BMC Genomics

RESEARCH

Open Access

Genome assembly and annotation of the olive grass mouse *Abrothrix olivacea* reveal transcriptomic and cellular adaptations across contrasting biomes

Melisa E. Magallanes Alba[1*], Agustin Baricalla[2], Matias Feijoo[3], Guillermo D'Elia[4], Daniel E. Naya[1] and Enrique P. Lessa[1]

Abstract

Abrothrix olivacea (Waterhouse, 1837), the olive grass mouse, is a widely distributed sigmodontine rodent that inhabits a broad range of environments, from the hyper arid deserts of southernmost Perú and northern Chile to the Patagonian steppe to the humid temperate rainforests of southern South America. Its extensive ecological breadth, coupled with physiological adaptations to water scarcity, makes it an ideal model for studying environmental responses and phenotypic plasticity. Here, we present the first de novo scaffold-level genome assembly of *A. olivacea*, generated from short-read DNA sequencing. The 2.25 Gb assembly achieved a scaffold N50 of 123 Mb and a BUSCO completeness score of 98.61%, indicating high sequence completeness. Genome annotation identified 21,476 protein-coding genes, providing a valuable resource for evolutionary, ecological, and functional genomics. As a case study, we used this reference genome to explore gene expression and genetic divergence in kidney tissue from individuals inhabiting contrasting environments: the southern Andean rainforest and the Patagonian steppe. By integrating single-cell transcriptomic data from *Mus musculus*, we performed cell type deconvolution, revealing environment-specific expression patterns linked to renal function. This new genomic resource opens avenues for investigating local adaptation, population structure, and conservation genetics in one of South America's most ecologically versatile and widely distributed rodents.

Keywords Genome assembly, Kidney transcriptomics, Adaptation, Comparative genomics

Background

The olive grass mouse, *Abrothrix olivacea* [1], a sigmodontine of the tribe Abrotrichini, exhibits remarkable ecological adaptability, thriving in diverse habitats in southern South America, across various regions in southernmost Perú, Chile and Argentina, from hyper arid deserts in southernmost Peru and northern Chile to the Valdivian and Magellanic temperate rainforests in the south, as well as the Chilean Mediterranean region, the Patagonian steppes and Fuegian grasslands [2]. This widespread distribution across heterogeneous landscapes offers valuable opportunities to explore how

*Correspondence:
Melisa E. Magallanes Alba
melisa.eliana.magallanes@gmail.com
[1]Departamento de Ecología y Evolución, Facultad de Ciencias, Universidad de La República, Montevideo, Uruguay
[2]Instituto Multidisciplinario de Investigaciones Biológicas de San Luis (IMIBIO-SL), Universidad Nacional de San Luis - CONICET, San Luis, Argentina
[3]Departamento de Sistemas Agrarios y Paisajes Culturales, Centro Universitario Regional del Este, Universidad de La República, Rocha, Uruguay
[4]Facultad de Ciencias, Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia, Chile

phenotypic variability is structured and whether it correlates with environmental features [3]. *A. olivacea* exhibits exceptional urine concentration capability and a relative medullary thickness comparable to desert-adapted species [4, 5]. Higher tolerance to water shortage in populations from xeric habitats has already been demonstrated [6]. Based on variation in morphology, coloration patterns, and DNA sequence data [7–10], distinct subspecies of *A. olivacea* have been described, and six divergent mitochondrial phylogeogroups have been found along its distribution [11, 12]. More recently, studies of both mitochondrial DNA and nuclear SNPs (single-nucleotide polymorphisms) have further documented substantial geographical variation in this species, both latitudinally and across biomes [13]. All these characteristics make *A. olivacea* an excellent focal case for the study of geographical variation in response to environmental changes.

While a de novo transcriptome of *A. olivacea* has previously been reported [14, 15], a complete genome sequence for this species is not yet available. Therefore, we sequenced the entire genome of the olive mouse and subsequently performed its assembly and annotation. Here, we report the first de novo scaffold-level genome assembly for the olive mouse based on short-read DNA sequencing. The genome assembly yielded a 2.25 Gb sequence with a scaffold N50 of 123 Mb. Benchmarking with Glires universal single-copy gene orthologs (BUSCOs) showed 98.8% completeness. Genome annotation resulted in the identification of 21,476 coding sequences. This breakthrough provides essential groundwork for numerous advancements. It enables genome-wide analyses, detection of structural variations, improves gene annotation accuracy, and facilitates comparative genomics. Furthermore, this genomic data is crucial for conservation efforts, aiding in the identification of genetically distinct populations, understanding their evolutionary potential, and developing strategies for their preservation amidst environmental challenges.

We leveraged the newly generated genome assembly to investigate gene expression differences in the kidneys of individuals from the southern Andean rainforests and the Patagonian steppe and to assess genetic divergence among populations. Furthermore, employing a deconvolution technique with the kidney single-cell profile of the house mouse (*Mus musculus*), we connected the observed gene expression variations to known kidney cell type functions. Also, beyond *A. olivacea*, other rodent models have been investigated for their physiological and transcriptomic responses to water limitation. In particular, the cactus mouse (*Peromyscus eremicus*), a desert specialist from North America, and the lesser Egyptian jerboa (*Jaculus jaculus*), native to the Saharan and Middle Eastern deserts, have been studied under experimental dehydration challenges [16, 17]. The availability of genome assemblies and transcriptomic data for these species offers a valuable comparative framework to contextualize the responses observed in *A. olivacea*, enabling us to explore shared and lineage-specific mechanisms of kidney adaptation to arid environments.

Table 1 Statistics of genome sequencing of the Olive mouse *Abrothrix olivacea*

Library	Insert size	Read length	Total bases (Gbp)	Coverage percent
Pair-end libraries	400 bp	125	60.5	98.56
Mate-pair libraries	3 kb	125	5.5	92.21
	8 kb	125	13.3	90.41

Coverage was calculated based on the final assembled genome size

Methods

Study site and sample collection

The individual used for genome sequencing was a male *A. olivacea* (field catalog number GD1411; housed at Colección de Mamíferos Universidad Austral de Chile under number UACH 8896). The specimen was collected on April 21, 2011, at Fundo San Martín, San José de la Mariquina, Los Ríos Region, Chile (39°38.954′ S, 73°11.553′ W).

Kidney expression profiles were sourced from earlier studies that aimed to characterize the kidney transcriptome and deepen insights into the molecular adjustments to water availability [14, 15]. The raw sequences can be accessed in NCBI under BioProjects PRJNA471316 and PRJNA231324, as well as in Table S1.

Genomic DNA extraction, sequencing, and assembly

High molecular weight genomic DNA was extracted from frozen liver tissue using a standard DNA purification kit (Qiagen) following the manufacturer's recommendations. The extracted DNA was treated with RNase A to remove RNA contaminants. DNA quality and concentration were assessed with a NanoDrop spectrophotometer and a Qubit fluorometer, and integrity was confirmed by agarose gel electrophoresis. A whole-genome shotgun approach on the Illumina HiSeq™ 2500 platform was employed to sequence the genome. This process yielded a total of 60.5 Gb as a paired-end library with insert sizes of about 400 base pairs (bp) and 18.8 Gb as two mate-paired libraries with insert sizes of 3 kb and 8 kb (Table 1).

Initially, Jellyfish [18] and Genomescope 2.0 [19] were used to perform k-mer analysis by short insert size library reads before assembly; the genome size was estimated to be 2.25 Gb. Quality control on the raw sequence data was checked with FastQC v0.12.1 [20]; low-quality reads and adaptors were removed using Trim Galore v.0.6.10 [21].

The initial assembly was created with MaSuRCA v.4.1.4 [22]. Genome assembly was further aided by the

Redundans pipeline [23]. Scaffolding improvements were achieved using RagTag in scaffold mode [24], leveraging the *Phyllotis vaccarum* reference genome (GCA_033318265.1) to guide structural organization. Finally, abyss-sealer [25] was used to fill gaps in the scaffolds. Figure 1 illustrates this process.

Assembly quality assessment

To evaluate the base-level accuracy and completeness of the genome assembly independently of a reference genome, we utilized Merqury v1.3 [26]. A k-mer database (k = 21) was constructed from the high-quality, filtered Illumina short reads using Meryl. The k-mer spectrum of the reads was then compared against the *A. olivacea* assembly to calculate the consensus Quality Value (QV) and overall k-mer completeness.

To assess the assembly quality and derive comparative insights of *A. olivacea*, a comparative genomic framework was established using available assemblies from *Phyllotis vaccarum* (GCA_033318265.1), *Sigmodon hispidus* (GCA_047301775.1), *Holochilus nanus* (GCA_050578075.1), *Peromyscus maniculatus* (GCA_049852395.1), and *Mus musculus* (GCA_000001635.9) retrieved from the NCBI repository.

To provide a standardized comparison of genome contiguity and composition, we evaluated all assemblies using QUAST v5.3.0 [27]. The evaluation was performed simultaneously for all species using the –eukaryote and –large parameters, which are optimized for mammalian-sized genomes. The highly fragmented assembly of *H. nanus* (> 117,000 sequences) was included in this statistical comparison but subsequently excluded from the structural synteny analysis to prevent computational artifacts and severe visual overplotting caused by excessive short contigs.

Macrochromosomal synteny analysis

To trace structural rearrangements, a multi-genome macrochromosomal synteny analysis was conducted using the ntSynt pipeline v1.05 [28]. To focus on large-scale structural conservation, all genome assemblies were first filtered to retain only scaffolds and chromosomes larger than 50 kb using SeqKit v2.13.0 [29].

The filtered assemblies (*A. olivacea*, *P. vaccarum*, *S. hispidus*, *P. maniculatus*, and *M. musculus*) were jointly analyzed in ntSynt to compute a multi-genome synteny block graph. Given the phylogenetic span, the sequence divergence parameter (-d) was set to 0.20. For visualization, the synteny blocks were parsed using the ntSynt-viz tool v1.0.2 [30].

Also, genome completeness was assessed using BUSCO v.5.8.0 [31, 32] and compleasm v.0.2.6 [33] with the orthodb (v12) glires lineage dataset. Summary statistics were calculated with gfastats v1.3.6 [34]. To evaluate assembly quality, we analyzed the genome assembly with BlobToolKit v. 4.4.6 [35].

Structural and functional annotation

Before genome annotation, repeat elements were identified and characterized using RepeatModeler v2.0.5 [36] to generate a de novo repeat library. Subsequently, repetitive elements were masked using RepeatMasker v4.1.5 [37], incorporating both the de novo library and known repeats from RepBase. Two independent gene

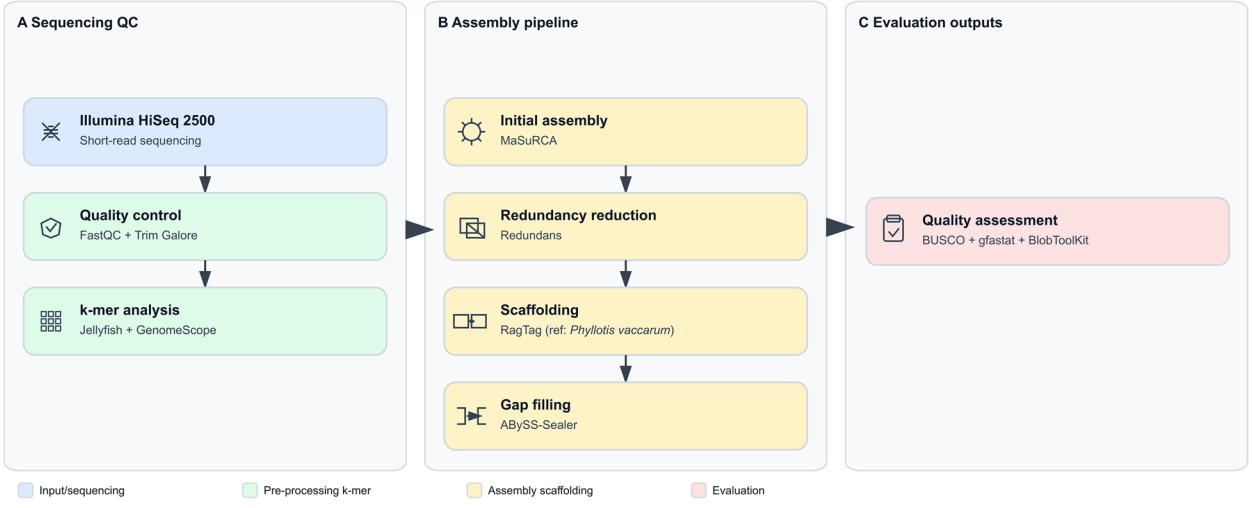

Fig. 1 Workflow of the genome assembly and annotation of the olive mouse. Panel **A** shows sequencing and pre-processing steps, including Illumina short-read sequencing, quality control, and k-mer analysis. Panel **B** summarizes the assembly pipeline, consisting of initial assembly with MaSuRCA, redundancy reduction with Redundans, scaffolding with RagTag using *Phyllotis vaccarum* as reference, and gap filling with ABySS-Sealer. Panel **C** presents the evaluation stage, where assembly completeness and quality were assessed with Compleasm gfastats and BlobToolKit. Colors indicate different categories: input/sequencing (blue), pre-processing (green), assembly/scaffolding (yellow), and evaluation (pink)

annotations were generated. First, BRAKER3 v3.0.8 [38] was run using RNA-seq data aligned to the genome as transcriptomic evidence, together with protein homology evidence from the Glires RefSeq protein database. Second, gene prediction transfer was performed using LiftOn v1.0.5 [39], employing *Mus musculus* (GCF_000001635.27) as reference genome annotation. Finally, both annotations were merged using AGAT tools (agat_sp_merge.pl and agat_sp_fix_features_locations_duplicated.pl) [40] to generate the final consensus annotation. The structural annotation obtained was also analyzed using BUSCO v.5.8.0.

Differential expression

Kidney RNA-Seq data were initially quality-trimmed using TrimGalore v0.6.10. High-quality reads were subsequently aligned to the assembled genome using HISAT2 v2.2.1 [41]. Gene-level read counts were obtained with FeatureCounts [42]. To minimize noise and ensure robust results, only genes with counts per million (CPM) ≥ 1 in at least one experimental group were retained for downstream analyses. Consistency among biological replicates was evaluated using principal component analysis (PCA) implemented via the prcomp function in R [43]. Read counts were normalized using the Trimmed Mean of M-values method (TMM) and differential expression analysis was conducted using the edgeR [44] and limma [45] packages. The model included a single main factor (biome), comparing forest and steppe samples. Genes with adjusted p-values < 0.05 [46] were considered significantly differentially expressed (DEGs).

In parallel, the publicly available datasets from *Peromyscus eremicus* (BioProject PRJEB19330; [16]) and *Jaculus jaculus* (BioProject PRJNA935758; [17]) were reanalyzed but using the newly available genomes as reference (GCA_949786415.1 and, GCA_020740685.1, respectively), and all available samples from that study were processed for comparative purposes.

Biological and functional analysis of DEGs

The ClusterProfiler v4.12.0 package [47] was utilized for gene set enrichment analysis (GSEA) of DEGs. This analysis was conducted on the complete set of genes using the Mouse MSigDB [48, 49]. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) were tested separately for each data set. Furthermore, to identify the source of DEGs at the cell type level from bulk RNA-seq data, a user-defined criterion of cell type signature gene set was created containing cluster marker genes for cell types identified in single-cell sequencing studies of mouse kidney [50–52]. Normalized enrichment score (NES) was used to indicate the strength of enrichment. |NES| ≥ 1.0 and adjusted or nominal p-value < 0.05 were defined as statistical significance.

Transcriptomic deconvolution of bulk kidney RNA-seq

To estimate the cellular composition of kidney bulk RNA-seq samples from *Abrothrix olivacea*, we performed transcriptomic deconvolution using CIBERSORTx [53]. As a reference, we used the Tabula Muris Droplet dataset [54], a publicly available scRNA-seq atlas of the house mouse (*Mus musculus*). We filtered the dataset to retain only kidney-derived cells. To generate the cell-type signature matrix, we used the AggregateExpression function from Seurat v5 [55] to compute the average expression per annotated cell type, based on raw counts. This matrix was used as input for CIBERSORTx. Bulk RNA-seq counts were obtained using featureCounts [42] and preprocessed with edgeR [44]. Genes with zero counts across all samples were removed. Normalization to counts per million (CPM) was performed using edgeR. The resulting CPM matrix was used as input for CIBERSORTx along with the cell-type signature matrix derived from mouse kidney scRNA-seq data. Deconvolution was run with 100 permutations using default settings. Given the phylogenetic distance between *M. musculus* and *A. olivacea*, gene identifiers were manually curated to retain only orthologous genes shared across species, ensuring compatibility between bulk and single-cell matrices.

Mitogenome assembly and annotation

The mitogenome was assembled, annotated and visualized using MitoZ v3.6 [56]. The mitogenome of *Akodon montensis* was used as the starting sequence (NC_025746.1). The resulting circularised mitogenome was 16428 bp in length and contained the standard 37 vertebrate mitochondrial genes (13 protein-coding, 22 tRNAs, and 2 rRNAs) (Figure S1).

Results and discussion

Genome assembly and annotation

We obtained a high-contiguity de novo genome assembly for *A. olivacea* with a total length of ~ 2.25 Gb and a heterozygosity rate of 0.942% (Fig. 2). The largest scaffold reached 200.1 Mb, and the assembly showed an excellent scaffold N50 of 122.5 Mb with an L50 of 8. The assembly contains 630,661 gaps, totaling 7.78 Mb, which represents only 0.35% of the entire sequence. The resulting assembly exhibits a high level of completeness, as assessed by compleasm, which indicated that 98.81% of the expected Glires single-copy orthologs were successfully recovered (Table 2). This project has been deposited at NCBI under the number PRJNA471316.

To rigorously evaluate the quality of our assembly in an evolutionary context, we performed a standardized comparative analysis against publicly available genomes of related rodents (Table 2). When comparing genomic metrics, it is essential to consider that the differences in genome size and GC content observed

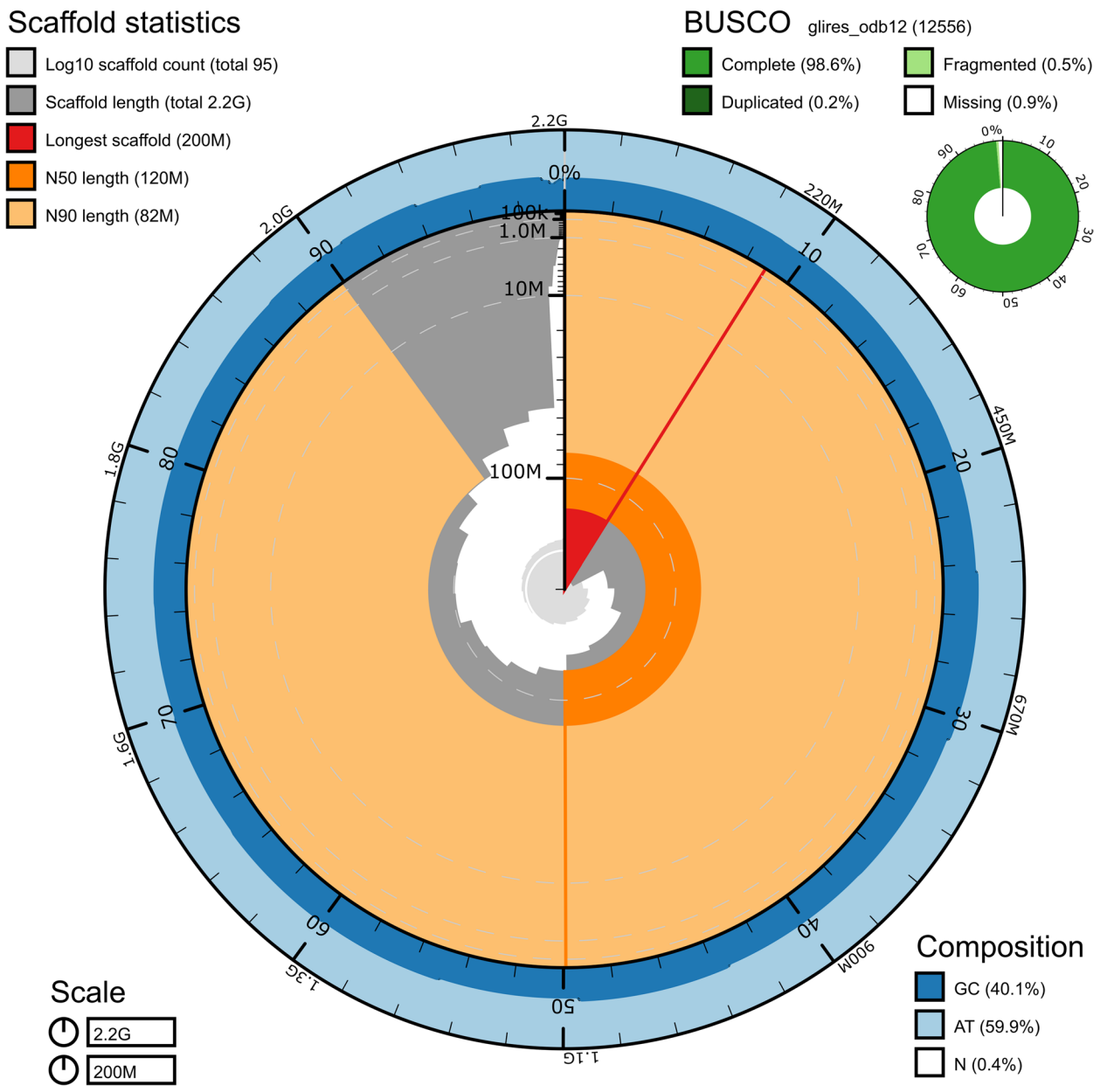

Fig. 2 BlobToolKit snail plot summarizing key statistics of the genome assembly, including scaffold length distribution, N50, GC content, and completeness (BUSCO scores). This visualization provides an overview of the assembly quality and characteristics

between *A. olivacea* and the selected reference species are heavily influenced by the sequencing strategy (Illumina vs. PacBio/Nanopore) rather than being purely driven by evolutionary divergence. Short-read sequencing (Illumina) inherently struggles to resolve long, complex repetitive elements, such as transposable elements and GC-rich centromeric or telomeric satellites. These regions frequently collapse during the assembly graph resolution. Consequently, short-read assemblies like ours often present an artificially compressed total genome size (~ 2.2 Gb) and a slightly lower overall GC content (~ 40%) compared to third-generation long-read assemblies, which can successfully span and incorporate these massive repetitive blocks. While true evolutionary expansions or contractions of repetitive landscapes likely exist among these rodents, disentangling biological divergence from short-read assembly collapse requires cautious interpretation.

Furthermore, reference-free k-mer analysis using Merqury demonstrated a high base-level accuracy. The assembly achieved a highly reliable consensus Quality Value (QV) of 51.3, corresponding to a remarkably low error rate of approximately 7.36 errors per million bases. The overall k-mer completeness was 84.88%.

Table 2 Comparative statistics of genome assembly and completeness across selected species

Feature	*Sigmodon hispidus*	*Abrothrix olivacea*	*Phyllotis vaccarum*	*Mus musculus*	*Holochilus nanus*
# contigs	39	95	303	60	86,722
Largest contig	243,842,572	200,137,661	206,300,460	195,154,279	534,501
Total length	2,668,580,129	2,247,830,787	2,660,325,576	2,728,220,475	2,151,540,032
GC (%)	39.96	40.15	40.01	41.67	40.5
N50	118,707,662	122,490,069	147,873,771	130,530,862	44,122
N90	64,898,139	81,833,982	59,266,063	95,294,699	10,740
auN	126,873,257	128,825,928	126,465,570	137,559,669	60,883.4
L50	9	8	8	9	13,595
L90	21	16	19	18	51,663
# N's per 100 kbp	0.14	437.01	0.49	2697.75	1899.75
Compleasm(%)	99.65	98.81	99.62	99.98	69.25
Sequencing technology	PacBio Sequel	Illumina	PacBio; Illumina	Multiple	Illumina NovaSeq

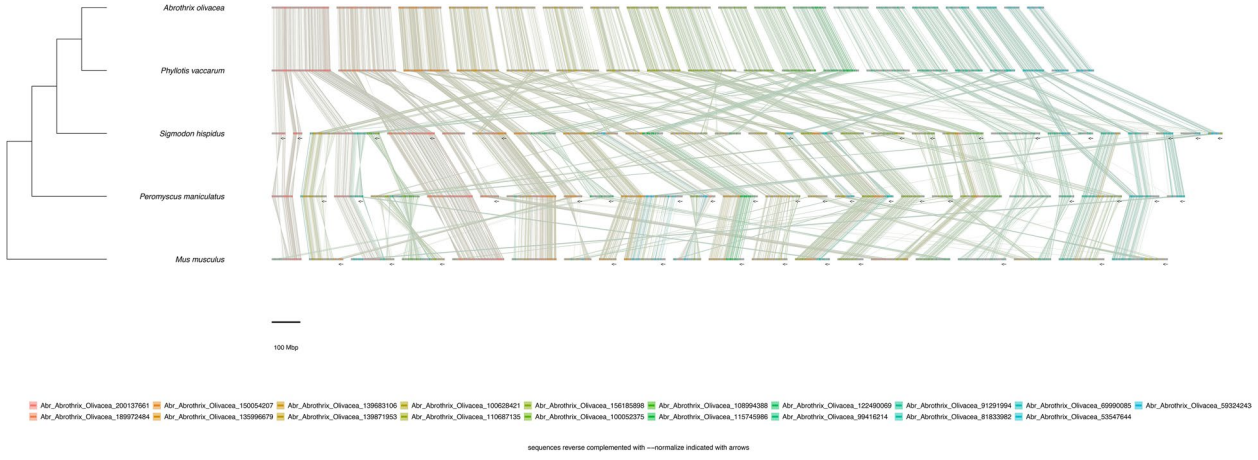

Fig. 3 Macrochromosomal synteny. The ribbon plot illustrates the conservation and rearrangement of syntenic blocks between *A. olivacea* and related rodent species. Genomes are ordered according to the phylogenetic topology shown on the left (Newick tree)

This completeness value is consistent with the expected performance of short-read assemblies in mammalian genomes, as it reflects the accurate reconstruction of the unique euchromatic space while capturing the inherent collapse of complex, highly repetitive heterochromatic regions.

To explore the structural evolution of these genomes, we performed a macrochromosomal synteny analysis (Fig. 3). The analysis revealed a highly complex pattern of structural rearrangements and fragmented syntenic blocks across the evaluated lineages. While strong macrochromosomal colinearity was observed between *A. olivacea* and *P. vaccarum,* this high degree of conservation must be interpreted cautiously, as it largely reflects the methodological reliance on the *P. vaccarum* assembly for the reference-guided scaffolding of our *A. olivacea* contigs.

Overall, the complex structural divergence observed underscores a major ongoing limitation in Neotropical rodent genomics, the severe lack of genomic data for intermediate phylogenetic nodes. The Sigmodontinae subfamily represents one of the most explosive mammalian radiations, likely accompanied by rapid karyotypic evolution [57]. Until a broader representation of chromosome-level genomes from diverse sigmodontine tribes becomes available, accurately reconstructing the precise evolutionary history of chromosomal rearrangements across this clade remains a significant challenge.

In addition, RNA-seq read mapping confirmed that the assembled transcripts were well represented within the *A. olivacea* scaffolds. As an example, Fig. 4 shows RNA-seq read alignments across the *Atp1a2* gene, which is expressed at moderate levels and has a typical multiexonic structure. The uniform read coverage across exons and the presence of clear exon–intron boundaries support the quality and completeness of the genome assembly. A total of 21,476 protein-coding genes were predicted in the assembled genome with an average coding DNA sequence length of 1,711 bp (Table 3).

Dataset for ecotype-associated targets

To assess the impact of using a complete reference genome on differential expression analyses, we revisited

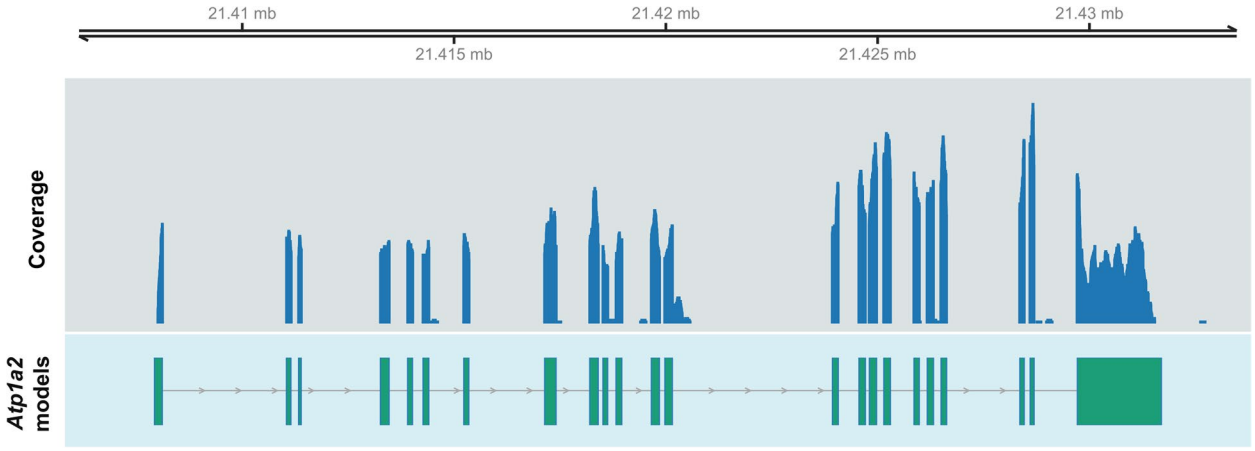

Fig. 4 RNA-seq reads coverage across the *Atp1a2* gene in *A. olivacea*. The top panel shows the genomic region containing the *Atp1a2* locus, while the middle panel displays the RNA-seq read alignments. The bottom panel shows transcript models predicted from the RNA-seq data. Exon coverage and splice junctions are well supported, illustrating the completeness of the assembled genome and its utility for transcriptomic analyses

Table 3 Summary of protein-coding gene annotation

Feature	Value
Number of genes	32,647
Number of transcripts	36,625
Number of protein-coding sequences (CDS)	21,476
Number of exons	268,728
Genes with a single exon	8,212
Average gene length (bp)	28,900
Average CDS length (bp)	1,711
Average transcript length (bp)	30,716
Average exons per transcript	7.3
Average introns in CDS per transcript	4.8
% genome covered by genes	42.00%
% genome covered by CDS	1.60%
Longest gene (bp)	2,100,374
Longest CDS (bp)	88,110
Longest intron (bp)	768,181

the data previously analyzed by Giorello et al. (2018), who employed a de novo transcriptome assembly for *A. olivacea* to assess differential expression profiles in forest and steppe samples. In our study, we used the newly assembled reference genome of the species for read mapping and quantification. Using the genome-based approach, we identified 985 differentially expressed genes (FDR < 0.05) between forest and steppe individuals, whereas the previous de novo transcriptome-based analysis reported 2,860 differentially expressed transcripts under the same significance threshold. Among these, 523 genes were shared between both approaches. The observed differences likely reflect the higher fragmentation and potential redundancy inherent to de novo assemblies, as well as the improved read assignment and gene annotation provided by the reference genome. These results highlight the importance of having a high-quality genome to optimize comparative transcriptomic analyses in non-model species (Fig. 4). However, it is important to note that many genes were exclusively detected as differentially expressed in the previous analysis by Giorello et al. (2018). These may include lineage-specific or poorly annotated transcripts, isoform variants, or tissue-specific genes that are underrepresented or fragmented in the current genome annotation. Therefore, both strategies are complementary, the reference genome enhances confidence and interpretability, while transcriptome-based assemblies may still capture biologically relevant elements missed by automated annotation. Future annotation efforts integrating long-read sequencing and manual curation will be essential to reconcile these two views and fully capture the transcriptional complexity of *A. olivacea*.

To further investigate the sources and pathways associated with these DEGs, we conducted GSEA using the C5 ontology gene sets as references where 143 significant clusters were detected (Table 4 and Table S2). The analysis revealed that DEGs associated with the forest ecotype were significantly enriched in several immune-related processes, including *antigen processing and presentation*, *regulation of T cell mediated cytotoxicity*, and *response to interferon beta*. This suggests an upregulation of both innate and adaptive immune responses in forest-dwelling individuals, potentially reflecting increased exposure to pathogens in more humid environments or ecotype-specific immune strategies. In contrast, DEGs associated with the steppe ecotype (i.e., gene sets with negative enrichment scores) were predominantly enriched in developmental and metabolic processes such as *kidney morphogenesis*, *epithelial cell differentiation*, *lipid metabolism*, and *thermogenesis*. These functions are consistent with adaptive responses to arid and colder environments, where efficient water reabsorption and non-shivering thermogenesis may be particularly

Table 4 Top 10 positively and negatively enriched gene sets identified by GSEA

Description	setSize	enrichmentScore	NES	*P* value	p.adjust
ANTIGEN PROCESSING AND PRESENTATION	47	0.67	2.12	0	0
ANTIGEN PROCESSING AND PRESENTATION OF PEPTIDE ANTIGEN	25	0.7	1.92	0	0.02
REGULATION OF T CELL MEDIATED CYTOTOXICITY	16	0.77	1.91	0	0.03
ANTIGEN PROCESSING AND PRESENTATION OF PEPTIDE ANTIGEN VIA MHC CLASS I	16	0.76	1.86	0	0.03
HORMONE BIOSYNTHETIC PROCESS	23	0.69	1.85	0	0.03
RESPONSE TO INTERFERON BETA	16	0.75	1.84	0	0.04
STEROID HORMONE BIOSYNTHETIC PROCESS	20	0.7	1.83	0	0.05
LEUKOCYTE MEDIATED CYTOTOXICITY	51	0.57	1.82	0	0.04
XENOBIOTIC METABOLIC PROCESS	67	0.53	1.81	0	0.02
LEUKOCYTE MEDIATED IMMUNITY	163	0.46	1.81	0	0.01
DETECTION OF ABIOTIC STIMULUS	45	−0.74	−2.28	0	0
DETECTION OF LIGHT STIMULUS	17	−0.86	−2.21	0	0
SENSORY PERCEPTION OF LIGHT STIMULUS	71	−0.65	−2.17	0	0
DETECTION OF STIMULUS	71	−0.64	−2.16	0	0
TRIGLYCERIDE METABOLIC PROCESS	54	−0.67	−2.15	0	0
SENSORY PERCEPTION	191	−0.53	−2.11	0	0
REGULATION OF PLASMA LIPOPROTEIN PARTICLE LEVELS	44	−0.68	−2.1	0	0
NEUTRAL LIPID BIOSYNTHETIC PROCESS	31	−0.72	−2.08	0	0.01
PROTEIN LIPID COMPLEX ASSEMBLY	17	−0.82	−2.08	0	0.01
REGULATION OF LIPID STORAGE	27	−0.72	−2.04	0	0.01

To facilitate interpretation, only the 10 gene sets with the highest and lowest normalized enrichment scores (NES) are shown. Positive NES values indicate enrichment in genes upregulated in forest ecotype samples, while negative NES values correspond to genes upregulated in steppe ecotype samples. GSEA was performed using GO: BP terms from the C5 collection of MSigDB

advantageous. Overall, the contrasting enrichment profiles between forest and steppe individuals underscore the ecological pressures shaping gene expression in each habitat, with the forest ecotype showing enhanced immune activity and the steppe ecotype exhibiting traits related to physiological adaptation to abiotic stress.

Comparison between A. olivacea and desert-adapted species DEGs

In order to compare our findings with those from another desert-adapted species, we conducted the DGE and GSEA on a dataset from the mouse *Peromyscus eremicus* exposed to experimental dehydration to examine the effects on desert-adapted animals [16]. To reanalyze the *P. eremicus* dataset, we used the newly available reference genome instead of the de novo transcriptome assembly employed in the original study. This allowed us to assess how the choice of reference influences the detection of DEGs. A similar pattern to that observed in *A. olivacea* emerged: 455 DEGs were identified by both approaches, whereas 751 were uniquely detected using the de novo transcriptome and 395 were exclusive to the genome-guided analysis.

When comparing the DEGs identified between the natural ecotypes of *A. olivacea* and those obtained from the experimental dehydration treatments in *P. eremicus*, only 65 genes were shared (Table 5 and Table S4). This limited overlap suggests that the molecular responses distinguishing the forest and steppe ecotypes of *A. olivacea* likely cannot be explained solely by dehydration stress. Instead, these differences may reflect broader ecological, physiological, and evolutionary processes, including long-term adaptation to multiple abiotic factors such as temperature, food availability, and photoperiod, as well as biotic interactions. However, among those shared DEGs, we highlight *Agt*, which encodes angiotensinogen, the precursor of the renin–angiotensin–aldosterone system (RAAS). This pathway plays a central role in water and electrolyte balance. The consistent differential expression of *Agt* in steppe-dwelling *A. olivacea* and dehydrated *P. eremicus* highlights a shared molecular signature of water restriction. Whereas in *P. eremicus* this reflects acute transcriptomic plasticity to experimental stress, in *A. olivacea* it may represent either a sustained physiological adjustment or an early local adaptation to xeric habitats. This parallel points to RAAS-related genes as core components of the mammalian response to water limitation. In addition to *Agt*, members of the solute carrier (Slc) family stand out due to their consistent expression patterns. Specifically, *Slc15a2*, *Slc3a*1, and *Slc7a13* were upregulated in forest-dwelling *A. olivacea* and in hydrated *P. eremicus*, suggesting their expression is associated with water availability. In contrast, *Slc2a1* was upregulated in steppe individuals and in dehydrated mice, pointing to a role in metabolic adjustments under water restriction. These cross-species signatures suggest that both acute dehydration stress and prolonged exposure to xeric habitats can recruit similar molecular pathways,

Table 5 Core candidate differentially expressed genes (DEGs) shared between *A. olivacea* (forest vs. steppe) and *P. eremicus* (hydrated vs. dehydrated)

Gene Symbol	logFC (*A. olivacea*)	FDR (*A. olivacea*)	logFC (*P.eremicus*)	FDR (*P.eremicus*)	Functional Description	Functional Category
Agt	−1.23	0.00536	−1.59	6.64E-05	angiotensinogen	RAAS pathway/Water balance
Slc15a2	1.96	1.38E-09	0.63	0.0148	solute carrier family 15 member 2	Solute transport
Slc3a1	0.71	0.000036	0.75	0.00517	solute carrier family 3 member 1	Solute transport
Slc7a13	2.75	0.000218	1.01	0.0124	solute carrier family 7 member 13	Solute transport
Slc2a1	−0.55	0.00238	−0.58	0.00061	solute carrier family 2 (glucose transporter)	Solute transport/ Metabolism
Fkbp5	−0.89	0.0372	−2.45	1.57E-07	FK506 binding protein 5	Cellular stress response
Hspa2	−0.52	0.00218	−1.54	8.64E-05	heat shock protein 2	Cellular stress response
Dnaja1	0.45	0.0477	0.31	0.0369	DnaJ heat shock protein family member A1	Cellular stress response
Dgat2	−1.51	0.000037	−0.62	0.0304	diacylglycerol O-acyltransferase 2	Lipid metabolism
Sqle	0.69	0.00321	1.48	1.29E-08	squalene epoxidase	Lipid metabolism
Rbp4	−2.19	0.000913	−2.23	5.44E-05	retinol binding protein 4 plasma	Lipid metabolism
Col15a1	0.9	0.0352	2.73	3.49E-09	collagen type XV alpha 1	ECM/Tissue remodeling
Fermt1	0.99	0.000143	0.75	0.00263	fermitin family member 1	ECM/Tissue remodeling
Itgb6	−0.54	0.0312	0.69	0.0177	integrin beta 6	ECM/Tissue remodeling
Fst	−0.96	0.000346	−0.66	0.02	follistatin	ECM/Tissue remodeling
Ptn	−0.57	0.0458	1.38	0.000188	pleiotrophin	ECM/Tissue remodeling

with Slc-mediated solute transport emerging as a central mechanism in mammalian water balance. Furthermore, stress-related genes (*Fkbp5*, *Hspa2*, *Dnaja1*) indicate the activation of cellular stress responses, while metabolic regulators (*Dgat2*, *Sqle*, *Rbp4*) point to adjustments in lipid metabolism. Together, these shared DEGs reinforce the notion that key pathways related to renal function, metabolism, and stress response are consistently recruited under xeric conditions.

Finally, genes associated with extracellular matrix and tissue remodeling (*Col15a1*, *Fermt1*, *Itgb6*, *Fst*, *Ptn*) suggest that structural changes in renal tissue could accompany these physiological responses. Together, these shared DEGs reinforce the notion that, although the overall transcriptomic overlap between *A. olivacea* and *P. eremicus* is limited, key pathways related to renal function, metabolism, tissue remodeling, and stress response are consistently recruited under xeric conditions.

In addition to *P. eremicus*, data from the desert rodent *Jaculus jaculus* under experimental dehydration further highlight shared transcriptional responses to water limitation [17]. Several key genes overlapped with those identified in *A. olivacea*, including *Ren* (renin) and *Aqp2*, central components of the RAAS pathway and collecting duct water reabsorption. Likewise, the glucose transporter *Slc2a1* and stress-related genes such as *Dnaja1* exhibited aridity-associated regulation. The recurrence of these genes across independent rodent lineages suggests that renal physiological responses to aridity may recruit a conserved molecular toolkit (Table S3).

Deconvolution

A kidney cell type signature gene set was constructed and implemented to identify the source of DEGs at the cell type level from olive mouse bulk RNA-seq data across both ecotypes (Fig. 5). The deconvolution analysis provided a complementary perspective on the renal transcriptomes of *A. olivacea*. As expected, epithelial cells from proximal tubules and the loop of Henle dominated the inferred cellular composition, reflecting their central roles in water and solute handling. Notably, forest individuals displayed higher relative contributions of proximal tubule cells, whereas steppe individuals showed increased signals from the loop of Henle and mesangial cells. Integrating deconvolution with differential expression analysis highlighted that the cellular basis of transcriptional divergence between forest and steppe ecotypes involves both renal and immune compartments. In the steppe, the overexpression of the collecting duct marker *Fxyd4* suggests a greater functional contribution of collecting duct epithelial cells, which may reflect adjustments in water and electrolyte handling under drier or more osmotically challenging conditions. In contrast, the forest exhibited higher expression of the proximal tubule marker *Enpep*, pointing to enhanced activity of proximal tubule epithelial cells, which play central roles in nutrient reabsorption and metabolic homeostasis. Additionally, the upregulation of macrophage-associated genes (*Csf1r* and *Cd68*) indicates an increased presence or activation of immune cells in the forest environment, potentially linked to differences in pathogen exposure or immune surveillance. Together, these results suggest that

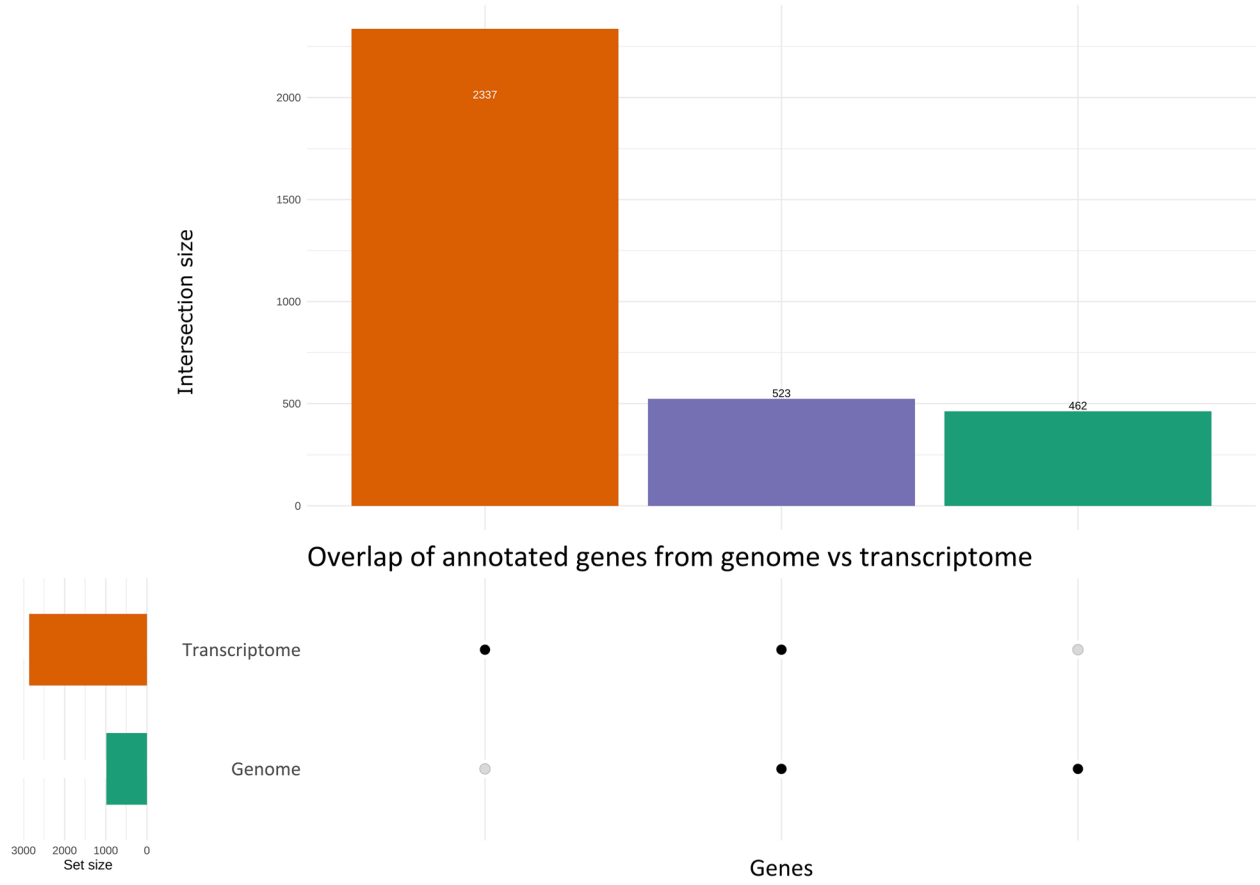

Fig. 5 Overlap of annotated genes identified from the genome (this study) and transcriptome analyses. UpSet plot showing the distribution of gene symbols recovered from genome-based annotation (green) and transcriptome-based annotation (orange). The intersection (purple) highlights genes detected by both approaches, while the vertical bars indicate the size of each set or intersection. Unique sets correspond to genes identified exclusively by one of the approaches. This comparison highlights the complementarity of genome and transcriptome data for gene discovery

ecological differences between biomes shape kidney transcriptomes by modulating the relative abundance and activity of distinct renal and immune cell populations. These cellular shifts parallel the observed transcriptional differences in solute carriers and metabolic genes, underscoring that ecotypic divergence involves not only gene-level changes but also the engagement of distinct cellular programs within the kidney. Such variation may underlie differences in urine concentrating ability and water conservation strategies, highlighting the multi-layered nature of adaptation to extreme environments (fig. 6).

Conclusion

This study presents the first high-quality genome assembly of the sigmodontine tribe Abrotrichini, the one that includes *Abrothrix olivacea*. This species thrives across contrasting biomes of southern South America. The newly generated genome represents a major resource for evolutionary and ecological genomics in sigmodontine rodents, providing a platform for studies of adaptation, physiology, and conservation. Beyond its potential value, we illustrated its utility through a case study of kidney transcriptomes of *A. olivacea* inhabiting forests and steppes.

Our analyses revealed striking ecological contrasts. Forest-dwelling individuals displayed strong enrichment for immune-related pathways, with upregulation of genes linked to macrophages, natural killer cells, and fibroblasts. Renal macrophages are known to play pivotal roles in immune surveillance and tissue homeostasis [58], and their increased signature in forest animals may reflect ecological pressures such as higher pathogen exposure in humid environments. Interestingly, experimental models of salt-sensitive hypertension in rodents demonstrate that dietary protein source modulates kidney injury through immune–microbiota interactions [59]. As such, the forest-associated immune activation in *A. olivacea* could represent not pathology but an adaptive adjustment, potentially shaped by dietary and environmental inputs. Parallels with human studies are also striking; even brief exposure to forest environments enhances innate immune activity, increasing NK cell function and reducing stress hormones [60]. Together, these findings support the notion that forest habitats

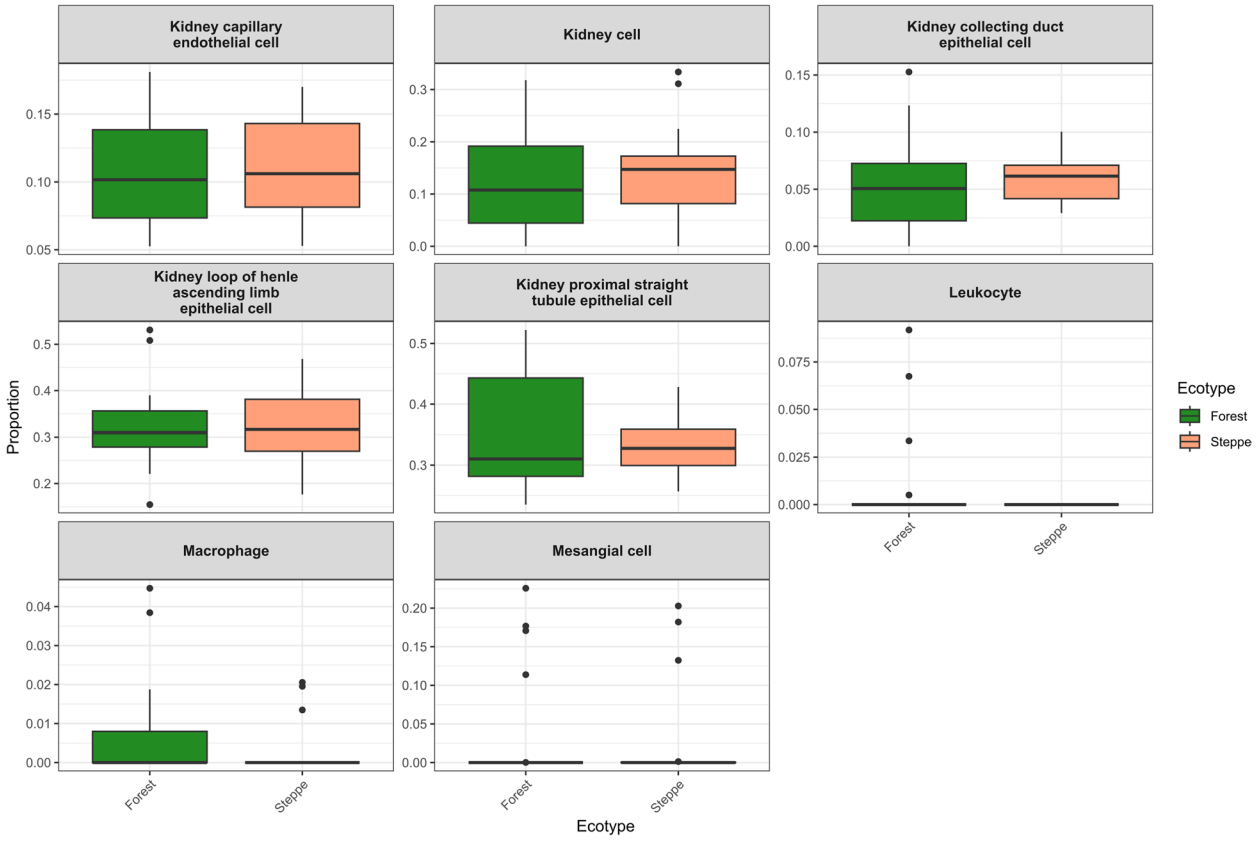

Fig. 6 Cell type deconvolution of kidney transcriptomes from *A. olivacea* using CIBERSORTx. Stacked barplots show the relative proportions of inferred renal cell types in individuals from forest and steppe ecotypes. Major contributions were observed from the proximal tubule and loop of Henle epithelial cells, with differences between ecotypes suggesting variation in cell-type representation linked to water availability. These results integrate with differential expression findings, indicating that ecotypic divergence involves both transcriptional and cellular-level changes in renal physiology

prime immunological states in mammals, a pattern now observed both in humans and in wild rodents.

Conversely, steppe individuals showed enrichment for developmental and metabolic pathways, including processes related to kidney morphogenesis, epithelial differentiation, lipid metabolism, and thermogenesis traits consistent with adaptation to arid, cold environments where efficient water reabsorption and energy balance are critical. These results align with convergent responses observed in desert rodents such as *Peromyscus eremicus*, in which RAAS-related genes and solute carriers underlie responses to water stress.

Finally, evidence from other mammalian systems underscores the broader eco-physiological framework. In wild woodrats, populations feeding on toxic creosote bush host a unique functional microbiome adapted for detoxification [61]. Analogously, environmental differences between forest and steppe may shape not only the renal transcriptome of *A. olivacea* but also its gut microbiome, selecting microbial consortia with metabolic capacities that support water and osmotic balance. Considering the interplay of microbiome function, immune cell activity, and renal gene expression highlights the complexity of adaptation in natural populations.

In sum, the genome of *A. olivacea* not only fills a critical gap in the genomic resources for sigmodontine rodents but also enables the dissection of multilevel adaptive processes, from gene regulation to cell-type composition and host–microbiome interactions. These findings exemplify how high-quality genomic resources in non-model species can illuminate the molecular architecture of adaptation to heterogeneous environments.

Limitations and future directions

While this study provides a comprehensive genomic and transcriptomic foundation for *Abrothrix olivacea*, several limitations must be acknowledged. First, regarding the genome assembly, our reliance on second-generation sequencing technologies (Illumina short-reads) inherently limits the resolution of highly repetitive and GC-rich heterochromatic regions. Consequently, metrics such as total genome size and global GC content may be underestimated due to assembly graph collapse. Future efforts utilizing third-generation long-read sequencing (e.g., PacBio HiFi or Oxford Nanopore) and Hi-C

scaffolding will be necessary to resolve the complete repetitive landscape and achieve a gapless, complete chromosome-level assembly.

Second, our bulk RNA-seq cellular deconvolution relied on a *Mus musculus* single-cell reference atlas. Although we mitigated cross-mapping artifacts by strictly restricting the reference matrix to highly conserved 1:1 structural orthologs (capturing core cell-identity markers), this approach inherently cannot resolve Abrothrix-specific cell states, novel transient subtypes, or lineage-specific gene expansions. Developing a species-specific single-cell RNA-seq atlas for *A. olivacea* remains a critical next step.

Finally, while our cross-species transcriptomic comparisons provide valuable exploratory insights into convergent molecular pathways for water conservation, comparing the intraspecific plasticity of *A. olivacea* with the deeply fixed adaptations of obligate desert specialists (*P. eremicus* and *J. jaculus*) requires cautious interpretation. The shared gene expression profiles observed may partially reflect conserved physiological stress responses rather than equivalent evolutionary adaptations. Future research should prioritize in vivo functional validations and population-level genomic sequencing to definitively decouple transient phenotypic plasticity from fixed genetic adaptations in these emerging rodent models.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12853-1.

Supplementary Material 1. Figure S1. Assembly and annotation of the mitochondrial genome of *A. olivacea*. The complete circularized mitogenome was reconstructed from Illumina reads and annotated using standard mitochondrial gene models. Protein-coding genes, rRNAs, and tRNAs are indicated with their respective orientation along the circular genome.

Supplementary Material 2. Table S1. Metadata for all RNA-seq samples utilized in this study.

Supplementary Material 3. Table S2. Complete results of Gene Set Enrichment Analysis (GSEA) contrasting kidney transcriptomes between forest and steppe ecotypes of *A. olivacea*. The analysis was performed using the MSigDB C5 collection (GO: Biological Processes). For each gene set, enrichment statistics are reported, including set size, enrichment score (ES), normalized enrichment score (NES), and associated significance values. Positive NES values indicate enrichment in forest samples, while negative NES values correspond to enrichment in steppe samples.

Supplementary Material 4. Table S3. List of differentially expressed genes (DEGs) detected in *Jaculus jaculus* in response to contrasting hydration states. For each gene, identifiers, log fold change values, adjusted *p*-values.

Supplementary Material 5. Table S4. Differentially expressed genes (DEGs) shared between *A. olivacea* (forest vs. steppe ecotypes) and *P. eremicus* (hydrated vs. dehydrated treatments). For each gene, log fold change (logFC) values in both species are shown together with a brief functional description. Despite the limited overlap (65 DEGs), several candidates are functionally relevant for dehydration response and renal physiology, including Agt, solute carriers, stress-related proteins, and genes involved in tissue remodeling.

Acknowledgements
We thank the Comisión Sectorial de Investigación Científica (CSIC), Universidad de la República (Uruguay), for financial support. Additional funding was provided by a postdoctoral fellowship from the Agencia Nacional de Investigación e Innovación (ANII, Uruguay).

Authors' contributions
M.M. conceived the study, performed genome assembly and annotation, and carried out transcriptomic analyses. A.B. contributed to all bioinformatics analyses, data interpretation, and manuscript revision. M.F. contributed to the differential expression and GSEA analyses. G.D. performed the deconvolution analysis and assisted in data interpretation. D.E.N. provided field samples, ecological background, and contributed to data interpretation. E.P.L. contributed to data interpretation, provided intellectual guidance, and critically revised the text. M.M. wrote the first draft of the manuscript. All authors read and approved the final version of the manuscript.

Funding
This work was supported by the Comisión Sectorial de Investigación Científica (CSIC), Universidad de la República (Uruguay), and by a postdoctoral fellowship from the Agencia Nacional de Investigación e Innovación (ANII, Uruguay).

Data availability
The genome assembly of *Abrothrix olivacea* has been deposited together with the raw Illumina sequencing reads available in the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA471316.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 16 September 2025 / Accepted: 7 April 2026
Published online: 16 April 2026

References

1. Waterhouse GR. Characters of new species of the genus Mus, from the collection of Mr. Darwin. In: Proceedings of the Zoological Society of London. 1837;(5)15–21.
2. Patton JL, Pardiñas UFJ, D'Elía G. Mammals of South America, Volume 2: Rodents. University of Chicago Press; 2015. Available from: http://www.biblio vault.org/BV.landing.epl?ISBN=9780226169576 https://doi.org/10.7208/chica go/9780226169606.001.0001. Cited 2024 May 27
3. Quiroga-Carmona M, D'Elía G. Climate influences the genetic structure and niche differentiation among populations of the olive field mouse Abrothrix olivacea (Cricetidae: Abrotrichini). Sci Rep. 2022;12(1):22395. https://doi.org/10.1038/s41598-022-26937-x.
4. McNab BK. The physiological ecology of vertebrates: a view from energetics. Ithaca, NY: Cornell Univ. Press; 2002. p. 576.
5. Al-kahtani MA, Zuleta C, Caviedes-Vidal E, Garland JT. Kidney mass and relative medullary thickness of rodents in relation to habitat, body size, and phylogeny. Physiol Biochem Zool. 2004;77(3):346–65. https://doi.org/10.1086/420941.
6. Bozinovic F, Rojas JM, Gallardo PA, Palma RE, Gianoli E. Body mass and water economy in the South American olivaceous field mouse along a latitudinal gradient: implications for climate change. J Arid Environ. 2011;75(5):411–5. https://doi.org/10.1016/j.jaridenv.2010.11.011.
7. Mann G. Los pequeños mamíferos de Chile [Small mammals of Chile]. Gayana. 1978;40:1–342.
8. Rodríguez-Serrano E, Cancino RA, Palma RE. Molecular phylogeography of *Abrothrix olivaceus* (Rodentia: Sigmodontinae) in Chile. J Mammal. 2006;87(5):971–80. https://doi.org/10.1644/05-MAMM-A-393R2.1.

9. Pizarro R, Valdés R, García-Chevesich P, Vallejos C, Sanguesa C, Morales C, et al. Latitudinal analysis of rainfall intensity and mean annual precipitation in Chile. Chil J Agric Res. 2012;72(2):252–61.
10. Quiroga-Carmona M, Teta P, D'Elía G. The skull variation of the olive field mouse Abrothrix olivacea (Cricetidae: Abrotrichini) is localized and correlated to the ecogeographic features of its geographic distribution. PeerJ. 2023;11:e15200. https://doi.org/10.7717/peerj.15200.
11. Lessa EP, D'Elía G, Pardiñas UFJ. Genetic footprints of late Quaternary climate change in the diversity of Patagonian-Fueguian rodents: phylogeography of Patagonian-Fueguian rodents. Mol Ecol. 2010;19(15):3031–7. https://doi.org/10.1111/j.1365-294X.2010.04734.x.
12. Quiroga-Carmona M, Abud C, Lessa EP, D'Elía G. The mitochondrial genetic diversity of the olive field mouse *Abrothrix olivacea* (Cricetidae; Abrotrichini) is latitudinally structured across its geographic distribution. J Mamm Evol. 2022;29(2):413–30. https://doi.org/10.1007/s10914-021-09582-5.
13. Giorello FM, D'Elía G, Lessa EP. Genomic footprints of Quaternary colonization and population expansion in the Patagonian-Fuegian region rules out a separate southern refugium in Tierra del Fuego. J Biogeogr. 2021;48(10):2656–70. https://doi.org/10.1111/jbi.14231.
14. Giorello FM, Feijoo M, D'Elía G, Valdez L, Opazo JC, Varas V, et al. Characterization of the kidney transcriptome of the South American olive mouse Abrothrix olivacea. BMC Genomics. 2014;15(1):446. https://doi.org/10.1186/1471-2164-15-446.
15. Giorello FM, Feijoo M, D'Elía G, Naya DE, Valdez L, Opazo JC, et al. An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea)*. Mol Ecol. 2018;27(16):3274–86. https://doi.org/10.1111/mec.14778.
16. MacManes MD. Severe acute dehydration in a desert rodent elicits a transcriptional response that effectively prevents kidney injury. Am J Physiol-Ren Physiol. 2017;313(2):F262–72. https://doi.org/10.1152/ajprenal.00067.2017.
17. Gillard BT, Amor N, Iraizoz FA, Pauža AG, Campbell C, Greenwood MP, et al. Mobilisation of jerboa kidney gene networks during dehydration and opportunistic rehydration. iScience. 2023;26(9):107574. https://doi.org/10.1016/j.isci.2023.107574.
18. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k* -mers. Bioinformatics. 2011;27(6):764–70. https://doi.org/10.1093/bioinformatics/btr011.
19. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11(1):1432. https://doi.org/10.1038/s41467-020-14998-3.
20. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
21. Krueger F, James F, Ewels P, Afyounian E, Weinstein M, Schuster-Boeckler B, et al. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. Zenodo; 2023. https://doi.org/10.5281/ZENODO.5127898. Cited 2024 Jun 6
22. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–77. https://doi.org/10.1093/bioinformatics/btt476.
23. Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 2016;44(12):e113. https://doi.org/10.1093/nar/gkw294.
24. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022;23(1):258. https://doi.org/10.1186/s13059-022-02823-7.
25. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 2017;27(5):768–77. https://doi.org/10.1101/gr.214346.116.
26. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):245. https://doi.org/10.1186/s13059-020-02134-9.
27. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142. https://doi.org/10.1093/bioinformatics/bty266.
28. Coombe L, Kazemi P, Wong J, Birol I, Warren RL. NtSynt: multi-genome synteny detection using minimizer graph mappings. BMC Biol. 2025;23(1):367. https://doi.org/10.1186/s12915-025-02455-w.
29. Shen W, Sipos B, Zhao L. SeqKit2: a swiss army knife for sequence and alignment processing. iMeta. 2024;3(3):e191. https://doi.org/10.1002/imt2.191.
30. Coombe L, Warren RL, Birol I. ntSynt-viz: Visualizing synteny patterns across multiple genomes. Genomics; 2025. https://doi.org/10.1101/2025.01.15.633221 Cited 2026 Mar 3
31. 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):3210–2. https://doi.org/10.1093/bioinformatics/btv351.
32. Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res. 2025;53(D1):D516–22. https://doi.org/10.1093/nar/gkae987.
33. Huang N, Li H. Compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023;39(10):btad595. https://doi.org/10.1093/bioinformatics/btad595.
34. Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. Bioinformatics. 2022;38(17):4214–6. https://doi.org/10.1093/bioinformatics/btac460.
35. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. Blobtoolkit – interactive quality assessment of genome assemblies. G3 Genes|Genomes|Genet. 2020;10(4):1361–74. https://doi.org/10.1534/g3.119.400908.
36. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020;117(17):9451–7. https://doi.org/10.1073/pnas.1921046117.
37. Tarailo-Graovac M, Chen N. Using repeatmasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinforma. 2009;25(1). https://doi.org/10.1002/0471250953.bi0410s25
38. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. Bioinformatics; 2023. https://doi.org/10.1101/2023.06.10.544449 Cited 2025 Jun 10
39. Chao KH, Heinz JM, Hoh C, Mao A, Shumate A, Pertea M, et al. Combining DNA and protein alignments to improve genome annotation with LiftOn. Genome Res. 2024;genome;gr.279620.124v2. https://doi.org/10.1101/gr.279620.124
40. Jacques Dainat, Darío Hereñú, Dr. K. D. Murray, Ed Davis, Ivan Ugrin, Kathryn Crouch, et al. NBISweden/AGAT: AGAT-v1.4.1. Zenodo; 2024. https://doi.org/10.5281/ZENODO.3552717. Cited 2024 Sep 30
41. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60. https://doi.org/10.1038/nmeth.3317.
42. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30. https://doi.org/10.1093/bioinformatics/btt656.
43. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2025. Available from: https://www.R-project.org/
44. Robinson MD, McCarthy DJ, Smyth GK. EdgeR : a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616.
45. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
47. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7. https://doi.org/10.1089/omi.2011.0118.
48. Dolgalev I. msigdbr: MSigDB gene sets for multiple organisms in a tidy data format. 2022. Available from: https://CRAN.R-project.org/package=msigdbr
49. Castanza AS, Recla JM, Eby D, Thorvaldsdóttir H, Bult CJ, Mesirov JP. Extending support for mouse data in the Molecular Signatures Database (MSigDB). Nat Methods. 2023;20(11):1619–20. https://doi.org/10.1038/s41592-023-02014-7.
50. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science. 2018;360(6390):758–63. https://doi.org/10.1126/science.aar2131.
51. Agarwal S, Sudhini YR, Polat OK, Reiser J, Altintas MM. Renal cell markers: lighthouses for managing renal diseases. Am J Physiol-Ren Physiol. 2021;321(6):F715–39. https://doi.org/10.1152/ajprenal.00182.2021.
52. Balzer MS, Rohacs T, Susztak K. How many cell types are in the kidney and what do they do? Annu Rev Physiol. 2022;84(1):507–31. https://doi.org/10.1146/annurev-physiol-052521-121841.

53. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773–82. https://doi.org/10.1038/s41587-019-0114-2.
54. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature. 2018;562(7727):367–72. https://doi.org/10.1038/s41586-018-0590-4 PubMed PMID: 30283141; PubMed Central PMCID: PMC6642641.
55. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304. https://doi.org/10.1038/s41587-023-01767-y.
56. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019;47(11):e63. https://doi.org/10.1093/nar/gkz173.
57. Parada A, Pardiñas UFJ, Salazar-Bravo J, D'Elía G, Palma RE. Dating an impressive Neotropical radiation: Molecular time estimates for the Sigmodontinae (Rodentia) provide insights into its historical biogeography. Mol Phylogenet Evol. 2013;66(3):960–8. https://doi.org/10.1016/j.ympev.2012.12.001.
58. Bell RMB, Conway BR. Macrophages in the kidney in health, injury and repair. In: International review of cell and molecular biology. Elsevier; 2022. p. 101–47. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1937644822000053 https://doi.org/10.1016/bs.ircmb.2022.01.005. Cited 2024 Jul 31
59. Mattson DL, Dasinger JH, Abais-Battad JM. Dietary protein, chronic salt-sensitive hypertension, and kidney damage. Kidney360. 2023;4(8):1181–7. https://doi.org/10.34067/KID.0000000000000210.
60. Li Q, Kobayashi M, Inagaki H, Hirata Y, Li YJ, Hirata K, et al. A day trip to a forest park increases human natural killer activity and the expression of anti-cancer proteins in male subjects. J Biol Regul Homeost Agents. 2010;24(2):157–65 (**PubMed PMID: 20487629**).
61. Stapleton TE, Lindsey LM, Sundar H, Dearing MD. Rodents consuming the same toxic diet harbor a unique functional core microbiome. Anim Microbiome. 2024;6(1):43. https://doi.org/10.1186/s42523-024-00330-7.

Publisher's Note