

Genome Report: chromosome-scale genome assembly of the African spiny mouse (*Acomys cahirinus*)

Elizabeth Dong Nguyen,^{1,2,3,*} Vahid Nikoonejad Fard,⁴ Bernard Y. Kim,⁵ Sarah Collins,² Miranda Galey,⁶ Branden R. Nelson,⁷ Paul Wakenight,⁷ Simone M. Gable,⁴ Aaron McKenna,⁸ Theo K. Bammler,⁹ Jim MacDonald,⁹ Daryl M. Okamura,^{1,2} Jay Shendure,^{3,10,11,12,13} David R. Beier,^{1,2} Jan Marino Ramirez,^{7,14} Mark W. Majesky,^{1,2,13,15} Kathleen J. Millen,^{1,3,7} Marc Tollis,^{4,‡} Danny E. Miller^{3,6,15,*‡}

¹Department of Pediatrics, University of Washington, Seattle, WA 98195, USA

²Center for Developmental Biology & Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA 98101, USA

³Brotman Baty Institute for Precision Medicine, University of Washington, Seattle, WA 98195, USA

⁴School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA

⁵Department of Biology, Stanford University, Stanford, CA 94305, USA

⁶Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA

⁷Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA 98101, USA

⁸Department of Molecular & Systems Biology, Dartmouth Geisel School of Medicine, Lebanon, NH 03755, USA

⁹Department of Environmental & Occupational Health Sciences, University of Washington, Seattle, WA 98195, USA

¹⁰Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

¹¹Allen Discovery Center for Cell Lineage Tracing, Seattle, WA 98195, USA

¹²Howard Hughes Medical Institute, Seattle, WA 98195, USA

¹³Institute of Stem Cell & Regenerative Medicine, University of Washington, Seattle, WA 98195, USA

¹⁴Department of Neurological Surgery, University of Washington, Seattle, WA 98195, USA

¹⁵Department of Laboratory Medicine & Pathology, University of Washington, Seattle, WA 98195, USA

*Corresponding author. Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. Email: elizabeth.nguyen@seattlechildrens.org; *Corresponding author. Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. Email: dm1@uw.edu

‡These authors contributed equally.

There is increasing interest in the African spiny mouse (*Acomys cahirinus*) as a model organism because of its ability for regeneration of tissue after injury in skin, muscle, and internal organs such as the kidneys. A high-quality reference genome is needed to better understand these regenerative properties at the molecular level. Here, we present an improved reference genome for *A. cahirinus* generated from long Nanopore sequencing reads. We confirm the quality of our annotations using RNA sequencing data from 4 different tissues. Our genome is of higher contiguity and quality than previously reported genomes from this species and will facilitate ongoing efforts to better understand the regenerative properties of this organism.

Keywords: genome assembly; *Acomys cahirinus*; spiny mouse; regenerative wound healing; Nanopore sequencing

Introduction

African spiny mice (genus *Acomys*) are a rodent species native to Africa and the Middle East. Their origin dates back to the late Miocene period ~8.7 MYA in the savannas of East Africa (Aghová et al. 2019). Unique adaptations to their environment have made them distinct from other rodents, as they are the first rodent species to exhibit menstruation (Bellofiore et al. 2017, 2021) and have the unique ability to concentrate urine to survive their arid environments (Dickinson et al. 2007). The African spiny mouse inhabits what is known as Evolution Canyon in lower Nahal Oren, Mount Carmel, Israel which consists of 2 distinct microenvironments, the hot and dry African slope and the temperate, humid, and forest European slope (Hadid et al. 2014). The spiny mouse has thus been an evolutionary model of sympatric speciation, with populations of animals demonstrating divergence in karyotype (Volobouev et al. 2007), mitochondrial DNA

(Hadid et al. 2014), and genome methylation patterns (Wang et al. 2022).

More recently, *Acomys cahirinus* (Desmarest, 1819), a member of the African spiny mouse family, has emerged as a model organism for the study of organ regeneration. Members of this family have adapted for survival in unique ways, including the ability for scarless healing of complex tissue after injury as adults. Spiny mice can shed their dorsal skin to escape the grasp of predators and then fully regenerate the lost skin without fibrotic scarring (Seifert et al. 2012). This scarless healing is accompanied by complete regeneration of skin including hair follicles, sebaceous glands, cartilage, adipose tissue, nerves, and blood vessels in the correct architecture for restoration of structure and function of skin tissue (Seifert et al. 2012; Brant et al. 2015; Gawriluk et al. 2016; Matias Santos et al. 2016; Jiang et al. 2019; Maden and Brant 2019; Brewer et al. 2021; Harn et al. 2021). The spiny mouse also demonstrates the ability to restore skeletal muscle after

damage induced by cardiotoxin (Garry et al. 2016; Maden et al. 2018). These healing properties extend to internal organs; kidney damage induced using aggressive models of obstructive and ischemic injury is followed by complete regeneration of functional kidney tissue without scarring (Okamura et al. 2021). The spiny mouse has also been shown to exhibit resistance to myocardial ischemia and minimal scarring, as well as improvement in cardiac function after injury (Koopmans et al. 2021; Peng et al. 2021; Qi et al. 2021). Regeneration to this degree has been demonstrated in other mammalian species (albeit rarely), including humans, particularly in fetal tissues (Colwell et al. 2003; Drenckhahn et al. 2008; Porrello et al. 2011; Pratsinis et al. 2019; Abrams et al. 2021). This suggests that the potential pathways directing regeneration exist in the mammalian genome in a repressed state. A deeper understanding of the spiny mouse genome would help uncover its wound healing properties and possible reversal in nonregenerative mammalian species.

Here, we report a long-read-based chromosome-level assembly for the African spiny mouse *A. cahirinus*, a member of the *Acomys* family that is known to be capable of organ regeneration (Brewer et al. 2021; Okamura et al. 2021). We found that the genome of *A. cahirinus* is 2.3 Gb in length and contains >40% repetitive DNA. While previously published reference genomes for the species (Wang et al. 2022) contained a reported 94% gene completeness and 108-Mb scaffold contiguity, our assembled *A. cahirinus* genome is more contiguous, with a scaffold N50 of 127 Mb, as well as more complete in terms of gene content (98.5%). The *A. cahirinus* genomic resources provided here will contribute to the better understanding of their unique organismal adaptations broadly, while accelerating further discovery of mechanisms underlying their novel adult regenerative capabilities.

Materials and methods

Karyotype and banding

Chromosome analysis was performed on fibroblasts grown from ear tissue, anticoagulated blood, and bone marrow from the femur of a male *A. cahirinus*. Fibroblasts were grown to 70–80% confluency in DMEM/F12 with 10% fetal bovine serum (FBS) and 1% Pen-Strep, with rounded cells indicating active mitosis from passage 1, 2, or 3. Anticoagulated blood was grown in RPMI (Gibco #11875093) supplemented with 10% FBS and 1% Pen-Strep and 200 μ L of PHA (Gibco #10576015) for 3 days. Femurs were cut open and rinsed multiple times with 1–2 mL of RPMI supplemented with 10% FBS and 1% Pen-Strep. Bone marrow cells were then placed into 10-mL cultures for 24 hours.

Samples were placed in 50 μ L of ethidium bromide (1 mg/mL) and 50 μ L of Karyomax Colcemid (10 μ g/mL) (Gibco #15212-012) for 1 hour. Cells were then spun down at 500 g for 10 minutes. Cells were gently resuspended in 0.56% KCl and incubated at room temperature for 20 minutes. Cells spun down again at 500 g for 10 minutes. Cells were gently resuspended in Carnoys Fixative (3:1 methanol:acetic acid) and incubated for 45 minutes. This was repeated twice, with incubation shortened to 10 minutes. Cells were then resuspended in a small volume of fresh Carnoys and dropped onto clean slides. Slides were kept at 37° for a minimum of 24 hours before banding.

For GTG banding, slides were dipped in Trypsin 2.5% (Gibco #15090-046) with NaCl for 15–60 seconds, then rinsed in NaCl with FBS, then NaCl again. Slides were then stained for 10 minutes in Karyomax Giemsa Stain R66 Solution (Gibco #10092-013) with

50 mL of Gurr Buffer Tablets 6.8ph (Gibco #10582-013). After rinse with ddH₂O, slides were dried and imaged.

Nanopore sequencing and preassembly filtering

Genomic DNA was extracted from blood from a single male *A. cahirinus* animal using a Monarch HMW DNA Extraction Kit for Cells & Blood (T3050, New England Biolabs, Ipswich, MA) following the manufacturer's recommended protocol. DNA was quantified prior to library construction using the Qubit DNA HS Assay (ThermoFischer, Waltham, MA), and DNA fragment lengths were assessed using the Agilent Femto Pulse System (Santa Clara, CA). Libraries were prepared for sequencing using the Oxford Nanopore ligation kit (SQK-LSK110) following the manufacturers' instructions, except that DNA repair and A-tailing were performed for 30 minutes and the ligation was allowed to continue for 1 hour. Prepared libraries were quantified using a Qubit fluorometer, and 30 fmol of the library was loaded onto a Nanopore version R.9.4.1 flow cell on the PromethION platform running MinKNOW version 21.05.20. To increase output, the flow cell was washed after approximately 24 hours of sequencing then an additional 12 fmol of library was loaded and run for an additional 48 hours. Basecalling was performed using Guppy 5.0.12 (Oxford Nanopore) using the super accuracy model (dna_r9.4.1_450bps_sup_prom.cfg). Reads of quality 6 or less were discarded, and NanoPlot was used to collect read statistics (Supplementary Table 1, Supplementary Fig. 1).

Assembly and polishing

FASTQ files for assembly were extracted from unaligned bam files using samtools (Li et al. 2009) then Flye version 2.9 for assembly using the --nano-hq flag (Kolmogorov et al. 2019). Haplotigs and overlaps in the assembly were purged using purge_dups (Guan et al. 2020). The assembly was then polished using Medaka version 1.4.2 (<https://github.com/nanoporetech/medaka>) followed by a second polishing step with pilon version 1.24 (Walker et al. 2014). Assembly statistics at each step were generated using Quast (Gurevich et al. 2013) and BUSCO version 5.2.2 using the vertebrata_odb10 database (Simão et al. 2015) (Supplementary Table 2).

Hi-C scaffolding

The primary contigs assembled from the Nanopore data were anchored to pseudo-chromosomes using 505,210,505 read pairs of a Hi-C library isolated from another *A. cahirinus* individual of unknown sex, downloaded from the NCBI Short Read Archive (SRX13258644) (Wang et al. 2022). After aligning the Hi-C reads with the ArimaHi-C Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline), YaHS v1.0 (Zhou et al. 2023) was used with default error correction for scaffolding, and Juicebox v1.11.08 (Dudchenko et al. 2018) was used to generate a Hi-C contact map.

Annotation

Progressive Cactus was used (Armstrong et al. 2020) to perform a whole-genome alignment of the scaffolded *A. cahirinus* draft assembly to the *Mus musculus* GRCm39 reference genome (RefSeq GCF_000001635.27_GRCm39). Comparative annotation of the draft genomes was then performed using the Comparative Annotation Toolkit (CAT) (Fiddes et al. 2018). Briefly, the *M. musculus* RefSeq annotation GFF was parsed and validated with the "parse_ncbi_gff3" and "validate_gff3" programs (respectively) from CAT. The *M. musculus* reference transcript cDNA sequences were downloaded and mapped to the *M. musculus* draft genome with minimap2 (Li 2018) and provided to CAT as long-read RNA-seq

reads in the “[ISO_SEQ_BAM]” field of the configuration file. For *A. cahirinus*, bulk RNA-seq data obtained from multiple pooled organs were downloaded from NCBI SRA BioProject PRJNA342864 (Bellofiore et al. 2017) and mapped to the draft assembly with STAR (Dobin et al. 2013) then provided to CAT in the “[BAMS]” field. CpG islands were identified using the cpg_lh utility from the UCSC suite of tools (Kent et al. 2002).

We modeled repeats de novo for the *A. cahirinus* scaffolds with RepeatModeler v2.0 (Flynn et al. 2020), and used RepeatMasker v4.1.3 (Smit et al. 1996) to (1) classify the de novo repeat family consensus sequences and (2) annotate all classified repeats in the genome assembly based on the “rodentia” repeat library from RepBase v4.0.7 (Bao et al. 2015).

RNA isolation and mapping

Tissues (blood, heart, liver, and testis) were collected from an adult male *A. cahirinus* and homogenized, and RNA isolation, library generation, and sequencing were performed as previously described (Brewer et al. 2021; Okamura et al. 2021). Briefly, total RNA was extracted in Trizol solution (Ambion), DNase treated, and purified (PureLink RNA Mini Kit, Thermo Fisher Scientific). RNA was processed with KAPA’s Stranded mRNA-Seq kit (Illumina) following the manufacturer’s protocol in duplicate for each sample. The resulting libraries were assessed for library quality using fragment length and number of cycles in real-time PCR. Passing samples were sequenced on a NextSeq 500 using a 300-cycle mid-output kit, with paired 150-bp reads.

RNA was mapped to the final assembly using bwa (version 0.7.17-r1188) (Li and Durbin 2009). Reads mapping to genomic features defined in the GTF file were counted using featureCounts using the simplified file format (Liao et al. 2014). For each gene, the transcripts per million (TPM) value was calculated using only mapped reads (Supplementary File 2). Venn diagrams were created in R and show overlap in genes from each tissue with TPM values greater than 2 (Fig. 2b).

Comparative genomics

Synteny analysis

To understand evolutionary change between the *A. cahirinus* and *M. musculus* genomes, we used SynMap2 (Haug-Baltzell et al. 2017) on the CoGe platform (Lyons and Freeling 2008) to visualize whole-genome synteny between *A. cahirinus* scaffolds and the *M. musculus* reference genome (mm39). We used lastz (Harris 2007) to map *A. cahirinus* coding sequences to both genomes, DAGChainer (Haas et al. 2004) to compute chains of syntenic genes, and CodeML (Yang and Nielsen 2002) to calculate the rate of nonsynonymous (Kn) and synonymous (Ks) substitutions, as well as their ratios (Kn/Ks), with default parameters.

Repeat analysis

To estimate the amount of evolutionary divergence within repeat families, we generated repeat family-specific alignments using the -a flag in RepeatMasker and calculated the average Kimura 2-parameter (K2P) sequence divergence between each annotated repeat insertion and its family consensus sequence. To correct for higher mutation rates at CpG sites, we weighted 2 transition mutations as 1% of a single transition. These steps were undertaken using the calcDivergenceFromAlign.pl tool in RepeatMasker. We compared the resulting repeat landscape obtained for *A. cahirinus* to a parallel analysis we conducted for *M. musculus* (mm10).

Orthologous gene analysis

To further examine genomic differences between *A. cahirinus* and *M. musculus*, we generated pairwise genome alignments. We first aligned *A. cahirinus* scaffolds as queries to the mouse reference genome (mm39) with lastz (Harris 2007) using parameters $K = 2400$, $L = 3000$, $Y = 9400$, $H = 2000$, which are sensitive enough to detect orthologous exons in placental mammals (Sharma and Hiller 2017), and a default scoring matrix, followed by chaining and netting (Kent et al. 2003). To analyze protein-coding genes, we downloaded mm39 RefSeq gene annotations for each mouse chromosome in whole gene BED format from the UCSC Genome Browser (Kent et al. 2002) and used the “stitch gene blocks” tool available on Galaxy (usegalaxy.org, last accessed February 2023) to reconstruct sequence alignments for each mouse protein-coding gene ID containing the prefix “NM_” (Blankenberg et al. 2011). We then removed gaps in the reference alignments, removed codons with missing nucleotides which produce unknown amino acids, removed premature stop codons, and converted all filtered FASTA alignments into axt format with AlignmentProcessor.py (<https://tinyurl.com/23y38664>, last accessed February 2023).

Finally, we used OrthoFinder v2.5.5 (Emms and Kelly 2019) to detect orthologs and identify gene duplication events in the evolutionary history of 12 therian mammalian genomes. In addition to *A. cahirinus*, we included the proteomes from the NCBI-annotated genome assemblies for opossum (*Monodelphis domestica*, GCA_027917375.1); African savannah elephant (*Loxodonta africana*, GCA_030077915.1); blue whale (*Balaenoptera musculus*, GCA_008658375.2); cow (*Bos taurus*, GCA_905123515.1); dog (*Canis lupus familiaris*, GCA_000002285.4); rhesus macaque (*Macaca mulatta*, GCA_003339765.3); human (*Homo sapiens*, GCA_000001405.29); guinea pig (*Cavia porcellus*, GCA_000151735.1); black rat (*Rattus rattus*, GCA_011064425.1); house mouse (mm10, above); and golden spiny mouse (*Acomys russatus*, GCA_903995435.1). All alignments and OrthoFinder output are included in Supplementary Data and are publicly available.

Functional analysis

To examine protein-coding differences that may point to selection pressures acting on genes since the divergence of *A. cahirinus* and *M. musculus*, we estimated the pairwise synonymous K_a and non-synonymous K_s substitution rate, as well as the rate ratio K_a/K_s for all filtered axt gene alignments with KaKs_calculator2.0 (Wang et al. 2010), accounting for variable mutation rates across sites with a maximum likelihood method MS (Supplementary File 1). We concatenated the results of the K_a/K_s test for each gene ID and applied the false discovery rate (FDR = 0.05) to reduce false positives (Supplementary Data). We functionally annotated all unique gene IDs with $K_a/K_s > 1.0$ and an adjusted P -value < 0.05 using DAVID (Sherman et al. 2022) and Gene Ontology enrichment (Gene Ontology Consortium 2015), applying Benjamini-Hochberg and FDR corrections to adjust for multiple testing.

Results and discussion

A. cahirinus from our colony have a chromosomal count of 38 (19 pairs). Most autosomes are metacentric or submetacentric with a large acrocentric X, small acrocentric Y, and 2 pairs of small acrocentric autosomes (Fig. 1a). The *A. cahirinus* karyotype, while similar in its combination of metacentric and acrocentric chromosomes to other rodent species, is karyotypically divergent from the completely acrocentric pairs of 20 chromosomes in *M. musculus*. The high karyotypic diversity in rodents enables chromosomal

numbers and morphology to be a useful tool in species identification. Although the geographic origin of our animals is unknown, we find that our results match the *A. cahirinus* karyotype from Moreshet, Israel, which is distinct from the *A. cahirinus* karyotype generated from animals in Sinai, Egypt, which have 36 chromosomes (Volobouev et al. 2007).

Using a single male individual from our colony, we generated 87.5 Gb of Nanopore data for primary assembly with a read length N50 of 63 kb and a mean read quality of 13. The initial primary assembly after purging duplicates and polishing contained 181 contigs with a contig N50 of 58.8 Mb, a longest contig of 126.8 Mb, and a total length of 2.3 Gb (Table 1). The contigs were anchored to 19

pseudo-chromosomes based on the Hi-C scaffolding, matching expectations from the karyotype (Fig. 1b). Hi-C scaffolding reduced the number of assembled sequences to 129, with a scaffold N50 of 127 Mb and a total length of 2,289,268,912 bp and 79 gaps. All nanopore contigs were scaffolded. Fifty percent of the scaffolded assembly resides on 8 scaffolds (L50). According to the BUSCO analysis of the scaffolded assembly, 98.5% of complete and partial single-copy mammalian orthologs are present, indicating a higher level of completeness than previously published reference genomes for *A. cahirinus* (Wang et al. 2022)

We estimate that approximately 37% of the *A. cahirinus* genome consists of repetitive sequences, an identical proportion to what

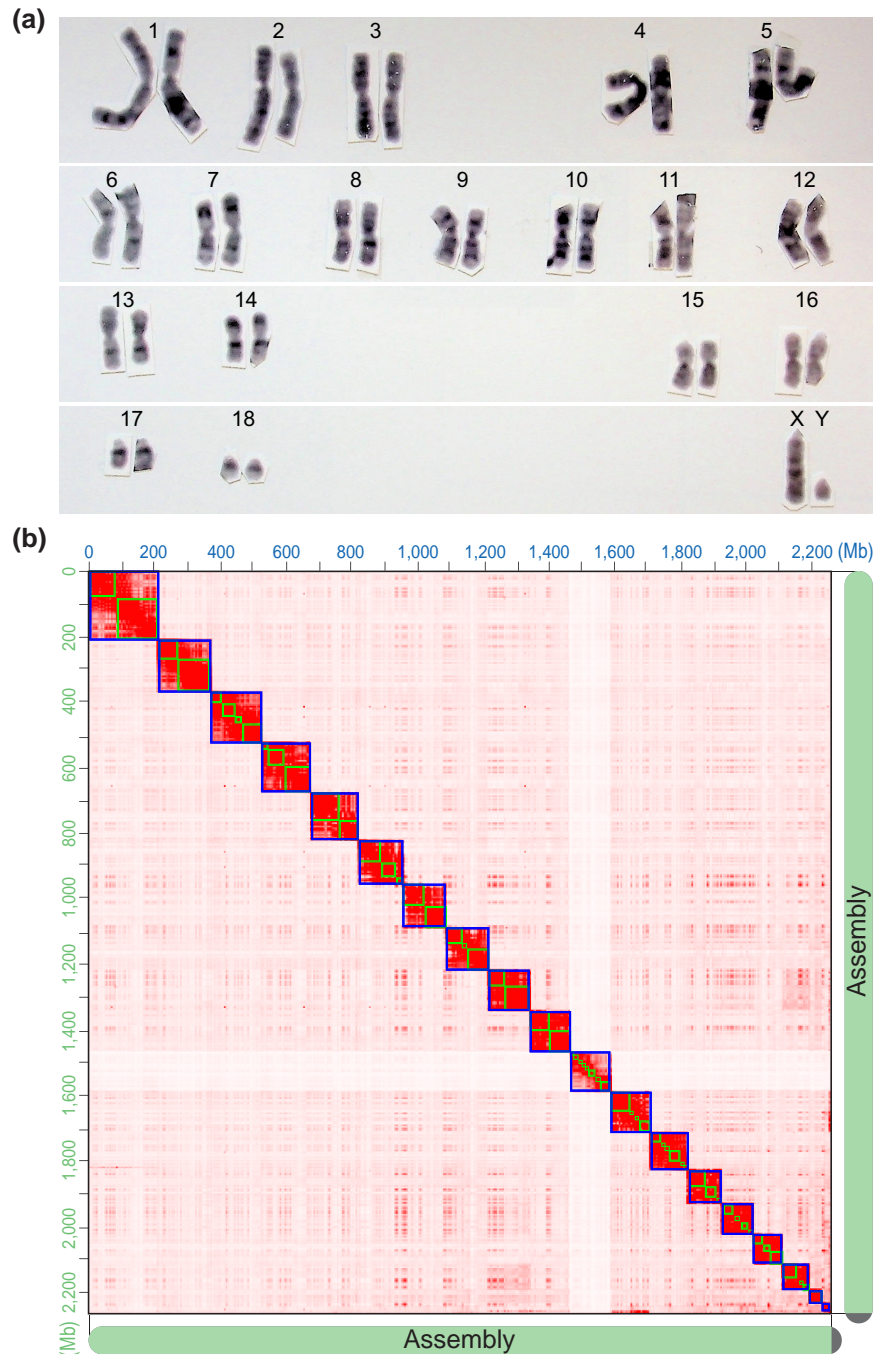


Fig. 1. Karyotype of *Acomys cahirinus* and results of Hi-C scaffolding of the assembly. a) Representative karyotype of a male *A. cahirinus* from the same colony that DNA and RNA were obtained from for sequencing and assembly. Autosomes are arranged by size and numbered accordingly. b) Hi-C interaction contact heatmap of 19 *A. cahirinus* pseudo-chromosomes matches the karyotype (bin size is 1 Mb).

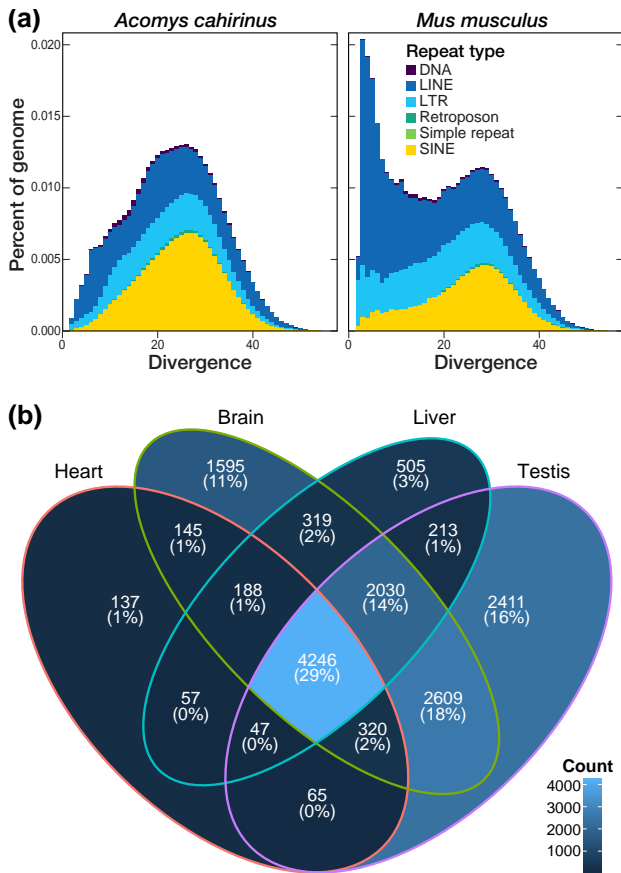


Fig. 2. a) Repeat landscapes for *Acomys cahirinus* and *Mus musculus*, visualizing the percent of each genome comprised of different types of repeats according to their Kimura 2-parameter divergence from family consensus. DNA, DNA transposon; LINE, long interspersed nuclear element; LTR, long terminal repeat retrotransposon; SINE, short interspersed nuclear element. b) Venn diagram showing the number of genes per tissue with a TPM value greater than 2. The majority of genes (4,246; 29%) are expressed at this level in all 4 tissues, while testis had the highest number of genes (2,411; 16%) with a TPM greater than 2 not seen in other tissues.

we found in *M. musculus* (Table 2). Thirty four percent of the *A. cahirinus* genome consisted of interspersed repeats such as transposable elements, with most of these belonging to retrotransposons, which accounted for 30% of the genome alone. Compared to *M. musculus*, *A. cahirinus* contains more SINE retrotransposons (8.3% of the genome vs 11.4%, respectively), while *M. musculus* contains more long interspersed nuclear elements (LINEs) (19% of the genome vs 11%, respectively). These differences can be attributed to a recent burst of LINE-1 retrotransposon activity in *M. musculus* (Sookdeo et al. 2013), as demonstrated by a relatively taller peak of LINE elements in the *M. musculus* genome at $\leq 10\%$ K2P divergence compared to *A. cahirinus* (Fig. 2a). The pseudochromosome 11 (chr11), which is 4,798,714 bp in length and made of 11 scaffolds and 12 contigs, contained relatively fewer repeats (2,441) compared to other scaffolds.

The average amino acid similarity across gene blocks between *M. musculus* and *A. cahirinus* was 80%. Almost all forms of structural variations such as inversions, duplications, and insertions/deletions were detected in the synteny analysis between *A. cahirinus* and *M. musculus* (Supplementary Fig. 2). While this suggests dynamic structural changes to the rodent genome, the mutation

rate analysis indicated a major peak at $K_s \ll 0.0$, suggesting the majority of found gene pairs predate the *M. musculus*-*A. cahirinus* divergence (Supplementary Fig. 3).

The average nucleotide divergence between *A. cahirinus* and *M. musculus* transcripts was 12%. We estimated K_a/K_s for 33,197 mouse gene IDs that passed our alignment filtering. The vast majority of the genes had K_a/K_s values between 0 and 1 (mean 0.16, Supplementary Fig. 4), indicating purifying selection acting on protein-coding genes across rodents of the Muridae family. Of the rest, 38 significant gene IDs had both $K_a/K_s > 1.0$ and an adjusted P -value < 0.05 ; 34 of these contained DAVID IDs, many of which are predicted or known DNA chromatin/transcription (GO:0140110) and signal transducer regulation factors (GO:0060089, Panther GO-Slim molecular function). For example, 3 genes (*Dmrtc1b*, *Dmrtc1c1*, and *Dmrtc1c2*) were annotated by Uni-Prot and InterPro as being involved in the doublesex and mab-3 related transcription factor-like families, and 4 were enriched with the GO term meiotic cell cycle (GO:0051321), along with *Obox5* (oocyte specific homeobox 5) and *Rhox4* (reproductive homeobox 4C) transcription factors. Thirty-one of the significant genes with $K_a/K_s > 1.0$ were enriched with PANTHER GO-Slim terms for biological processes using the 33,197 aligned *M. musculus*-*A. cahirinus* genes as background. Enriched gene ontology terms included spermatid development (GO:0007286, 34.2-fold enrichment, $FDR = 0.0184$) and germ cell development (GO:0007281, 46.9-fold enrichment, $FDR = 8.59E^{-05}$). These results suggest that important differences at the amino acid level between *M. musculus* and *A. cahirinus* contribute to post-speciation differences in reproductive development. This is a common result in comparative genomics analyses across mammalian species (Chai et al. 2021), and yet it may be indicative of *A. cahirinus*' adaptations underlying their novel menstrual cycle, longer gestational times, and precocial births vs *M. musculus*' more rapid estrous cycles and altricial birthing strategies (Bellofiore et al. 2017). Other *A. cahirinus* $K_a/K_s > 1.0$ genes with molecular function and biological process enriched terms of interest include catalytic activity (GO:0003824), metabolic processes (GO:0008152), and immune system processes (GO:0002376). For example, *Tcl1b3/4* (T cell leukemia/lymphoma 1B 3 and 4) is a protein serine/threonine kinase activator, and *Wfdc10* (WAP four-disulfide core domain 10) is an endopeptidase inhibitor with predicted roles in local immune responses in reproductive tissues based on mouse ENCODE RNA expression (Yue et al. 2014). *Gimap4* (GTPase of the immune associated nucleotide binding protein 5) regulates T lymphocyte activation and long-term survival (Limoges et al. 2021), *Ccnb3* (cyclin B3) is a known cell cycle and proliferation regulator, and *Slamf7* (signaling lymphocytic activation molecule family 7) is a signaling receptor that regulates innate and adaptive immune cell activation more broadly, which also underlies certain cancer progression in humans (Farhangnia et al. 2023). Thus, even our initial comparison between *A. cahirinus* and *M. musculus* protein coding level changes successfully revealed novel variants in key genes that regulate cellular processes. Understanding how these naturally selected protein coding changes differentially affect signal control and chromatin regulation mechanisms in *A. cahirinus* compared to other mammals will be of interest as this genome is investigated further to decode this animal's unique biology.

Out of 254,113 total genes across the 12 species, we assigned 247,445 (97.4%) genes to orthogroups, indicating that with our species sampling we were able to capture a high degree of orthologous gene relationships. In particular, we found that 97.1% of *A. cahirinus* genes were successfully assigned to orthogroups, indicative of the quality of our annotation. The highest percentage of genes assigned

Table 1. Comparison of metrics from available genome assemblies for *Acomys cahirinus*.

	Broad Institute (GCA_004027535.1)		Wang et al. (2022)		Current study	
	Contigs	Scaffolds	Contigs	Scaffolds	Contigs	Scaffolds
Total length	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb
Number of sequences	391,811	371,342	120	108	181	129
N50	42.5 kb	65.4 kb	55.0 Mb	nr	58.8 Mb	127.8 Mb
L50	15,859	10,134	na	nr	16	8
Number of gaps (≥ 5 bp)	20,469			nr		79
Complete + partial BUSCOs (Mammalia orthoDBv10)	83.1%			94% ^a		98.5%

bp, base pairs; kb, kilobase pairs; Mb, megabase pairs; Gb, gigabase pairs; nr, not reported.
^a Unknown BUSCO database.

Table 2. Genome annotation information for *Acomys cahirinus* (current study) and *Mus musculus* (mm10).

Parameters	Current study (<i>Acomys cahirinus</i>)	<i>Mus musculus</i> (mm10)
Total length	2.3 Gb	2.7 Gb
GC content	42.8%	41.7%
Annotated protein-coding genes	19,818	22,192
Average number of exons per gene	14.75	5.91
Average gene length	50,140 bp	28,506 bp
Average number of isoforms per gene	4.34	3.59
Bases masked	36.7%	36.7%
Total interspersed repeats	33.9%	39%
Retroelements	30.2%	37.1%
SINEs	11.7%	8.3%
LINEs	11.4%	18.9%
LTR elements	7.2%	9.9%
DNA transposons	0.63%	0.45%
Unclassified	3.0%	1.5%

Gb, gigabase pairs; bp, base pairs; SINEs, short interspersed nuclear elements; LINEs, long interspersed nuclear elements; LTR, long terminal repeat.

to an orthogroup was for guinea pig (98.9%), and the lowest percentage of genes assigned to an orthogroup was for opossum (94.5%), likely due to its being the only marsupial in our species sampling. We inferred the number of unique gene duplication events in the evolutionary history of *Acomys*, including 537 at the origin of therian mammals, 333 at the origin of eutherians, 4,061 for human, 828 for murine rodents, 2,493 for *M. musculus*, 2,246 for *A. russatus*, and 6,679 for *A. cahirinus*. We found 126 unique genes in 51 families unique to *A. cahirinus*. These results demonstrate that our *A. cahirinus* genome will be a useful tool in comparative genomics studies outside the context of pairwise comparisons to mouse.

To confirm the quality of our assembly and annotation as well as to identify a broad range of expressed transcripts, we performed RNA sequencing of several tissues. We aligned short-read RNA isolated from heart, liver, brain, and testis to the assembled genome. TPM for each annotated gene was calculated and used to determine expression levels among the 4 tissues for the 19,818 annotated genes. More genes with a TPM level > 2 were observed in the brain (11,450) and testis (11,938) compared to liver (7,593) and heart (5,194), and testis had the largest number of uniquely expressed transcripts (2,411) (Fig. 2b). This result is consistent with other studies that have demonstrated a high number of unique transcripts in brain in both *M. musculus* and *Rattus norvegicus* (Söllner et al. 2017) as well as a unique number of expressed transcripts in testis (Djureinovic et al. 2014; Uhlén et al. 2015). These results provide

support for the high-quality of our genome assembly and demonstrate that tissue-specific expression analysis is feasible in order to better understand the regenerative capabilities of this species.

Diverse scientific disciplines have long studied *A. cahirinus* for their unique organismal and behavioral adaptations. Most recently, *A. cahirinus* have emerged as an exciting and experimentally tractable adult regenerative mammalian model, as their naturally selected capacity for antifibrotic scarless epidermal wound healing extends across multiple internal systems and different injury contexts. Hence, our highly contiguous, high-quality genome presented here will broadly benefit the growing *A. cahirinus* community and will accelerate more detailed investigations into the genetic and epigenetic mechanisms underlying *A. cahirinus*' novel capacity to maintain organ regeneration as adult mammals.

Data availability

The scaffolded genome assembly, RNA sequencing data, and original Nanopore data are available at NCBI under bioproject PRJNA935753. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAULSH000000000. The version described in this paper is version JAULSH010000000. The scaffolded genome assembly and gff3 files are also available at <https://doi.org/10.5281/zenodo.7761277>. Additional annotation, alignment, and results from Ka/Ks analysis are available at <https://doi.org/10.5281/zenodo.7734822>. Orthofinder results are available at <https://doi.org/10.6084/m9.figshare.23528349>.

Supplemental material available at G3 online.

Acknowledgments

We thank Angela Miller for the help with editing and figure preparation.

Funding

W. M. Keck Foundation (MWM and KJM). National Institutes of Health R01DK114149 (MWM). NIH R21OD023838 (KJM and BRN). NIH R21OD030107 (KJM). Impetus Longevity Award (BRN). NIH DP5OD033357 (DEM). Brotman Baty Institute (EDN and DEM). National Institutes of Health U54CA217376 (MT). NIH 5P50HD103524-03 (JM and TKB). Howard Hughes Medical Institute (JS). Seattle Children's Research Institute Center for Developmental Biology and Regenerative Medicine (DRB). Seattle Children's Research Institute Center for Integrative Brain Research (JMR).

Conflicts of interest

DEM is engaged in a research agreement with Oxford Nanopore, and they have paid for him to travel to speak on their behalf. DEM is on a scientific advisory board at Oxford Nanopore. The other authors have no conflicts to declare.

Author contributions

Conception: EDN, BRN, MWM, KJM, MT, DEM. Analysis: EDN, VN, SMG, BYK, AM, TKB, JM, MT, DEM. Experiments: PW, MG, AM, SC, DMO, DEM. Funding: BRN, KJM, JS, MWM, MT, DEM, DRB, JMR. Writing: EDN, BRN, KJM, MT, DEM.

Literature cited

- Abrams MJ, Tan FH, Li Y, Basinger T, Heithe ML, Sarma Anish, Lee IT, Condiotte ZJ, Raffiee M, Dabiri JO, et al. A conserved strategy for inducing appendage regeneration in moon jellyfish, *Drosophila*, and mice. *Elife*. 2021;10:e65092. doi:10.7554/eLife.65092.
- Aghová T, Palupčiková K, Šumbera R, Frynta D, Lavrenchenko LA, Meheretu Y, Sádlová J, Votýpka J, Mbau JS, Modrý D, et al. Multiple radiations of spiny mice (Rodentia: Acomys) in dry open habitats of Afro-Arabia: evidence from a multi-locus phylogeny. *BMC Evol. Biol.* 2019;19(1):69. doi:10.1186/s12862-019-1380-9.
- Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature*. 2020;587(7833):246–251. doi:10.1038/s41586-020-2871-y.
- Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6(1):11. doi:10.1186/s13100-015-0041-9.
- Bellofiore N, Ellery SJ, Mamrot J, Walker DW, Temple-Smith P, Dickinson H. First evidence of a menstruating rodent: the spiny mouse (*Acomys cahirinus*). *Am J Obstet Gynecol.* 2017;216(1):40.e1–40.e11. doi:10.1016/j.ajog.2016.07.041.
- Bellofiore N, McKenna J, Ellery S, Temple-Smith P. The spiny mouse—a menstruating rodent to build a bridge from bench to bedside. *Front Reprod Health.* 2021;3:784578. doi:10.3389/frph.2021.784578.
- Blankenberg D, Taylor J, Nekrutenko A; Galaxy Team. Making whole genome multiple alignments usable for biologists. *Bioinformatics.* 2011;27(17):2426–2428. doi:10.1093/bioinformatics/btr398.
- Brant JO, Lopez M-C, Baker HV, Barbazuk WB, Maden M. A comparative analysis of gene expression profiles during skin regeneration in *Mus* and *Acomys*. *PLoS One.* 2015;10(11):e0142931. doi:10.1371/journal.pone.0142931.
- Brewer CM, Nelson BR, Wakenight P, Collins SJ, Okamura DM, Dong XR, Mahoney WM Jr, McKenna A, Shendure J, Timms A, et al. Adaptations in Hippo-Yap signaling and myofibroblast fate underlie scar-free ear appendage wound healing in spiny mice. *Dev Cell.* 2021;56(19):2722–2740.e6. doi:10.1016/j.devcel.2021.09.008.
- Chai S, Huang X, Wu T, Xu S, Ren W, Yang G. Comparative genomics reveals molecular mechanisms underlying health and reproduction in cryptorchid mammals. *BMC Genomics.* 2021;22(1):763. doi:10.1186/s12864-021-08084-1.
- Colwell AS, Longaker MT, Lorenz HP. Fetal wound healing. *Front Biosci.* 2003;8(6):s1240–s1248. doi:10.2741/1183.
- De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C. Nanopack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34(15):2666–2669. doi:10.1093/bioinformatics/bty149.
- Dickinson H, Moritz K, Wintour EM, Walker DW, Kett MM. A comparative study of renal function in the desert-adapted spiny mouse and the laboratory-adapted C57BL/6 mouse: response to dietary salt load. *Am J Physiol Renal Physiol.* 2007;293(4):F1093–F1098. doi:10.1152/ajprenal.00202.2007.
- Djureinovic D, Fagerberg L, Hallström B, Danielsson A, Lindskog C, Uhlén M, Pontén F. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol Hum Reprod.* 2014;20(6):476–488. doi:10.1093/molehr/gau018.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21. doi:10.1093/bioinformatics/bts635.
- Drenckhahn J-D, Schwarz QP, Gray S, Laskowski A, Kiriazis H, Ming Z, Harvey RP, Du X-J, Thorburn DR, Cox TC. Compensatory growth of healthy cardiac cells in the presence of diseased cells restores tissue homeostasis during heart development. *Dev Cell.* 2008;15(4):521–533. doi:10.1016/j.devcel.2008.09.005.
- Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv* 254797. <https://doi.org/10.1101/254797>, 2018, preprint: not peer reviewed.
- Emms DM, Kelly S. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. doi:10.1186/s13059-019-1832-y.
- Farhangnia P, Ghomi SM, Mollazadehghomi S, Nickho H, Akbarpour M, Delbandi AA. SLAM-family receptors come of age as a potential molecular target in cancer immunotherapy. *Front Immunol.* 2023;14:1174138. doi:10.3389/fimmu.2023.1174138.
- Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, et al. Comparative Annotation Toolkit (CAT)-simultaneous clade and personal genome annotation. *Genome Res.* 2018;28(7):1029–1038. doi:10.1101/gr.233460.117.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 2020;117(17):9451–9457. doi:10.1073/pnas.1921046117.
- Garry GA, Antony ML, Garry DJ. Cardiotoxin induced injury and skeletal muscle regeneration. *Methods Mol Biol.* 2016;1460:61–71. doi:10.1007/978-1-4939-3810-0_6.
- Gawriluk TR, Simkin J, Thompson KL, Biswas SK, Clare-Salzler Z, Kimani JM, Kiama SG, Smith JJ, Ezenwa VO, Seifert AW. Comparative analysis of ear-hole closure identifies epimorphic regeneration as a discrete trait in mammals. *Nat Commun.* 2016;7(1):11164. doi:10.1038/ncomms11164.
- Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Res.* 2015;43(D1):D1049–D1056. doi:10.1093/nar/gku1179.
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 2020;36(9):2896–2898. doi:10.1093/bioinformatics/btaa025.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072–1075. doi:10.1093/bioinformatics/btt086.
- Haas BJ, Delcher AL, Wortman JR, Salzberg SL. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics.* 2004;20(18):3643–3646. doi:10.1093/bioinformatics/bth397.
- Hadid Y, Pavlíček T, Beiles A, Ianovici R, Raz S, Nevo E. Sympatric incipient speciation of spiny mice *Acomys* at “Evolution Canyon,”

- Israel. Proc Natl Acad Sci U S A. 2014;111(3):1043–1048. doi:10.1073/pnas.1322301111.
- Harn HI-C, Wang S-P, Lai Y-C, Van Handel B, Liang Y-C, Tsai S, Schiessl IM, Sarkar A, Xi H, Hughes M, et al. Symmetry breaking of tissue mechanics in wound induced hair follicle regeneration of laboratory and spiny mice. Nat Commun. 2021;12(1):2595. doi:10.1038/s41467-021-22822-9.
- Harris RS. Improved pairwise alignment of genomic DNA. 2007.
- Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E. Synmap2 and SynMap3D: web-based whole-genome synteny browsers. Bioinformatics. 2017;33(14):2197–2198. doi:10.1093/bioinformatics/btx144.
- Jiang T-X, Harn HI-C, Ou K-L, Lei M, Chuong C-M. Comparative regenerative biology of spiny (*Acomys cahirinus*) and laboratory (*Mus musculus*) mouse skin. Exp Dermatol. 2019;28(4):442–449. doi:10.1111/exd.13899.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci U S A. 2003;100(20):11484–11489. doi:10.1073/pnas.1932072100.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12(6):996–1006. doi:10.1101/gr.229102.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–546. doi:10.1038/s41587-019-0072-8.
- Koopmans T, van Beijnum H, Roovers EF, Tomasso A, Malhotra D, Boeter J, Psathaki OE, Versteeg D, van Rooij E, Bartscherer K. Ischemic tolerance and cardiac repair in the spiny mouse (*Acomys*). NPJ Regen Med. 2021;6(1):78. doi:10.1038/s41536-021-00188-2.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–3100. doi:10.1093/bioinformatics/bty191.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.
- Liao Y, Smyth GK, Shi W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–930. doi:10.1093/bioinformatics/btt656.
- Limoges MA, Cloutier M, Nandi M, Ilangumaran S, Ramanathan S. The GIMAP family proteins: an incomplete puzzle. Front Immunol. 2021;12:679739. doi:10.3389/fimmu.2021.679739.
- Lyons E, Freeling M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 2008;53(4):661–673. doi:10.1111/j.1365-3113.2007.03326.x.
- Maden M, Brant JO. Insights into the regeneration of skin from *Acomys*, the spiny mouse. Exp Dermatol. 2019;28(4):436–441. doi:10.1111/exd.13847.
- Maden M, Brant JO, Rubiano A, Sandoval AGW, Simmons C, Mitchell R, Collin-Hooper H, Jacobson J, Omairi S, Patel K. Perfect chronic skeletal muscle regeneration in adult spiny mice, *Acomys cahirinus*. Sci Rep. 2018;8(1):8920. doi:10.1038/s41598-018-27178-7.
- Matias Santos D, Rita AM, Casanellas I, Brito Ova A, Araújo IM, Power D, Tiscornia G. Ear wound regeneration in the African spiny mouse *Acomys cahirinus*. Regeneration (Oxf). 2016;3(1):52–61. doi:10.1002/reg2.50.
- Okamura DM, Brewer CM, Wakenight P, Bahrami N, Bernardi K, Tran A, Olson J, Shi X, Yeh S-Y, Piliponsky, et al. Spiny mice activate unique transcriptional programs after severe kidney injury regenerating organ function without fibrosis. iScience. 2021;24(11):103269. doi:10.1016/j.isci.2021.103269.
- Peng H, Shindo K, Donahue RR, Gao E, Ahern BM, Levitan BM, Tripathi H, Powell D, Noor A, Elmore GA, et al. Adult spiny mice (*Acomys*) exhibit endogenous cardiac recovery in response to myocardial infarction. NPJ Regen Med. 2021;6(1):74. doi:10.1038/s41536-021-00186-4.
- Porrello ER, Mahmoud AI, Simpson E, Hill JA, Richardson JA, Olson EN, Sadek HA. Transient regenerative potential of the neonatal mouse heart. Science. 2011;331(6020):1078–1080. doi:10.1126/science.1200708.
- Pratsinis H, Mavrogonatou E, Kletsas D. Scarless wound healing: from development to senescence. Adv Drug Deliv Rev. 2019;146:325–343. doi:10.1016/j.addr.2018.04.011.
- Qi Y, Dasa O, Maden M, Vohra R, Batra A, Walter G, Yarrow JF, Aranda JM Jr, Raizada MK, Pepine CJ. Functional heart recovery in an adult mammal, the spiny mouse. Int J Cardiol. 2021;338:196–203. doi:10.1016/j.ijcard.2021.06.015.
- Seifert AW, Kiama SG, Seifert MG, Goheen JR, Palmer TM, Maden M. Skin shedding and tissue regeneration in African spiny mice (*Acomys*). Nature. 2012;489(7417):561–565. doi:10.1038/nature11499.
- Sharma V, Hiller M. Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. Nucleic Acids Res. 2017;45(14):8369–8377. doi:10.1093/nar/gkx554.
- Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216–W221. doi:10.1093/nar/gkac194.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–3212. doi:10.1093/bioinformatics/btv351.
- Smit AFA, Hubley R, Green P. Repeat-Masker Open-3.0. 1996–2010. 1996. <http://www.repeatmasker.org>.
- Söllner JF, Leparç G, Hildebrandt T, Klein H, Thomas L, Stupka E, Simon E. An RNA-Seq atlas of gene expression in mouse and rat normal tissues. Sci Data. 2017;4(1):170185. doi:10.1038/sdata.2017.185.
- Sookdeo A, Hepp CM, McClure MA, Boissinot S. Revisiting the evolution of mouse LINE-1 in the genomic era. Mob DNA. 2013;4(1):3. doi:10.1186/1759-8753-4-3.
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419. doi:10.1126/science.1260419.
- Volobouev V, Auffray JC, Debat V, Denys C, Gautun JC, Trasnier M. Species delimitation in the *Acomys cahirinus*-*dimidiatus* complex (Rodentia, Muridae) inferred from chromosomal and morphological analyses. Biol J Linn Soc Lond. 2007;91(2):203–214. doi:10.1111/j.1095-8312.2007.00773.x.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963.
- Wang Y, Qiao Z, Mao L, Li F, Liang X, An X, Zhang S, Liu X, Kuang Z, Wan N, et al. Sympatric speciation of the spiny mouse from Evolution Canyon in Israel substantiated genomically and methylomically. Proc Natl Acad Sci U S A. 2022;119(13):e2121822119. doi:10.1073/pnas.2121822119.

- Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. Kaks_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. 2010;8(1):77–80. doi:[10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3).
- Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 2002;19(6):908–917. doi:[10.1093/oxfordjournals.molbev.a004148](https://doi.org/10.1093/oxfordjournals.molbev.a004148).
- Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, *et al*. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*. 2014 Nov 20;515(7527):355–364. doi:[10.1038/nature13992](https://doi.org/10.1038/nature13992).
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*. 2023;39(1):btac808. doi:[10.1093/bioinformatics/btac808](https://doi.org/10.1093/bioinformatics/btac808).

Editor: R. Mallarino