring in 2014 and was frozen immediately after collection. The sample was collected under Australian National University Animal Experimentation Ethics Committee ethics protocol A2011/017. DNA extraction used the QIAGEN Genomic Tips kit (Qiagen Catalogue # 10223) yielding 170ng/μl of DNA. Sequencing used the KAPA HyperPrep PCR free library kits (Roche Catalogue # KK8503) and two lanes of HiSeq Xten 150 bp PE sequencing (Illumina) at the Ramaciotti Centre for Genomics (UNSW, Sydney). Additional sequencing using the 10X Chromium Genomics library prep with >50 kb size selection was performed using 2 lanes of HiSeq Xten (150 bp PE) (Illumina). The sample was accessioned to the Australia Museum Accession number AM M.56404.

Genome assembly

Raw 10x data was assembled using Supernova v2.0.0 (RRID:SCR_016756) (Weisenfeld et al. 2017) with default settings and pseudohaplotype output. Haplotypes were tidied and filtered using Diploidocus v0.3.0 (RRID:SCR_021231) (Stuart et al. 2022), which filtered redundant and 100% unresolved scaffolds, and assigned the longest version of each scaffold to haplotype 1 [bettong.v1.0]. Scaffolds flagged as diploid (identical in both haplotypes) were then added back to haplotype 2 and each haplotype ran through Telociraptor v0.9.0 (<https://github.com/slimsuite/telociraptor>) to length-sort and rename scaffolds after modifying the ends to reveal telomeres, where appropriate [bettong.v1.1]. Scaffolds identified by Tiara v1.0.3 (Karlicki et al. 2022) as being from archaea, bacteria, prokaryota or organelle were filtered, along with scaffolds flagged for exclusion or review by FCS-GX [bettong.v1.2] (Astashyn et al. 2023).

Completeness was estimated using Benchmarking Universal Single-Copy Orthologues (BUSCO, RRID:SCR_015008) v5.4.3 (Simao et al. 2015) using the mammalia_odb10 (n:9226) and vertebrata_odb10 datasets (n:3354).

Repetitive elements of the genome were identified, classified and masked using a Pawsey Supercomputing Centre Nimbus cloud machine (256GB RAM, 64 vCPU, 3 TB storage) by building a database using RepeatModeler v2.0.1 (RRID:SCR_015027) (Flynn et al. 2020) with default settings; repeats were then masked using RepeatMasker v4.0.9 (RRID:SCR_012954) (Smit et al. 2013-2015).

Table 1. Assemblies used for GeMoMa annotations.

Common name	Scientific name	Assembly ID	Reference
Cow	<i>Bos taurus</i>	ARS-UCD1.2	(Rosen et al. 2020)
Human	<i>Homo sapiens</i>	GRCh38.p13	
Opossum	<i>Monodelphis domestica</i>	ASM229v1	(Mikkelsen et al. 2007)
Mouse	<i>Mus musculus</i>	GRCm39	
Tammar wallaby	<i>Notamacropus eugenii</i>	Meug_1.0	(Renfree et al. 2011)
Platypus	<i>Ornithorhynchus anatinus</i>	mOrnAna1.p.v.a	(Zhou et al. 2021)
Koala	<i>Phascolarctos cinereus</i>	phaCin_unsw_v4.1	(Johnson et al. 2018)
Tasmanian devil	<i>Sarcophilus harrisii</i>	mSarHar1.11	(Stammnitz et al. 2023)
Wombat	<i>Vombatus ursinus</i>	bare-nosed_wombat_genome_assembly	

Genome annotation

A homology-based annotation was created independently for each haplotype using GeMoMa v1.9 (RRID:SCR_012954) (Keilwagen et al. 2019) using the annotation from nine Ensembl mammalian genomes (cow [*Bos taurus*], human [*Homo sapiens*], opossum [*Monodelphis domestica*], mouse [*Mus musculus*], Tammar wallaby [*Macropus eugenii*], platypus [*Ornithorhynchus anatinus*], koala [*Phascolarctos cinereus*], Tasmanian devil [*Sarcophilus harrisii*], wombat [*Vombatus ursinus*]) (Table 1) and default settings. SAAGA v0.7.9 (<https://github.com/slimsuite/saaga.git>) was used to map annotated proteins onto a combined dataset of SwissProt (Edwards and Palopoli 2015, UniProt 2023) and Quest for Orthologues reference proteomes (Nevers et al. 2022) to add descriptions and extract the longest isoform per gene for completeness estimation using BUSCO v.5.4.3 (RRID:SCR_015008) in protein mode against the vertebrata_obd10 (n:3354) and mammalia_obd10 lineages (n:9226) (Simao et al. 2015).

The ‘genestats’ script (<https://github.com/darencard/GenomeAnnotation>) was used to obtain the average number of exons and introns and the average exon and intron length.

Results

Genome assembly

Sequencing generated 185M reads of short read data and 149M reads of 10× Genomics data. Genome assembly with Supernova estimated a 3.79 Gb genome size (46.88X raw coverage) and assembled a 3.57 Gb genome in 38,249 scaffolds (scaffold N50=2.77 Mb) (Silver 2024). Following Diploidocus cleanup, there were 27,408 primary scaffolds (3.46 Gb) with 1,681 alternative scaffolds (3.01 Gb). Telociraptor made five inversions and trimmed one contig. Contamination removal filtered 786 scaffolds (2.00 Mb) from each haplotype. This gave a final genome size of 3.46 Gb with 26,623 scaffolds (Table 2). The genome size is comparable to that of other marsupial genomes, including that of the closely related woylie (*Bettongia penicillate ogilbyi*) (Haouchar et al. 2016, Peel et al. 2021). BUSCO completeness of the final genome was 92.2% for mammalia_odb10 and over 96.9% for vertebrata_odb10 (96.8% for haplotype two) (Table 2). Whilst the BUSCO scores suggest a highly complete genome, the use of long read data such as PacBio or Oxford Nanopore would assist in increasing the contiguity of the assembly. Additionally, 53.08% of the genome was identified as repeats, which is similar to other marsupials, including the closely related woylie (53.05%) (Peel et al. 2021) (Table 3).

Table 2. Genome assembly statistics of the eastern bettong (*Bettongia gaimardi*) bettong.v1.2hap1 assembly.

Metric	
Assembly size (Gb)	3.46
Number of contigs	91,460
Contig N50 (kb)	87.36
Contig N90 (kb)	64.48
Contig L50	11,569

Table 2. *Continued*

Metric	
Contig L90	17,470
Longest contig (kb)	956.31
GC content (%)	38.65
Number of Scaffolds	26,623
Scaffold N50 (Mb)	2.93
Scaffold N90 (Mb)	2.21
Scaffold L50	358
Scaffold L90	524
Longest Scaffold (Mb)	15.08
Complete vertebrata_odb10 BUSCOs	96.9% (Single copy: 92.9%, Duplicated: 4.0%)
Fragmented vertebrata_odb10 BUSCOs	2.0%
Missing vertebrata_odb10 BUSCOs	1.1%
Complete mammalia_odb10 BUSCOs	92.2% (Single copy: 89.4%, Duplicated: 2.8%)
Fragmented mammalia_odb10 BUSCOs	1.8%
Missing mammalia_odb10 BUSCOs	6.0%
Gaps (%)	1.44

Table 3. Classification of repeat elements of the eastern bettong (*Bettongia gaimardi*) genome assembly.

Repeat element	Number of elements	% of sequence
SINEs	2,091,730	8.91
ALUs	19,968	0.10
MIRs	2,069,105	8.08
LINES	3,10,0486	33.21
LINE1	1,049,710	19.2
LINE2	1,146,482	7.49
L3/CR1	563,684	3.09
LTR elements	74,485	0.8
ERVL	15,049	0.18
ERV Class I	23,496	0.23
ERV Class II	23,617	0.28
DNA elements	761,393	2.67
<i>hAT-Charlie</i>	154,968	0.72
TcMar-Tigger	41,111	0.23
Unclassified	1,179,121	7.5
Total interspersed repeats	1,836,063,741bp	53.08
Small RNA	653	0
Satellites	23,609	0.13

Genome annotation

Genome annotation with GeMoMa predicted 36,068 and 36,015 protein coding genes for haplotype one and two, respectively, which is a large over-estimation with other marsupials having around 20,000 protein coding genes (Johnson et al. 2018, Brandies et al. 2020). In the future, generating transcriptomes for a variety of eastern bettong tissues would likely improve the accuracy of the annotation. The annotation is highly complete with 94.1% of mammalian protein BUSCOs complete (Table 4). The average protein length was 384.3 and 384.7 amino acids for haplotype one and two, respectively, with an average of 6.52 exons per gene (Table 4). On average, predicted proteins were 87.8% the length of

Table 4. Statistics of the annotation of the eastern bettong (*Bettongia gaimardi*).

Metrics	
Annotation	
Complete vertebrata_odb10 BUSCOs	97.3% (Single copy: 93.9%, Duplicated: 3.4%)
Fragmented vertebrata_odb10 BUSCOs	1.6%
Missing vertebrata_odb10 BUSCOs	1.1%
Complete mammalia_odb10 BUSCOs	94.1% (Single copy: 91.5%, Duplicated: 2.6%)
Fragmented mammalia_odb10 BUSCOs	1.7%
Missing mammalia_odb10 BUSCOs	4.2%
Average number of exons per gene	6.52
Average Protein Length (aa)	384.3/384.7 haplotype 1/haplotype 2

their best SwissProt/QFO hit, suggesting some fragmentation of the annotation, which might be inflating the numbers of annotated genes.

Ethical considerations

Samples were collected under Australian National University Animal Experimentation Ethics Committee ethics protocol A2011/017 (Approved 2011, expired Dec 2014). The sample was collected in during trapping in 2014.

Data availability

The raw data are publicly available through the Bioplatforms Australia Oz Mammals Genomes: <https://data.bioplatforms.com/organization/bpa-omg>. The assembled and annotated genome herein is hosted on the Australasian Genomes site (<https://awgg-lab.github.io/australasiangenomes/>) in addition to NCBI.

Raw genome sequences are available on:

NCBI's Short Read Archive (SRA): Raw DNA data for generation of genome. SRX26311185 and SRX26311186 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1095660>) (Silver et al. 2024).

The data produced as part of this study are stored on NCBI under BioProjects PRJNA1095660 (Silver et al. 2024). Databases of molecular data on the NCBI Web site include such examples as nucleotide sequences (GenBank), protein sequences, macromolecular structures, molecular variation, gene expression, and mapping data. They are designed to provide and encourage access within the scientific community to sources of current and comprehensive information. Therefore, NCBI itself places no restrictions on the use or distribution of the data contained therein.

Reporting guidelines

Figshare: ARRIVE checklist for A reference genome for the eastern bettong (*Bettongia gaimardi*), DOI: <https://doi.org/10.6084/m9.figshare.27144360.v1> (Silver 2024).

The project contains the following reporting guidelines:

- Author Checklist – ARRIVE

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Acknowledgements

The authors would like to acknowledge computational resource support from: Galaxy Australia, a service provided by the Australian Biocommons and its partners; the University of Sydney's High Performance Computing facility Artemis provided by the Sydney Informatics Hub; the University of New South Wales Katana High Performance Computing (doi:10.26190/669X-A286). Support for DNA sequencing was provided through the Oz Mammals Genomics (OMG) Initiative consortium (<https://ozmammalsgenomics.com/consortium/>), which was funded by Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy.

The eastern bettong reintroduction was conducted as part of and with support of the “Mulligans Flat – Gorooyarroo Woodland Experiment” (<https://www.coexistenceconservationlab.org/mulligans-flat-gorooyarroo-woodland-experiment>). Thanks to the ACT Government and Woodlands and Wetlands Trust and their staff for their support for the eastern bettong reintroduction project at Mulligans Flat Woodland Sanctuary. Thanks to Brittany Brocket for the initial DNA extraction.

References

- Astashyn A, Tvedte ES, Sweeney D, *et al.*: **Rapid and sensitive detection of genome contamination at scale with FCS-GX.** *bioRxiv.* 2023.
[Publisher Full Text](#)
- Batson W, Fletcher D, Portas T, *et al.*: **Re-introduction of eastern bettong to a critically endangered woodland habitat in the Australian Capital Territory, Australia.** *Global Re-introduction Perspectives.* 2016; pp. 172–177.
- Brandies PA, Tang S, Johnson RSP, *et al.*: **The first Antechinus reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies.** *GigaByte.* 2020; 2020: gigabyte7–gigabyte22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Burbidge AA, McKenzie N: **Patterns in the modern decline of Western Australia's vertebrate fauna: Causes and conservation implications.** *Biol. Conserv.* 1989; 50(1-4): 143–198.
[Publisher Full Text](#)
- Burbidge AA, Woinarski J, Johnson CN: **Bettongia giamardi.** *The IUCN Red List of Threatened Species.* 2016.
[Publisher Full Text](#)
- Edwards RJ, Palopoli N: **Computational prediction of short linear motifs from protein sequences.** *Methods Mol. Biol.* 2015; 1268: 89–141.
[Publisher Full Text](#)
- Flynn JM, Hubley R, Goubert C, *et al.*: **RepeatModeler2 for automated genomic discovery of transposable element families.** *Proc. Natl. Acad. Sci. USA.* 2020; 117(17): 9451–9457.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Haouchar D, Pacioni C, Haile J, *et al.*: **Ancient DNA reveals complexity in the evolutionary history and taxonomy of the endangered Australian brush-tailed bettongs (Bettongia: Marsupialia: Macropodidae: Potoroinae).** *Biodivers. Conserv.* 2016; 25(14): 2907–2927.
[Publisher Full Text](#)
- Johnson RN, O'Meally D, Chen Z, *et al.*: **Adaptation and conservation insights from the koala genome.** *Nat. Genet.* 2018; 50(8): 1102–1111.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Karlicki M, Antonowicz S, Karnkowska A, *et al.*: **Tiara: Deep learning-based classification system for eukaryotic sequences.** *Bioinformatics.* 2022; 38(2): 344–350.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Keilwagen J, Hartung F, Grau J: **GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data.** *Methods Mol. Biol.* 2019; 1962: 161–177.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lewin HA, Richards S, Lieberman Aiden E, *et al.*: **The Earth BioGenome Project 2020: Starting the clock.** *Proc. Natl. Acad. Sci. USA.* 2022; 119(4): e2115635118.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- McKenzie N, Burbidge A, Baynes A, *et al.*: **Analysis of factors implicated in the recent decline of Australia's mammal fauna.** *J. Biogeogr.* 2007; 34(4): 597–611.
[Publisher Full Text](#)
- Mikkelsen TS, Wakefield MJ, Aken B, *et al.*: **Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences.** *Nature.* 2007; 447(7141): 167–177.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Munro NT, McIntyre S, Macdonald B, *et al.*: **Returning a lost process by reintroducing a locally extinct digging marsupial.** *PeerJ.* 2019; 7: e6622.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nevers Y, Jones TEM, Jyothi D, *et al.*: **The Quest for orthologs orthology benchmark service in 2022.** *Nucleic Acids Res.* 2022; 50(W1): W623–W632.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peel E, Silver L, Brandies P, *et al.*: **A reference genome for the critically endangered woylie, *Bettongia penicillata ogilbyi*.** *GigaByte.* 2021; 2021: gigabyte35–gigabyte15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Renfree MB, Papenfuss AT, Deakin JE, *et al.*: **Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development.** *Genome Biol.* 2011; 12(8): R81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rosen BD, Bickhart DM, Schnabel RD, *et al.*: **De novo assembly of the cattle reference genome with single-molecule sequencing.** *Gigascience.* 2020; 9(3).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ross CE, McIntyre S, Barton PS, *et al.*: **A reintroduced ecosystem engineer provides a germination niche for native plant species.** *Biodivers. Conserv.* 2019a; 29(3): 817–837.
[Publisher Full Text](#)
- Ross CE, Munro NT, Barton PS, *et al.*: **Effects of digging by a native and introduced ecosystem engineer on soil physical and chemical properties in temperate grassy woodland.** *PeerJ.* 2019b; 7: e7506.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Short J: **The extinction of rat-kangaroos (Marsupialia: Potoroidae) in New South Wales, Australia.** *Biol. Conserv.* 1998; 86(3): 365–377.
[Publisher Full Text](#)
- Silver L: **ARRIVE checklist for A reference genome for the eastern bettong (Bettongia gaimardi).** Dataset. *figshare.* 2024.
[Publisher Full Text](#)
- Silver LW, Edwards EJ, Neaves L, *et al.*: **Bettongia gaimardi genome sequencing and assembly.** National Centre for Biotechnology Information: Sequence Read Archive; 2024.
- Simao FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; 31(19): 3210–3212.
[Publisher Full Text](#)
- Smit A, Hubley R, Green P: **RepeatMasker Open-4.0.** 2013-2015. Retrieved 19 December 2019.
[Reference Source](#)
- Stammnitz MR, Gori K, Kwon YM, *et al.*: **The evolution of two transmissible cancers in Tasmanian devils.** *Science.* 2023; 380(6642): 283–293.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Stuart KC, Edwards RJ, Cheng Y, *et al.*: **Transcript and annotation-guided genome assembly of the European starling.** *Mol. Ecol. Resour.* 2022; 22(8): 3141–3160.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- UniProt C: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; 51(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Weisenfeld NI, Kumar V, Shah P, *et al.*: **Direct determination of diploid genome sequences.** *Genome Res.* 2017; 27(5): 757–767.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhou Y, Shearwin-Whyatt L, Li J, *et al.*: **Platypus and echidna genomes reveal mammalian biology and evolution.** *Nature.* 2021; 592(7856): 756–762.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 28 January 2025

<https://doi.org/10.5256/f1000research.177228.r362834>

© 2025 Attila N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Németh Attila 

University of Debrecen, Debrecen, Hungary

Dear Author(s),

I accept the authors' responses. I have no further comments.

Best wishes.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: molecular phylogenetics, phylogeography, (molecular) taxonomy, conservation genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 17 January 2025

<https://doi.org/10.5256/f1000research.173363.r351906>

© 2025 Attila N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Németh Attila 

University of Debrecen, Debrecen, Hungary

Dear Editor(s),
Dear Authors,

I am sending below my comments on Manuscript, entitled "A reference genome for the eastern bettong (*Bettongia gaimardi*)," which was submitted for publication in the journal F1000Research.

I find the article important, of scientific value and interest, because it is the first to provide information on the genome of an animal species, a reference genome, as the authors say, for which we have no such information.

Nevertheless, some concerns have been raised that make me unsure whether the genomic data provided can be considered a true reference genome.

- Why was the reference genome based on a sample from only one individual? It is not an endangered species, so what was the reasoning behind this choice?
- Why were only short reads performed? The length of repetitive regions is typically much greater than that of short reads, which raises questions about the methodology used.
- How accurate do the authors estimate their chosen method is?

Although I am not a real expert on genome assembly, I recommend that the authors add a brief section in the manuscript discussing the limitations of their findings. This section could explain the rationale behind their decision to work with a single individual and the reasons for choosing this particular method, such as the use of short reads.

I also think it is important to briefly discuss what the authors think might explain why this animal has one and a half times as many protein-coding regions as the known marsupial genome. Could there be a biological or evolutionary reason for this?

Overall, I believe the manuscript is valuable and I support its indexing. While I am uncertain if the results can truly be classified as a reference genome, I am convinced that indexing them will greatly benefit the scientific community. At the same time, I feel that the suggested enhancements could significantly improve the quality of the manuscript and ensure the correct perception and subsequent further use of the results.

Best regards,

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: molecular phylogenetics, phylogeography, (molecular) taxonomy, conservation genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 19 Jan 2025

Luke Silver

1. Why was the reference genome based on a sample from only one individual? It is not an endangered species, so what was the reasoning behind this choice?

Response: A single individual was used as this is still the current best practice in wildlife for generating reference assemblies, we agree that in the near term future it will become more feasible to generate a pangenome for wildlife species to truly capture the entire variation of a species.

2. Why were only short reads performed? The length of repetitive regions is typically much greater than that of short reads, which raises questions about the methodology used.

Response: The authors agree that long read sequencing would be more preferable to cover the entire length of repetitive regions, however, at the time sequencing was undertaken (~2017) short reads and 10x were the technologies most accessible to conservation researchers.

3. How accurate do the authors estimate their chosen method is?

Response: Based on the similarity of the size of the genome assembly and repeat content we can be confident that the genome is highly complete. Additionally our BUSCO analysis suggests a highly complete genome assembly (BUSCO completeness >95%). We agree that utilising long read sequencing would result in an increase in contiguity of the assembly.

4. Although I am not a real expert on genome assembly, I recommend that the authors add a brief section in the manuscript discussing the limitations of their findings. This section could explain the rationale behind their decision to work with a single individual and the reasons for choosing this particular method, such as the use of short reads.

Response: We have added a sentence into the results of the improvements that could be realised with the use of a long read technology. "Whilst the BUSCO scores suggest a highly complete genome, the use of long read data such as PacBio or

Oxford Nanopore would assist in increasing the contiguity of the assembly”

5. I also think it is important to briefly discuss what the authors think might explain why this animal has one and a half times as many protein-coding regions as the known marsupial genome. Could there be a biological or evolutionary reason for this?

Response: The increase in the number of protein coding genes could be due to the unavailability of transcriptome data for the species. Which would provide increased support for which of the predicted genes are actually expressed in the species. We have added a sentence in the results stating this. “In the future, generating transcriptomes for a variety of eastern bettong tissues would likely improve the accuracy of the annotation.”

Competing Interests: No competing interests were disclosed.

Reviewer Report 14 January 2025

<https://doi.org/10.5256/f1000research.173363.r351899>

© 2025 Hughes J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jonathan J Hughes 

University of California, Riverside, USA

The authors describe a whole-genome assembly generated with 10X sequencing for the eastern bettong, an ecologically important marsupial of conservation concern. They report their protocols for extraction, sequencing, QC, and annotation. The manuscript is clearly written and the methodology is both appropriate and reasonably detailed. I have a few small concerns that could be addressed:

1. State the parameters used when running any given software. If only default parameters were used for all software, then state that.
2. In Table 2, use commas consistently for reporting large numbers.
3. The number of scaffolds reported in the results (26,663) is different to that reported in Table 2 (26,623).
4. I have been able to locate most of the raw data and the assembled genome through NCBI, but the genome does not appear to be currently hosted on the Australian Genomes website as stated.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, phylogenetics, mammalogy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 Jan 2025

Luke Silver

1. State the parameters used when running any given software. If only default parameters were used for all software, then state that.

We have stated in the methods the settings used for running the software

2. In Table 2, use commas consistently for reporting large numbers.

We have used comma's consistently throughout the manuscript

3. The number of scaffolds reported in the results (26,663) is different to that reported in Table 2 (26,623).

We have updated the results to correctly report 26,623 as the number of scaffolds

4. I have been able to locate most of the raw data and the assembled genome through NCBI, but the genome does not appear to be currently hosted on the Australian Genomes website as stated.

We have uploaded the genome and annotation onto the Australian Genomes website

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 January 2025

<https://doi.org/10.5256/f1000research.173363.r351903>

© 2025 Osada N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Naoki Osada 

Hokkaido University, Sapporo, Japan

This manuscript presents the results of whole-genome sequencing and assembly of *Bettongia gaimardi*. The authors conducted genome assembly, quality evaluation, and annotation, including repetitive sequence annotation. I have verified that the assembled genome sequence and raw data are publicly available. Overall, this dataset is valuable for future research, and I have only some minor comments:

1. In the abstract, showing scaffold N50 would be helpful.
2. In the Introduction section, the phrase "10x short reads" might be misleading, as it could imply a 10-fold coverage of short reads. I suggest using "10x Genomics short reads" or a similar term for clarity.
3. In the Method section, the sentence "Further sequencing using ... was sequenced with 2 lanes ..." requires rephrasing for better readability.
4. It would be helpful if the authors provided additional details about the amount of sequencing data obtained, such as the number of reads or total bases.
5. I assume the authors used default settings for running the software, but it should be stated in the Method section.
6. In the Data availability section, the first sentence would be to: "The raw data are ...".

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics, molecular evolution, genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 Jan 2025

Luke Silver

Reviewer Comment: In the abstract, showing scaffold N50 would be helpful.

Author Response: We have added in the scaffold N50 value to the abstract

Reviewer Comment: In the Introduction section, the phrase “10x short reads” might be misleading, as it could imply a 10-fold coverage of short reads. I suggest using “10x Genomics short reads” or a similar term for clarity.

Author Response: We have rephrased as suggested

Reviewer Comment: In the Method section, the sentence "Further sequencing using ... was sequenced with 2 lanes ..." requires rephrasing for better readability.

Author Response: We have rephrased the sentence to “Additional sequencing using the 10X Chromium Genomics library prep with >50 kb size selection was performed using 2 lanes of HiSeq Xten (150 bp PE) (Illumina).” for clarity

Reviewer Comment: It would be helpful if the authors provided additional details about the amount of sequencing data obtained, such as the number of reads or total bases

Author Response: We have added the amount of reads produced in the first sentence of the results “Sequencing generated 185M reads of short read data and 149M reads of 10x Genomics data”

Reviewer Comment: I assume the authors used default settings for running the software, but it should be stated in the Method section.

Author Response: We have stated in the methods the settings used for running the software

Reviewer Comment: In the Data availability section, the first sentence would be to: “The raw data are ...”.

Author Response: We have updated the data availability section as suggested

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research