bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Title: Reference genome of the endangered eastern quoll (*Dasyurus viverrinus*)

Gabrielle A. Hartley[1], Stephen R. Frankenberg[2], Natasha M. Robinson[3], Anna J. MacDonald[4], Rodrigo K. Hamede[5], Christopher P. Burridge[5], Menna E. Jones[5], Tim Faulkner[6], Hayley Shute[6], Robert Brewster[7], Rachel O'Neill[1,8], Marilyn B. Renfree[2], Andrew J. Pask[2,9], Charles Y. Feigin[2,10*].

1. Institute for Systems Genomics, University of Connecticut, Storrs, CT 06269, USA.
2. School of BioSciences, The University of Melbourne, VIC 3010, Australia.
3. Fenner School of Environment & Society, Australian National University, Canberra, ACT 2601, Australia.
4. Research School of Biology, Australian National University, Canberra, ACT 2601, Australia; Current address: Australian Antarctic Division, Department of Climate Change, Energy, the Environment and Water, Kingston, TAS 7050, Australia.
5. School of Natural Sciences, University of Tasmania, Hobart, TAS 7005, Australia.
6. Australian Reptile Park & Aussie Ark, NSW 2250, Australia.
7. WWF-Australia, PO Box 528, Sydney, NSW 2001, Australia.
8. Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA.
9. Department of Sciences, Museums Victoria, Carlton, VIC 3053, Australia.
10. Department of Molecular Biology, Princeton University, Princeton, NJ 08540, USA.

*Correspondence: charles.feigin@unimelb.edu.au.

Abstract

The eastern quoll (*Dasyurus viverrinus*) is an endangered marsupial mesopredator native to Australia. Since the extirpation of the last mainland Australian populations in the late 20th century, wild populations of this species have been restricted to two islands at the far southern end of its historical range. Eastern quolls are the subject of captive breeding programs and attempts have recently been made to re-establish a population in mainland Australia through translocations. However, few resources currently exist to guide the genetic management of this species. Here, we present a chromosome-scale reference genome for the eastern quoll, along with gene annotations supported by multi-tissue transcriptomes. Through comparisons with related species, we find that our reference genome is among the most complete marsupial

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

assemblies currently available. Using this assembly, we infer the species' demographic history and identify potential evidence of a long-term decline beginning in the late Pleistocene. Finally, we identify a deletion at the *ASIP* locus that likely drives differences in pelage color between the eastern quoll and the closely related Tasmanian devil (*Sarcophilus harrisii*). The genomic resources we present are valuable new tools for evolutionary and conservation genomic studies.

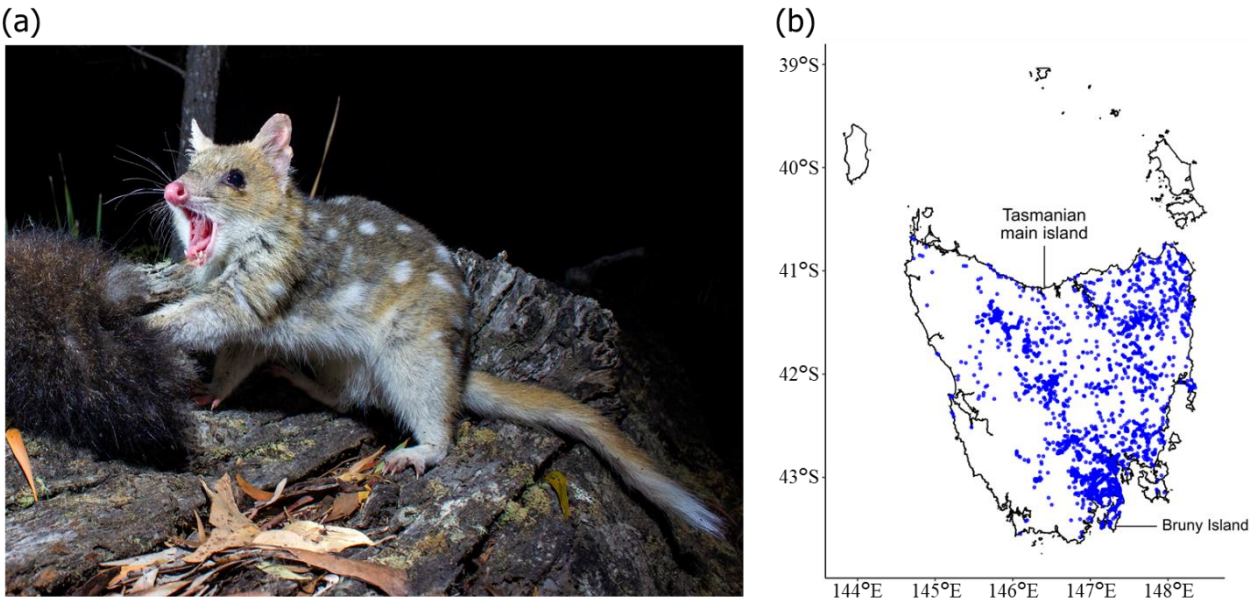
Introduction

The eastern quoll (*Dasyurus viverrinus*; Fig. 1a) is a marsupial carnivore native to Australia[3]. A nocturnal mesopredator, the eastern quoll's diet primarily consists of small mammals, reptile and invertebrate prey, as well as opportunistic scavenging on the remains of larger species[4]. The eastern quoll is morphologically, behaviorally, and ecologically distinct from the larger, parapatric spotted-tailed quoll (*Dasyurus maculatus*)[5]. Eastern quolls give birth to exceptionally altricial young that are comparable to mid-gestation fetuses of eutherian mammals[6; 7]. Indeed, neonates have structurally immature lungs and perform as much as 95% of their gas exchange through their skin[8-10]. Like other marsupials, their young travel from the birth canal into their mother's pouch on the day of birth and complete much of their development *ex utero*[3; 11].

Despite having once ranged across much of southeastern Australia, eastern quolls were likely extirpated from the mainland by the late 20th century[12]. Today, natural populations of this species are restricted to the state of Tasmania on two islands (the Tasmanian main island and Bruny Island; Fig. 1b), which lie at the southern end of their historical range. Moreover, several Tasmanian eastern quoll populations have also undergone significant and ongoing population declines in recent decades and the species was declared 'Endangered' by IUCN in 2016[13-15]. Suggested causes of decline include invasive predators and climatic fluctuation[16; 17].

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. 1.



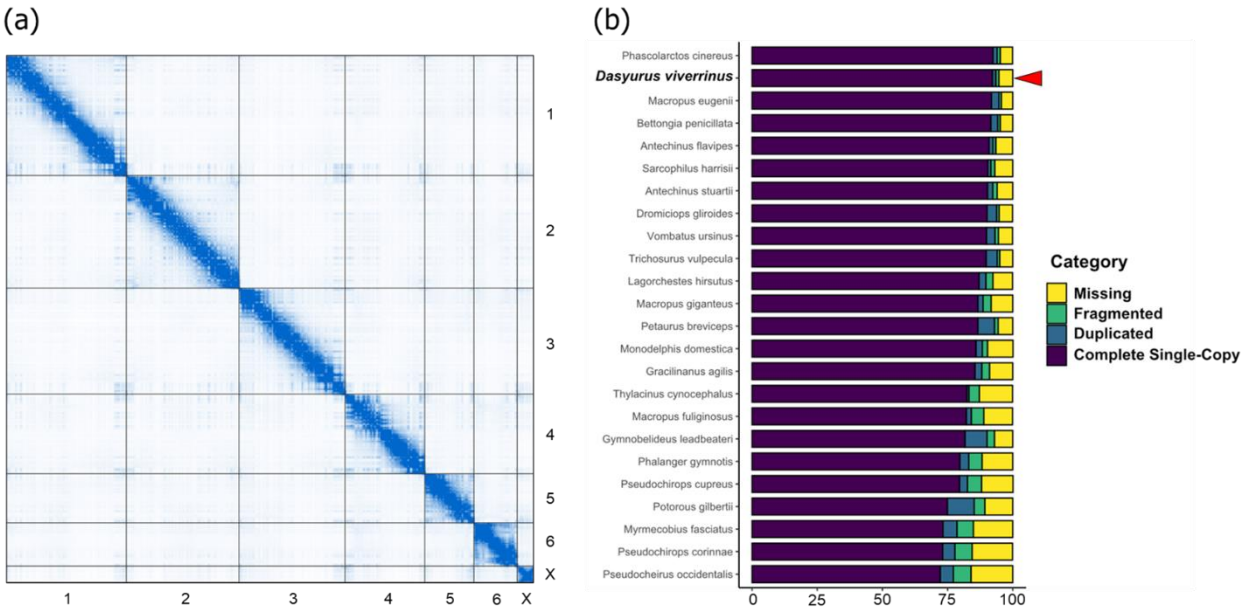
a) Photograph of an adult eastern quoll (photo credit Brett Vercoe). b) Map of the state of Tasmania, showing 50 years of eastern quoll sightings across the Tasmanian main island and Bruny Island recorded in the Tasmanian Natural Values Atlas. Individual sightings are shown as blue dots.

The eastern quoll has been flagged as one of 20 priority mammal species in the Australian Government's Threatened Species Strategy, with conservation efforts including population supplementation studies, an extensive captive breeding program spanning multiple sanctuaries, and fenced safe havens free of cats and foxes[18; 19]. The species has also recently been the subject of an ambitious project aimed at re-establishing wild populations in parts of its former mainland range, with two pilot releases of captive-bred animals in Booderee National Park on the south coast of New South Wales[20; 21]. However, potential challenges to the long-term fitness and adaptive viability of eastern quolls in both Tasmania and mainland Australia remain largely unexplored. The paucity of genomic resources also presents a barrier to conservation genetic management of the species[22]. To address this limitation, we present a high-quality eastern quoll reference genome with transcriptome-based gene annotations.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results and discussion

Fig. 2.



a) Contact map of Omni-C data against chromosome-scale scaffolds in the eastern quoll assembly. The enrichment level of chromosomal contacts is shown as blue pixel intensity b) Stacked bar plot comparing the recovery of mammalian BUSCO genes. Recovery in the eastern quoll is among the highest for sequenced marsupials, with low rates of duplicated or fragmented orthologs.

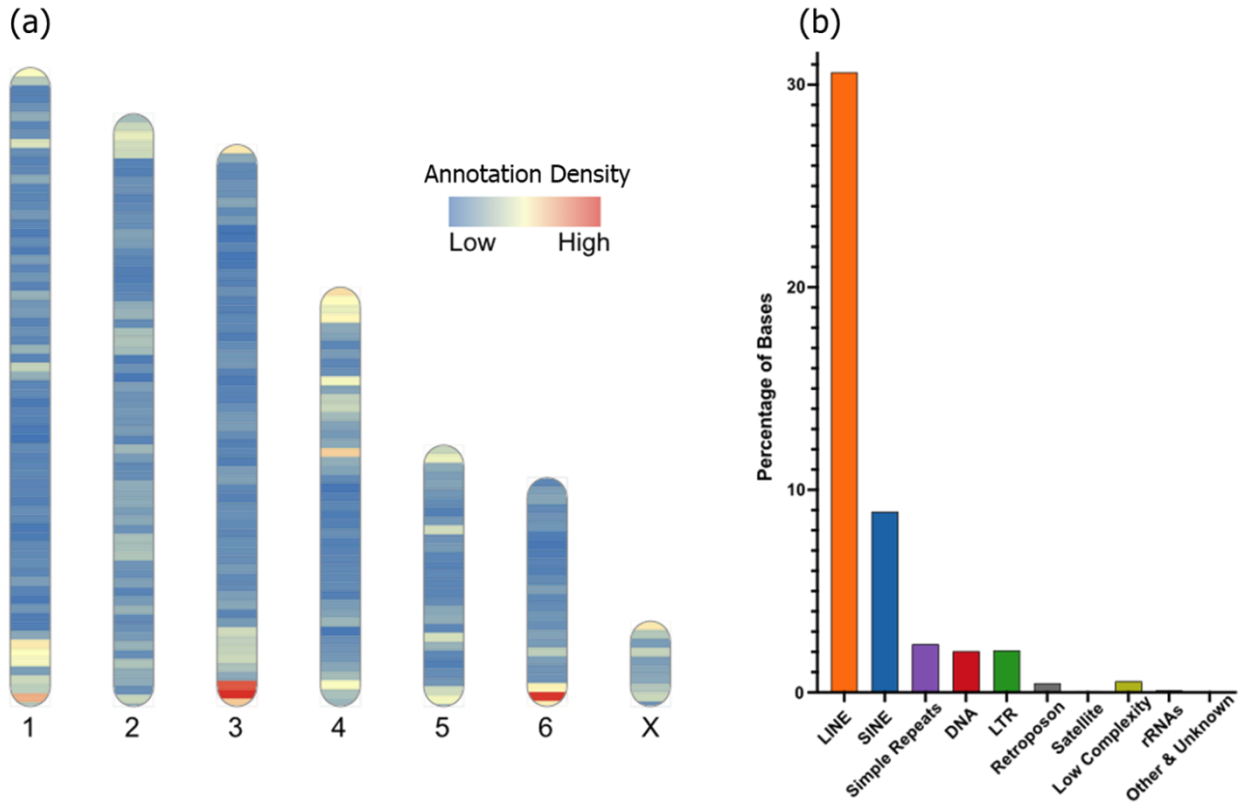
The eastern quoll reference genome

Eastern quoll samples were acquired opportunistically from an adult female postmortem, including several frozen tissues and live primary fibroblasts which were expanded in culture (Supplementary Table S1). Using these, we generated a chromosome-scale, de novo reference genome by assembling ~97.68 gigabases (Gb) of Pacific Biosciences high-fidelity (HiFi) long-reads and ~126 Gb of Omni-C long-range chromatin contact data. We called the resulting assembly DasViv_v1.0. The assembly size was ~3.14 Gb, comparable to that of related marsupials, and scaffolds were nearly free of internal gaps (Supplementary Table S2)[23; 24]. The assembly was composed of only 76 scaffolds, of which the seven largest corresponded to the conserved karyotype found across all known dasyurids: six autosomes plus the X chromosome (Fig. 2a)[25; 26]. Together, the seven chromosome-scale scaffolds accounted for 99.34% of the

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

total assembly size. Homology between eastern quoll chromosome-scale scaffolds and those of the related Tasmanian devil (*Sarcophilus harrisii*) and yellow-footed antechinus (*Antechinus flavipes*) was confirmed by the high overlap of orthologous gene annotations (>95%) and their similar total lengths (Supplementary Table S3)[23; 24]. Recovery of complete single-copy mammalian BUSCOs (Benchmarking Universal Single-Copy Orthologs) was 92.2%, second only to the koala (*Phascolarctos cinereus*) among marsupial reference genomes currently available on NCBI (Fig. 2b). Moreover, rates of duplicated and fragmented BUSCO genes were low (1.3% and 1.2%, respectively), reinforcing the completeness and integrity of our assembly.

Fig. 3.



a) Heat map illustrating the density of annotated genes across eastern quoll chromosome-scale scaffolds. b) Bar plot showing the distribution of annotated repeats by class.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Genome annotation

To accompany our assembly, we generated gene annotations by combining evidence from transcriptome data generated from five tissues (Supplementary Table S1), homologous proteins from related marsupials (Supplementary Table S4), as well as *ab initio* predictions. In total, we generated 29,622 gene models (Fig. 3a)[23; 24; 27-29].

We also annotated and characterized the repeat content of the eastern quoll genome using RepeatMasker[30]. In total, ~1.476 gigabases were masked as repetitive (47.2% of the assembly), comprising mainly LINEs (30.61%) and SINEs (8.93%; Fig. 3b). L1 repeats constituted the most abundant LINEs in the assembly (20.15%) and MIRs constituted the most abundant SINEs (7.67%). Notably, we identified a small fraction of bases (7,906 in total) corresponding to putative KERV elements, an endogenous retrovirus that has undergone radical expansion in the kangaroo genus *Macropus* (Supplementary Table S5)[31].

Historical demography

Australia has experienced substantial ecological changes over recent geological epochs. For instance, the late Pleistocene saw extensive climatic changes, the periodic isolation of Tasmania from the mainland, and the arrival and dispersal of humans into Australia[1; 32; 33]. This period also marked the extinction of the remaining megafauna (terrestrial vertebrates >40 kilograms)[34]. However, the effects that these potential historical stressors may have had on eastern quoll populations are poorly explored. Therefore, we next sought to examine historical trends in eastern quoll effective population size (N_e) with multiple sequentially Markovian coalescent (MSMC) analysis, using a mutation rate for the closely-related Tasmanian devil measured from parent-offspring trios and a generation time of 2 years[35; 36].

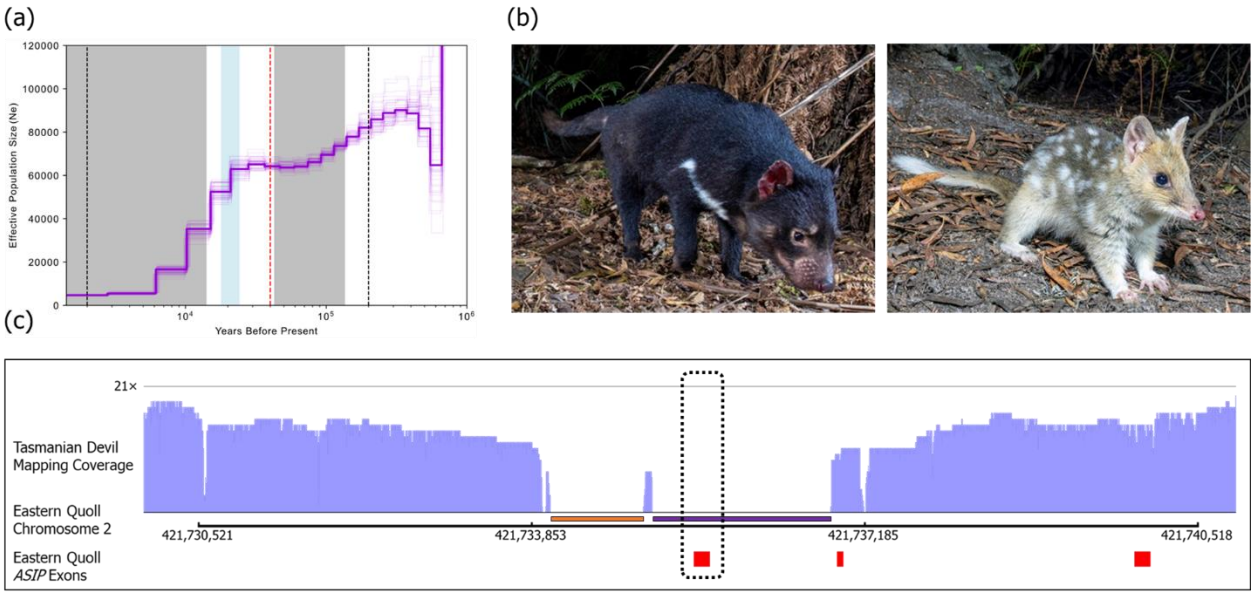
Our analysis indicated long-term decline in N_e, comprised of two phases (Fig. 4a). The first of the inferred decline phases is indicated to have begun ~300 thousand years ago (kya) and continued until approximately 70 kya. However, recent work on the properties of MSMC analysis

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in the related Tasmanian devil suggest that accuracy can decline markedly during very recent and very ancient periods (less than 1,000 and greater than 100,000 generations ago, respectively)[2]. Thus, we are cautious in over-interpreting the precise start of this decline phase, which falls outside of this range in our analysis (~2-200 kya). The more recent inferred decline began approximately 28 kya and extended into the Holocene (Fig. 4a). The arrival of humans in Tasmania ~40 kya seems an unlikely candidate to have driven this decline, as they co-existed with eastern quolls for some 12 kyr with apparently little impact on N_e[33]. Curiously, our analyses indicate that both declines may have initiated during periods in which Tasmania was connected to mainland Australia due to lowered sea levels[37], despite the potential for increased gene flow between Tasmanian and mainland populations. It is possible that unfavorable climates during these colder periods created suboptimal environmental conditions in Tasmania and it is unclear whether the Bassian Plain (the land bridge connecting Tasmania and mainland Australia) environments were conducive to gene flow. Indeed, bare-nosed wombat (*Vombatus ursinus*) populations with past territorial connections via the Bassian (i.e. on Tasmania, Flinders Island, and mainland Australia) appear to have maintained genetic isolation[38]. Future resequencing studies of eastern quolls from Tasmanian, Bruny Island, and preserved historical mainland specimens may provide further insights into historical gene flow and population structure, which can confound MSMC and related analyses.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. 4.



a) Step plot illustrating inferred changes in eastern quoll effective population size (Ne) over time. Grey regions indicate periods in which Tasmania was separated from mainland Australia (~135-43 kya and ~14 kya to present). The blue region indicates the coldest period of the Last Glacial Maximum (~24-18 kya)[1]. The dashed red line represents the arrival of humans in Tasmania (~40 kya). Area between the black dashed lines at 2,000 and 200,000 years represent the window within which MSMC analysis is expected to be accurate based on previous studies (1-100 thousand generations before present)[2]. b) Photographs comparing pelage color of the Tasmanian devil and eastern quoll (photo credits Brett Vercoe). c) Mapping coverage of Tasmanian devil long reads (in light purple) across the *ASIP* locus in the eastern quoll genome. Exons of the eastern quoll *ASIP* gene are shown as red blocks. The deletion region is indicated by the absence of mapped Tasmanian devil reads underlined by a purple bar and encompassing exon 1 (dashed box) of *ASIP*. The putative eastern quoll insertion is underlined by an orange bar, upstream of the Tasmanian devil deletion.

Comparative genomics of pigmentation

The closest living relative to the quolls (genus *Dasyurus*) is the Tasmanian devil. Notably, quolls and Tasmanian devils differ markedly in their coat color and patterning. The quoll background coat color consists of brown dorsal fur, punctuated by spots of white fur and white-to-yellow ventral fur. In contrast, Tasmanian devils have nearly uniform brown/black fur on both dorsum and ventrum, with many individuals bearing unpigmented white patches on the chest, shoulders and/or base of the tail (Fig. 4b). The color of the background dorsal fur of eastern quolls is produced by alternating bands of yellow pheomelanin and dark brown eumelanin in

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

individual hair shafts, a pattern called “agouti”[39]. The agouti pattern, which is common among diverse mammals and likely ancestral among dasyurid marsupials, is known to be regulated in part by the interactions of two key proteins, the melanocortin 1 receptor (MC1R) which promotes eumelanin production and agouti signaling peptide (ASIP), which antagonizes MC1R leading to pheomelanin synthesis[40]. Melanistic morphs of many animals have been shown to be caused by loss-of-function (LOF) mutations in the *ASIP* gene, including multiple independent polymorphisms among *Peromyscus* mice and in wild cats[41; 42]. Therefore, we hypothesized that the dark, eumelanin-bearing hair in Tasmanian devils might therefore have evolved through a comparable mechanism.

We first sought to compare ASIP orthologs from our eastern quoll genome (DasViv_v1.0) and the reference genomes of the Tasmanian devil (mSarHar1.11) with those from several other dasyuromorph species by aligning their coding sequences (Supplementary File S1 and Supplementary Table S6)[23; 24; 43-45]. Notably, while extracting orthologs from each reference genome, we failed to identify the first exon of ASIP from mSarHar1.11, despite the assembly being of high-quality and containing few gaps[24]. To rule out the possibility that this region is present in the Tasmanian devil genome, but not reliably assembled during the construction of mSarHar.1.11, we aligned the nanopore long reads used to produce this genome against our eastern quoll genome. A histogram of read coverage showed two regions in which no Tasmanian devil reads mapped (**Fig. 4c**). Among these, the second region (~1.8kb long) encompassed the entirety of *ASIP* exon 1. Together, these observations indicated the presence of a potential deletion of the first exon of this gene, including the start codon, supporting the notion of an *ASIP* LOF underlying the Tasmanian devil’s melanistic coat color.

To confirm that the putative *ASIP* exon 1 deletion was not unique to the individual animal used to produce the mSarHar1.11 assembly, we next extracted the orthologous genomic region from another Tasmanian devil genome assembly available on NCBI, (SarHar_Dovetail_2.0), which was generated from a different animal and used different sequencing and assembly

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

approaches. Additionally, we extracted this region from a third dasyurid species, the yellow-footed antechinus (AdamAnt_v2) for comparison. Interestingly, alignment of these regions revealed that the first, upstream region which had shown zero Tasmanian devil read mapping coverage likely represents an eastern quoll-specific insertion, as this sequence was not found in either Tasmanian devil or yellow-footed antechinus (Supplementary File S2). However, consistent with our previous observations, our alignments also revealed identical deletion breakpoints for the putative deletion region encompassing *ASIP* exon 1 in both Tasmanian devil individuals (Supplementary File S2).

Taken together, these results provide strong evidence for the fixation of a loss-of-function mutation in the Tasmanian devil ortholog of *ASIP*. Given the known effects of *ASIP* LOF in other mammals, such a mutation is expected to cause a melanistic (non-agouti) pigment phenotype, thus underpinning the stark difference in background coat color between the Tasmanian devil and quolls.

Conclusions

Here, we present new, high-quality genomic resources for the endangered eastern quoll. Our chromosome-scale genome assembly closely matches the known chromosome complement of the species which is conserved among dasyurid marsupials. The completeness and contiguity of the assembly exceeds that of most existing marsupial reference genomes as indicated by high BUSCO recovery and low gap percentage. Our comparative analysis of core pigmentation loci reveals the probable basis of pelage variation between the eastern quoll and its close relative the Tasmanian devil, and genome annotations provided here represent a tool for future comparative studies. Moreover, we identify preliminary evidence of historical demographic declines, reinforcing the value of future population genomic studies aimed at defining diversity, population structure, and genetic load in this species.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Data availability

The eastern quoll reference genome and all sequence data used in its generation are available on NCBI under BioProject PRJNA758704. Transcriptome data used in gene annotations are available under BioProject PRJNA963007. Gene annotation GFF files and all original code used in this study can be accessed in a permanent FigShare repository: https://doi.org/10.26188/23501301.v1.

Acknowledgements

We thank Revive & Restore for their funding and organizational support for this project with special thanks to Bridget Baumgartner, Cantata Bio LLC for computational support and sequencing services with special thanks to Mark Daly and Jordan Zhang, Karrie Rose for collecting tissues, Sandy Ingleby and Emma Peel for depositing samples at Australian Museum, the Computational Biology Core in the Institute for Systems Genomics at the University of Connecticut for the use of the Xanadu HPC, Brett Vercoe for the use of his eastern quoll and Tasmanian devil photographs and Elise Ireland for proofreading the manuscript.

Author contributions

C.Y.F., A.J.M., N.M.R., R.K.H., C.P.B. and M.E.J. conceived the study and acquired funding. C.Y.F. performed assembly quality, gene annotation, historical demographic, and comparative genomic analyses. G.A.H. performed genomic repeat characterization. S.R.F. isolated dermal fibroblasts. R.B., T.F., H.S. and N.M.R. acquired samples, M.B.R, R.O. and A.J.P. facilitated sample transport. C.Y.F., G.A.H. and S.R.F. wrote the manuscript with input from all authors.

Funding

This project was supported by a Wild Genomes grant to C.Y.F., A.J.M., N.M.R., R.K.H., C.P.B. and M.E.J. from Revive & Restore (Contract no. 2020-017). C.Y.F. was supported by

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

NIH NRSA fellowship F32GM139240-01. G.A.H. and R.J.O. are supported by a grant from the US National Institutes of Health R01GM123312-02.

Author notes

Conflicts of interest: None declared.

Materials and methods

Eastern quoll sightings map

An approximation of the eastern quoll's range within Tasmania was produced by accessing curated sighting data from studies recorded in the Tasmanian Natural Values Atlas (https://www.naturalvaluesatlas.tas.gov.au/), filtering to those within the 50-year period from 1/1/1973 to 1/1/2023. Locations were visualized on a map of Tasmania using a custom R script.

Samples

Samples of kidney, heart, liver, and spleen tissue were opportunistically collected as secondary use during the necropsy of an adult, brown morph female quoll named Manda, originating from the Aussie Ark captive breeding sanctuary and which had been euthanized for veterinary reasons at Taronga Zoo (Sydney, NSW Australia). Excess tissue samples from this individual were deposited in the Australian Museum along with the pelt and skeleton under accession number M.52159. Subsampled tissues were snap frozen in liquid nitrogen immediately upon collection and were stored at -80°C until used.

Additionally, slices of toe pad were taken and used to isolate primary dermal fibroblasts in culture. Briefly, small pieces of footpad tissue were scored into the base of a 30-mm plastic culture dish. 0.5 mL of DMEM (Gibco 10569044) containing 10% fetal bovine serum and antibiotic/antimycotic) was carefully added to the dish without disrupting the tissue pieces and incubated at 33˚C in 5% CO2. The next day, medium was carefully added up to a total of 2 mL, then replaced every two days. Once the fibroblast outgrowths were 90% confluent, cells were

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

expanded by trypsinization to a T25 flask, then five aliquots cryopreserved in fibroblast medium containing 10% DMSO. Cells were expanded in culture for eight passages before pellets were collected and frozen at -80°C for RNA-seq.

Genome sequencing and assembly

Samples of liver tissue were used by Dovetail Genomics (Scotts Valley, CA USA) for genome sequencing and assembly. High molecular weight DNA was extracted using the Qiagen Blood & Cell Culture DNA miniprep kit (catalogue no: 13323) and quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). PacBio (Menlo Park, CA, USA) SMRTbell libraries were constructed using the SMRTbell Express Template Prep Kit 2.0. Libraries were bound to polymerase using the Sequel II Binding Kit 2.0 and were sequenced on Sequel II 8M SMRT cells. This yielded approximately 97.68 gigabases of sequence comprised of 5,959,830 HiFi reads with an average length of ~16,390nt. HiFi reads were then assembled into contigs using Hifiasm v0.15.4-r343 with default parameters[46]. Unpurged duplicates were then removed using purge_dups (v1.2.5) using automated thresholds[47].

A 3D chromatin contact library was produced from additional liver subsamples using Dovetail's Omni-C kit and was sequenced on an Illumina HiSeqX, generating approximately 126 gigabases of sequence (~420 million read pairs in 2x150bp format)**.** Chromosome-scale scaffolds were then produced by first aligning Omni-C libraries to the *de novo* contigs with BWA mem and the HiRise Pipeline was used to make contig joins and break putative misjoins[48; 49]. Contact data were visualized by processing alignments with Pairtools (v1.0.2) and creating a .hic file with juicer_tools pre (v1.22.01) which was loaded into Juicebox (v1.11) and exported to generate a contact map[50-52].

Genome assessment

Genome assembly metrics related to contiguity, base composition and contig/scaffold lengths were generated using the stats.sh script contained in the bbmap v39.01package[53].

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Assembly completeness was further assessed via the recovery of benchmarking orthologs using BUSCO v5.4.6 in genome mode with the mammalia_odb10 database[54]. BUSCO gene recovery in the eastern quoll assembly was compared to that of all other marsupial whole-genome assemblies hosted on the NCBI Genomes database and labelled as the representative genome for their species as of April 18th 2023[23; 24; 28; 29; 43-45; 55-58].

Inference of chromosome homology

Homology between eastern quoll chromosomes and those of the related Tasmanian devil and yellow-footed antechinus were inferred using gene annotation overlap. Briefly, homologs of Tasmanian devil genes and antechinus genes respectively were identified via liftover to the eastern quoll genome using Liftoff v1.6.3 (parameters -d 4 -a 0.9 -s 0.9)[59]. Annotations were compared between chromosome-scale scaffolds in the eastern quoll and each reference dasyurid's genomes. Chromosomes sharing >= 95% of their gene content were deemed to be homologous. Annotation-based homology assignment was supported by the nearly identical relative sizes of presumptive homologous chromosomes between species, consistent with the exceptionally conserved karyotype previously reported among all examined dasyurids[25; 26].

RNA-sequencing and gene annotation

Frozen samples of eastern quoll heart, kidney, liver, spleen and a pellet of cultured dermal fibroblasts were provided to Psomagen Inc (Rockville, MD USA). RNA-seq libraries were prepared using the TruSeq Stranded mRNA LT sample prep kit (15031047 Rev. E) and sequenced on an Illumina NovaSeq 6000 in 2x150bp format. Residual adapters were removed, and reads were trimmed and filtered for quality using Trimmomatic v0.39 (parameters: SLIDINGWINDOW:5:15, MINLEN:50, AVGQUAL:20, ILLUMINACLIP:2:30:10)[60]. After processing, libraries ranged from approximately 29 to 40 million retained read pairs.

To annotate genes in the eastern quoll genome, filtered RNA-seq reads from all five tissues together with RefSeq homologous proteins from seven other marsupial species were provided

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

to the funannotate v1.8.14 pipeline which integrated these with augustus *ab initio* predictions to infer gene models[61]. Within the funannotate pipeline, the following modules were used: train (parameters: --no_trimmomatic, --stranded RF), predict (--augustus_species human, --busco_seed_species human, --optimize_augustus, --busco_db mammalia, --ploidy 1, --organism other, --min_intronlen 10, --max_intronlen 100000, --repeats2evm), and update (default parameters). This approach produced models for 29,622 genes, including 31,319 protein-coding transcripts. Of these, we were able to assign gene symbols to 14,293, based on high-confidence 1-to-1 orthology inferred using eggNOG-mapper's Mammalia database, a figure comparable to that of the Tasmanian devil and yellow-footed antechinus annotations produced by RefSeq (15,613 and 15,573 respectively)[62]. The density of annotated genes was visualized using the RIdeogram, providing a histogram of gene counts across 10 megabase (mb) windows on each chromosome[63].

Repeat masking and annotation

Repeats in the eastern quoll assembly were annotated with RepeatMasker (v4.1.3) using the NCBI BLAST derived search engine rmblast, sensitive settings (-s), and a combined Dfam (v3.6) and Repbase (v20181026) repeat database for marsupials (-species metatheria)[30; 64]. The repeat annotations produced were used to hardmask the assembly using Bedtools (v2.29.0)[65]. Subsequently, RepeatMasker was performed on the hardmasked assembly using a custom repeat library to identify KERV long terminal repeat elements (LTRs) and other marsupial-derived satellites absent in the above repeat databases. The resulting repeat annotations were combined and summarized using the RepeatMasker utility script buildSummary.pl.

Historical demography

Haplotype-phased variants for the reference eastern quoll were identified by parsing aligned Omni-C reads with samtools and pairtools and passing them to the Google Deepvariant pipeline which was run with default settings to call and filter variants[50; 66; 67]. Variant phasing was

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

307 performed using HapCUT2 (v1.2)[68]. The three largest chromosomes in the eastern quoll
308 genome exceed the maximum scaffold size limitations in the .bai alignment index required for
309 several steps in data processing. Therefore, a second copy of the assembly was made, in which
310 each of these scaffolds was split into two equal halves using samtools faidx. HiFi reads were
311 then mapped against this copy using minimap2 v 2.24-r1122 (parameters -a -x map-hifi) and
312 filtered with samtools view (parameters -h -q 20 -F 2304)[67; 69]. The alignment file was then
313 reduced to only chromosome-scale scaffolds corresponding to the six autosomes and the
314 average mapping depth was calculated with samtools depth. A mapping coverage mask file was
315 then generated using the bamcaller.py included in the MSMC2 package, providing the average
316 coverage depth calculated above. Additionally, the RepeatMasker bed file was used to identify
317 construct a mappability mask which excluded repetitive regions. These masks were provided
318 together with heterozygous SNPs with QUAL >= 10 to the msmc-tools script
319 generate_multihetsep.py (https://github.com/stschiff/msmc-tools). Finally, MSMC2 v2.1.4 was
320 used to infer the eastern quoll's demographic history with 50 EM-iterations (-i 50) and time
321 pattern (-p) 1*2+20*1+1*5, reducing the number of free parameters relative to default settings
322 and grouping the last five time segments to reduce overfitting at extremely ancient time periods.
323 Data were then plotted using the previously-reported generation time of two years and a per-
324 site, per-generation mutation rate of 5.95e-9[35; 36].

325 **Comparative genomics**

326 The *ASIP* sequence was annotated in all available dasyuromorph reference genomes on
327 NCBI via lift-over from the annotation of the yellow-footed antechinus (see: 'Inference of
328 chromosome homology'). Sequences were then extracted with gffread, translated into the
329 conserved reading frame using MACSE v2, and then aligned using the MAFFT web server
330 using default parameters[70-72].

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To generate coverage histograms, Tasmanian devil long reads used to generate mSarHar1.11 were accessed from NCBI SRA (Accession: ERR3930603) and aligned against the eastern quoll genome with minimap2 v2.24 (parameter: -x map-ont). Pileups of reads mapped to the *ASIP* locus were then visualized using IGV[73].

To perform alignments of genomic sequence surrounding the *ASIP* exon 1 region, *ASIP* exon 2 from the Tasmanian devil was aligned against the eastern quoll (DasViv_v1.0), yellow-footed antechinus (AdamAnt_v2), and Tasmanian devil (mSarHar1.11 and SarHar_Dovetail_2.0) assemblies using blastn with default settings[74]. Samtools faidx was then used to extract the region containing the putative deletion region (i.e., between conserved flank sequences found in all species). These regions were then and were then aligned using the MAFFT web browser with default parameters[67; 72].

References

1. De Deckker P, Moros M, Perner K, Blanz T, Wacker L, Schneider R, Barrows TT, O'Loingsigh T, Jansen E. 2020. Climatic evolution in the australian region over the last 94 ka - spanning human occupancy -, and unveiling the last glacial maximum. Quaternary Science Reviews. 249:106593.
2. Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK, Ruiz-Aravena M, Ryder O, McCallum HI, Jones ME et al. 2019. Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in tasmanian devils. Molecular Biology and Evolution. 36(12):2906-2921.
3. Jones ME, Rose RK. 2001. Dasyurus viverrinus. Mammalian Species. (677):1-9.
4. Jones ME, Barmuta LA. 1998. Diet overlap and relative abundance of sympatric dasyurid carnivores: A hypothesis of competition. Journal of Animal Ecology. 67(3):410-421.
5. Jones ME, Barmuta LA. 2000. Niche differentiation among sympatric australian dasyurid carnivores. Journal of Mammalogy. 81(2):434-447.
6. Cook LE, Newton AH, Hipsley CA, Pask AJ. 2021. Postnatal development in a marsupial model, the fat-tailed dunnart (sminthopsis crassicaudata; dasyuromorphia: Dasyuridae). Communications Biology. 4(1):1028.
7. Ferner K, Schultz JA, Zeller U. 2017. Comparative anatomy of neonates of the three major mammalian groups (monotremes, marsupials, placentals) and implications for the ancestral mammalian neonate morphotype. J Anat. 231(6):798-822.
8. Ferner K. 2021. Early postnatal lung development in the eastern quoll (dasyurus viverrinus). The Anatomical Record. 304(12):2823-2840.
9. Ferner K. 2021. Development of the skin in the eastern quoll (dasyurus viverrinus) with focus on cutaneous gas exchange in the early postnatal period. J Anat. 238(2):426-445.
10. Ferner K, Zeller U, Renfree MB. 2009. Lung development of monotremes: Evidence for the mammalian morphotype. Anatomical record (Hoboken, NJ : 2007). 292(2):190-201.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

11. Tyndale-Biscoe H, Renfree M. 1987. Reproductive physiology of marsupials. Cambridge: Cambridge University Press.
12. Frankham GJ, Thompson S, Ingleby S, Soderquist T, Eldridge MDB. 2017. Does the 'extinct' eastern quoll *dasyurus viverrinus* persist in barrington tops, new south wales? Australian Mammalogy. 39(2):243-247.
13. Burbidge AA, Woinarski J. 2016. Dasyurus viverrinus. The IUCN Red List of Threatened Species 2016. e.T6296A21947190.
14. Fancourt BA, Hawkins CE, Nicol SC. 2013. Evidence of rapid population decline of the eastern quoll (*dasyurus viverrinus*) in tasmania. Australian Mammalogy. 35(2):195-205.
15. Cunningham CX, Aandahl Z, Jones ME, Hamer R, Johnson CN. 2023. Regional patterns of continuing decline of the eastern quoll. Australian Mammalogy. 45(2):151-159.
16. Fancourt BA, Bateman BL, VanDerWal J, Nicol SC, Hawkins CE, Jones ME, Johnson CN. 2015. Testing the role of climate change in species decline: Is the eastern quoll a victim of a change in the weather? PLoS One. 10(6):e0129420.
17. Fancourt BA. 2016. Diagnosing species decline: A contextual review of threats,causes and future directions for management and conservation of the eastern quoll. Wildlife Research. 43(3):197-211.
18. Hamer RP, Robinson N, Brewster R, Barlow M, Guinane M, Humphrey M, Mifsud A, Hamilton DG, Kutt AS. 2023. Not waiting for the death knell: A pilot study to examine supplementation and survivorship in a declining population of tasmanian eastern quoll (*dasyurus viverrinus*). Australian Mammalogy. 45(2):171-180.
19. Fraser H, Legge SM, Garnett ST, Geyle H, Silcock J, Nou T, Collingwood T, Cameron KA, Fraser F, Mulcahy A et al. 2022. Application of expert elicitation to estimate population trajectories for species prioritized in australia's first threatened species strategy. Biological Conservation. 274:109731.
20. Robinson NM, Blanchard W, MacGregor C, Brewster R, Dexter N, Lindenmayer DB. 2020. Finding food in a novel environment: The diet of a reintroduced endangered meso-predator to mainland australia, with notes on foraging behaviour. PLOS ONE. 15(12):e0243937.
21. Robinson NM, Blanchard W, Macgregor C, Brewster R, Dexter N, Lindenmayer DB. 2021. Can evolutionary theories of dispersal and senescence predict postrelease survival, dispersal, and body condition of a reintroduced threatened mammal? Ecology and Evolution. 11:1002 - 1012.
22. Brandies P, Peel E, Hogg CJ, Belov K. 2019. The value of reference genomes in the conservation of threatened species. Genes (Basel). 10(11):846.
23. Tian R, Han K, Geng Y, Yang C, Shi C, Thomas PB, Pearce C, Moffatt K, Ma S, Xu S et al. 2022. A chromosome-level genome of antechinus flavipes provides a reference for an australian marsupial genus with male death after mating. Molecular ecology resources. 22(2):740-754.
24. Stammnitz MR, Gori K, Kwon YM, Harry E, Martin FJ, Billis K, Cheng Y, Baez-Ortega A, Chow W, Comte S et al. 2023. The evolution of two transmissible cancers in tasmanian devils. Science. 380(6642):283-293.
25. Deakin JE. 2018. Chromosome evolution in marsupials. Genes (Basel). 9(2):72.
26. Rofe R, Hayman D. 1985. G-banding evidence for a conserved complement in the marsupialia. Cytogenetics and cell genetics. 39(1):40-50.
27. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A et al. 2007. Genome of the marsupial monodelphis domestica reveals innovation in non-coding sequences. Nature. 447(7141):167-177.
28. Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, Grueber CE, Cheng Y, Whittington CM, Dennison S et al. 2018. Adaptation and conservation insights from the koala genome. Nature genetics. 50(8):1102-1111.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

29. Tian R, Han K, Geng Y, Yang C, Guo H, Shi C, Xu S, Yang G, Zhou X, Gladyshev VN et al. 2021. A chromosome-level genome of the agile gracile mouse opossum (gracilinanus agilis). Genome Biol Evol. 13(8).
30. Tarailo-Graovac M, Chen N. 2009. Using repeatmasker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics. Chapter 4:Unit 4.10.
31. Ferreri GC, Brown JD, Obergfell C, Jue N, Finn CE, O'Neill MJ, O'Neill RJ. 2011. Recent amplification of the kangaroo endogenous retrovirus, kerv, limited to the centromere. Journal of virology. 85(10):4761-4771.
32. Clarkson C, Jacobs Z, Marwick B, Fullagar R, Wallis L, Smith M, Roberts RG, Hayes E, Lowe K, Carah X et al. 2017. Human occupation of northern australia by 65,000 years ago. Nature. 547(7663):306-310.
33. Tobler R, Rohrlach A, Soubrier J, Bover P, Llamas B, Tuke J, Bean N, Abdullah-Highfold A, Agius S, O'Donoghue A et al. 2017. Aboriginal mitogenomes reveal 50,000 years of regionalism in australia. Nature. 544(7649):180-184.
34. Hocknull SA, Lewis R, Arnold LJ, Pietsch T, Joannes-Boyau R, Price GJ, Moss P, Wood R, Dosseto A, Louys J et al. 2020. Extinction of eastern sahul megafauna coincides with sustained environmental deterioration. Nature Communications. 11(1):2250.
35. Burbidge A, Harrison P, Woinarski J. 2014. The action plan for australian mammals 2012.
36. Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St. Leger J, Shao C et al. 2023. Evolution of the germline mutation rate across vertebrates. Nature. 615(7951):285-291.
37. Lambeck K, Chappell J. 2001. Sea level change through the last glacial cycle. Science. 292(5517):679-686.
38. Martin A, Carver S, Proft K, Fraser TA, Polkinghorne A, Banks S, Burridge CP. 2019. Isolation, marine transgression and translocation of the bare-nosed wombat (vombatus ursinus). Evolutionary applications. 12(6):1114-1123.
39. Dry FW. 1926. The coat of the mouse (mus musculus). Journal of Genetics. 16(3):287-340.
40. McRobie HR, King LM, Fanutti C, Symmons MF, Coussons PJ. 2014. Agouti signalling protein is an inverse agonist to the wildtype and agonist to the melanic variant of the melanocortin-1 receptor in the grey squirrel (sciurus carolinensis). FEBS Letters. 588(14):2335-2343.
41. Kingsley EP, Manceau M, Wiley CD, Hoekstra HE. 2009. Melanism in peromyscus is caused by independent mutations in agouti. PLoS One. 4(7):e6435.
42. Schneider A, David VA, Johnson WE, O'Brien SJ, Barsh GS, Menotti-Raymond M, Eizirik E. 2012. How the leopard hides its spots: Asip mutations and melanism in wild cats. PLoS One. 7(12):e50386.
43. Peel E, Silver L, Brandies P, Hayakawa T, Belov K, Hogg CJ. 2022. Genome assembly of the numbat (myrmecobius fasciatus), the only termitivorous marsupial. bioRxiv.2022.2002.2013.480287.
44. Feigin C, Frankenberg S, Pask A. 2022. A chromosome-scale hybrid genome assembly of the extinct tasmanian tiger (thylacinus cynocephalus). Genome Biology and Evolution. 14(4).
45. Brandies PA, Tang S, Johnson RSP, Hogg CJ, Belov K. 2020. The first antechinus reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies. Gigabyte. 2020:0.
46. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods. 18(2):170-175.
47. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 36(9):2896-2898.
48. Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 25(14):1754-1760.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

49. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26(3):342-350.
50. Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, Imakaev M, Venev SV. 2023. Pairtools: From sequencing data to chromosome contacts. bioRxiv.2023.2002.2013.528389.
51. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. Cell Syst. 3(1):95-98.
52. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for hi-c contact maps with unlimited zoom. Cell Syst. 3(1):99-101.
53. Bushnell B. 2014. Bbmap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).
54. Seppey M, Manni M, Zdobnov EM. 2019. Busco: Assessing genome assembly and annotation completeness. Methods Mol Biol. 1962:227-245.
55. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP et al. 2017. De novo assembly of the aedes aegypti genome using hi-c yields chromosome-length scaffolds. Science. 356(6333):92-95.
56. Peel E, Silver L, Brandies P, Hogg CJ, Belov K. 2021. A reference genome for the critically endangered woylie, bettongia penicillata ogilbyi. GigaByte. 2021:gigabyte35.
57. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 592(7856):737-746.
58. Feigin CY, Moreno JA, Ramos R, Mereby SA, Alivisatos A, Wang W, van Amerongen R, Camacho J, Rasweiler JJ, Behringer RR et al. 2023. Convergent deployment of ancestral functions during the evolution of mammalian flight membranes. Science Advances. 9(12):eade7511.
59. Shumate A, Salzberg SL. 2021. Liftoff: Accurate mapping of gene annotations. Bioinformatics. 37(12):1639-1643.
60. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for illumina sequence data. Bioinformatics. 30(15):2114-2120.
61. Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation (v1.8.1). Zenodo.
62. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. Eggnog-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Molecular Biology and Evolution. 38(12):5825-5829.
63. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. Rideogram: Drawing svg graphics to visualize and map genome-wide data on the idiograms. PeerJ Computer science. 6:e251.
64. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AFA, Wheeler TJ. 2016. The dfam database of repetitive DNA families. Nucleic acids research. 44(D1):D81-D89.
65. Quinlan AR, Hall IM. 2010. Bedtools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841-842.
66. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT et al. 2018. A universal snp and small-indel variant caller using deep neural networks. Nature biotechnology. 36(10):983-987.
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and samtools. Bioinformatics. 25(16):2078-2079.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.06.556354; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

68. Edge P, Bafna V, Bansal V. 2017. Hapcut2: Robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res. 27(5):801-812.
69. Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094-3100.
70. Pertea G, Pertea M. 2020. Gff utilities: Gffread and gffcompare [version 1; peer review: 3 approved]. F1000Research. 9(304).
71. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. Macse v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution. 35(10):2582-2584.
72. Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. Mafft: A novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Research. 30(14):3059-3066.
73. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nature biotechnology. 29(1):24-26.
74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215(3):403-410.