Supplementary Information

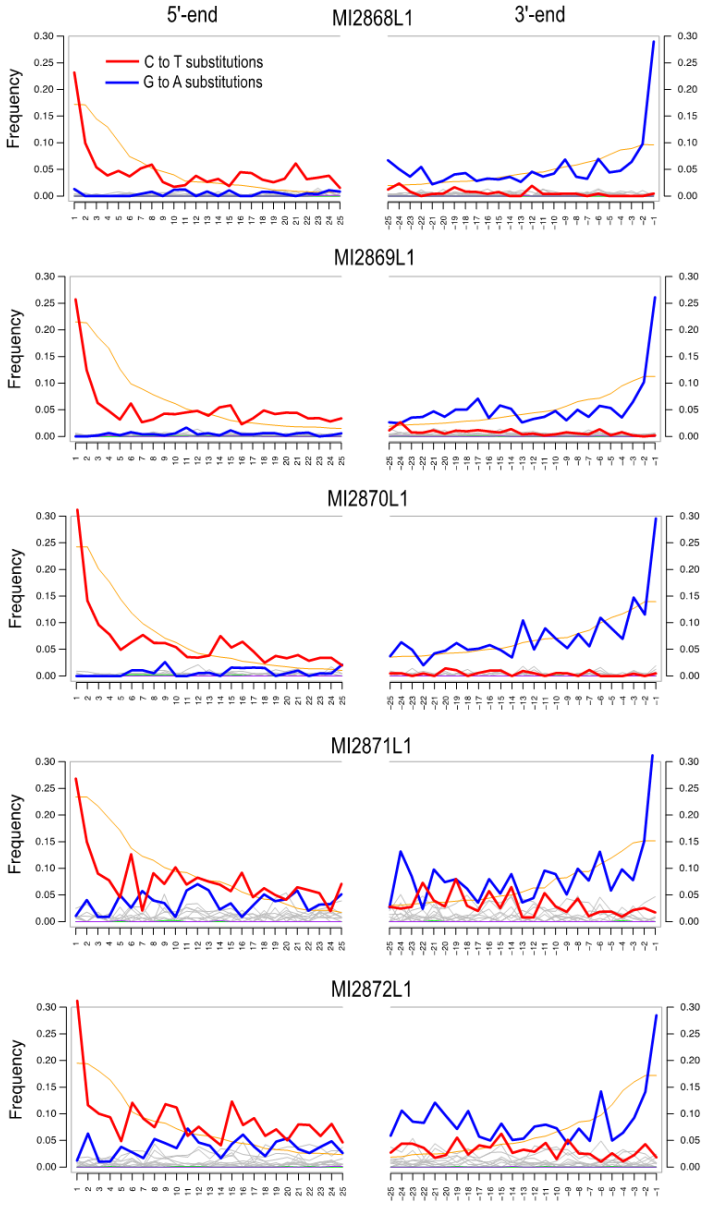

Figure S1. DNA damage patterns produced by mapDamage software for double-stranded libraries produced in this study. The double-stranded libraries were not USER treated.

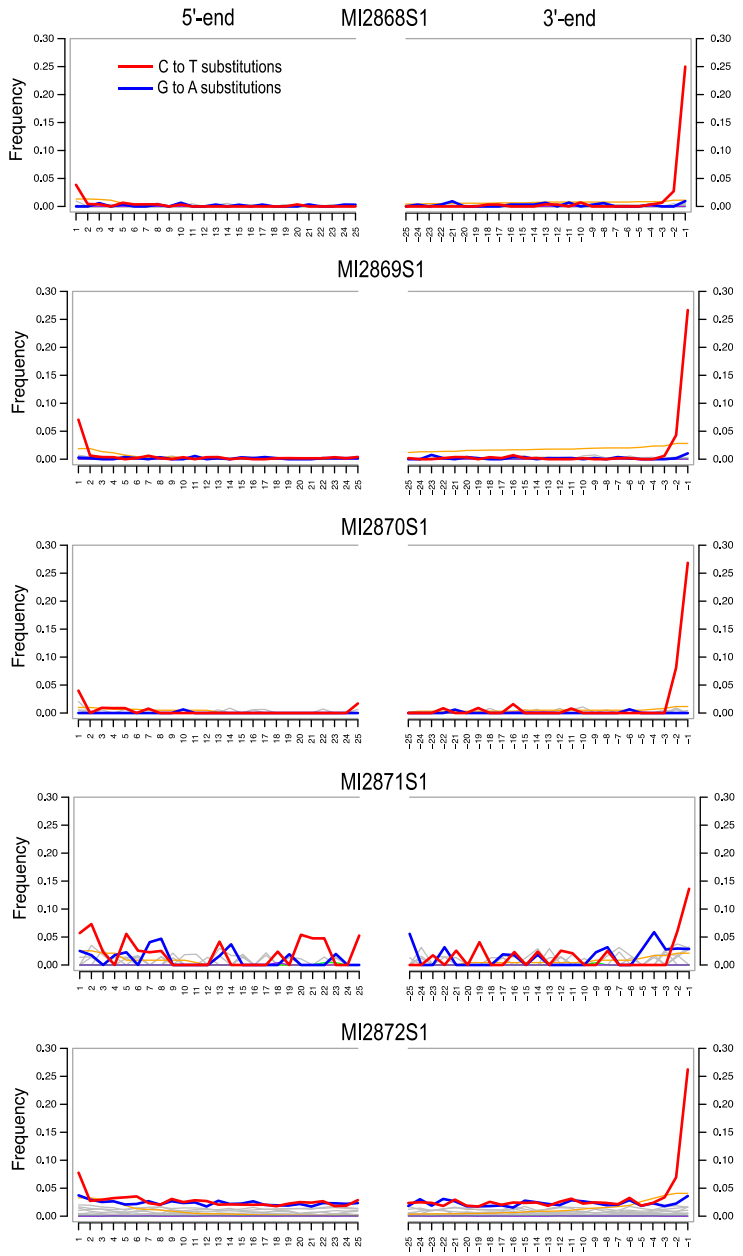

Figure S2. DNA damage patterns obtained using mapDamage software for single-stranded (SS) libraries produced in this study. SS libraries were treated with partial USER protocol, allowing to preserve excess of deamination only at terminal nucleotides.

1. Supplementary Materials and Methods

1.1. Double-stranded library preparation

Double-indexed, double-stranded sequencing libraries were constructed following the protocol of Meyer and Kircher [18] with minor modifications described in Baca et al. [6] using 20 μL of DNA extract as input. The blunt-end repair was performed in a 30 μL reaction containing 1× buffer Tango, 15 U T4 polynucleotide kinase (Thermo, Waltham, MA, USA), 3U T4 DNA polymerase, 100 μDNTPs, and 1 mM ATP. The reaction was incubated for 15 min at 25 °C, followed by 5 min at 12 °C and 20 min at 95 °C to inactivate the enzymes. An adapter ligation step was performed by adding 10 μL of the adapter ligation mix directly to the blunt-end repair reaction resulting in a final reaction volume of 40 μL containing 1× T4 DNA ligase buffer, 5% PEG-4000, 5 U T4 DNA ligase (Thermo Scientific, Waltham, MA, USA) and 1 μM of the P5 and P7 adapters. The reaction was incubated for 30 min at 22 °C and purified using magnetic beads. Adapter fill-in was performed by adding 20 μL of purified ligation product to 15 μL of the reaction master mix, resulting in a 35 μL reaction containing 9.6 U of BST polymerase (New England Biolabs, Ipswich, MA, USA), 1× Thermopol buffer and 0.25 μM of the dNTPs. The reactions were incubated in a thermocycler for 20 min at 37 °C, followed by heat inactivation at 80 °C for 20 min. DNA libraries were amplified in three replicates in 25 μL reaction volumes containing 10 μL of adapter-ligated DNA, 1× AmpliTaq Gold 360 Master Mix (Thermo Scientific, Waltham, MA, USA) and 0.2 μM of the P5 and P7 indexing primers under the following conditions: 95 °C for 12 min, 19 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min and 72 °C for 10 min. Each indexing primer contained a 7-bp long index. The amplification replicates were combined and purified using magnetic beads. The libraries were visualised by 2% agarose gel electrophoresis, quantified with the Denoxiv spectrophotometer and subjected to the target enrichment procedure.

1.2. Single-stranded library preparation

We prepared single-stranded libraries for all five specimens to confirm results obtained from double-stranded ones and to generate enough molecules mapping to vole mtDNA to call mitogenomes. Twenty microlitres of DNA extract were combined with 9 μl of water and 1 μL of USER enzyme (New England Biolabs, Ipswich, MA, USA) and incubated for 1 h at 37 °C. Further steps strictly followed the protocol outlined by Gansauge et al. [19]. The appropriate number of PCR cycles was determined with qPCR using the Illumina Library Quantification kit before indexing (KAPA). Indexing PCR was performed in duplicate using AccuPrime™ *Pfx* DNA Polymerase (Thermo Scientific, Waltham, MA, USA). Amplified libraries were combined, purified using magnetic beads, and subjected to the target enrichment procedure.

1.3. Target enrichment of mtDNA

Target enrichment was performed to enrich the libraries with vole mtDNA. Hybridisation bait was produced using the modern DNA of the following vole species: common vole (*Microtus arvalis*), field vole (*M. agrestis*), root vole (*Alexandromys oeconomus*), narrow-headed vole (*Stenocranius gregalis*) and bank vole (*Clethrionomys glareolus*). Total genomic DNA was extracted from tissue fragments using the Syngen Tissue DNA Mini Kit The mitogenomes were amplified in four overlapping fragments using PrimeSTAR GXL DNA Polymerase (Takara Bio, Shiga, Japan).

Table S1. Primers used to generate the mitogenomes from various vole species.

Primer ID	Sequence	Product length
MICMT01F	TGCAAGCATCCCATAAACAA	3.8 kb
MICMT01R	ATGGGCCCGATAGCTTTATT	
MICMT02F	CAAAATTCTCCGTGCTACCC	4.4 kb
MICMT02R	TTGTGTGGTTGGGGTAAATG	
MICMT03F	CGCCTCTTTCATTACCCCTA	4.2 kb
MICMT03R	TCYCAGCCGATGAAGAGTTG	
MICMT04F	ACCCHAACCTAAACCGATTC	4.5 kb
MICMT04R	ATAAGGCCAGGACCAAACCT	

Each fragment was amplified separately. The amplification reaction was carried out in 50 µl and consisted of 20–50 ng of genomic DNA, 1× PrimeSTAR GXL buffer, dNTPs (200 µM each), 0.2 µM primers and 2.5 U PrimeSTAR GXL DNA polymerase. The PCR conditions were 30 cycles at 98 °C for 10 s, 55 °C for 15 s and 68 °C for 30 to 50 s depending on the target length.

The PCR products were mixed in equimolar ratios and sonicated to a length of ca. 200 bp using the Covaris S220 sonicator. The sonicated DNA from various species was pooled and converted into DNA bait according to the protocol reported by Maricic et al. [81]. Target enrichment was carried out in solution according to the protocol of Horn [5]. Hybridisation was performed using the Oligo aCGH/ChiP-on-Chip hybridisation kit (Agilent Technologies, Palo Alto, CA, USA). Each reaction (50 µL) consisted of 12–15 µL pooled libraries (up to five libraries), 25 µL of 2× hybridisation buffer, 5 µL of blocking agent, 4 µL of blocking oligos (25 µM each) and 1–3 µL of DNA bait. The quantities of the pooled libraries and the DNA bait were adjusted so that the library–to–bait ratio was 10 to 1. Hybridisation was carried out for 20–24 h at 65 °C in a thermocycler. After incubation, the hybridisation reaction was incubated for 20 min with 5 µL of streptavidin coated beads (Dynabeads MyOne C1, Thermo Fisher, Waltham, MA, USA) to immobilise the enriched libraries. The beads-immobilised libraries were washed five times with BWT buffer (see Horn [20] for buffer composition), incubated for 2 min at 50 °C with HWE buffer, washed once with BWT buffer, transferred to a new tube, washed once with TET and resuspended in 35 µL of TE buffer. To separate the enriched library from the bait, the mixture was incubated for 5 min at 95 °C, the beads were collected on a magnet and the eluate was transferred to a new tube. The enriched library was amplified 15 cycles in three separate reactions of 20 µL each using Herculase II Fusion DNA polymerase (Agilent Technologies), which was purified using magnetic beads and subjected to the second round of hybridisation and amplification. Multiple enriched library pools were combined for sequencing ensuring that all P5 and P7 indices were unique in the pool, quantified using qPCR (Illumina Library Quantification kit, KAPA) and sequenced on the NextSeq550 at the CeNT UW NGS Core Facility using the 150 bp Mid Output kit and a 2 × 75 bp sequencing scheme. A custom Read1 primer was used for the single-stranded libraries, as described by Gansauge et al. [19].

1.4. Sequence processing

The raw reads were demultiplexed using bcl2fastq v. 2.19 (Illumina). Overlapping reads were collapsed and adaptor and quality trimmed using AdapterRemoval v. 2.2.2 [22] with the following parameters: --collapse, --minalignmentlength 4, --trims, --trimqualities, --gzip, --basename 'sample'. We indicated the sequence of the second adapter the single-stranded libraries as --adapter2 GGAAGAGCGTCGTG-TAGGGAAAGAGTGT. The reads were mapped to a range of vole mtDNA references using the BWA-MEM algorithm [23]. We used competitive mapping approach to filter out human contaminations [24]. In this approach human and vole mtDNA sequences were combined into a single reference with two chromosomes.

Duplicated, short (<30 bp) and low mapping quality reads (mapq <30) were removed by *samtools* v.1.9 [82 using the *samtools view-q 30*; *samtools sort* and *samtools rmdup* commands. Consensus sequence was called using in-house script utilizing bcftools mpileup and *ivar call* commands [26]. The BED file, which is used for masking low coverage positions, was generated using the *genomecov* command from BEDtools v. 2.27 [83] and filtered to retain only positions with coverage less than 3 using the *awk* script. Read alignments and vcf files were inspected manually using Tablet v. 1.17 software [27]. In the case of specimens identified as *S. anglicus* we called entire mtDNA genome while in the case of *M. arvalis* we called only 4.2 kb fragment as in Baca et al. [29].

1.5. Phylogenetic analyses

To confirm the taxonomic identification of the studied specimens based on the number of reads mapping to various vole mitogenomes we reconstructed Maximum Likelihood phylogeny of selected Arvicolinae species. We aligned a dataset of 25 mtDNA sequences of 24 species together with four specimens from Petit Guinards which yielded enough reads to call the mtDNA sequence. For the reconstruction we used fragment of mtDNA coding region (3.2 kb) that overlapped with a shorter fragment called for one specimen. We used IQ-tree 2 [30] for phylogeny reconstruction and to select the most appropriate partitioning scheme and assign best substitution model to each partition. We used ultrafast bootstrap to assess the branch support.

Intraspecific phylogenies were reconstructed in BEAST 1.10.4 [32] using previously published mitogenomic datasets [7,32]. The strategy and parameters used to estimate the age and phylogenetic position of each studied specimen was the same as previously [7,29]. Briefly, the age of each of the studied specimens was estimated separately using a dataset of radiocarbon dated and modern specimens. Then the resulting age estimate was used as a prior on age of this specimen in the joint analysis with all the available sequences of this species. See either Baca et al [7] or Baca et al. [29] for the detailed description on narrow-headed vole and common vole, respectively.

Additional References

81. Maricic, T.; Whitten, M.; Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **2010**, *5*, 9–13, doi:10.1371/journal.pone.0014004.
82. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.R.; Durbin, R.; Subgroup. 1000 genome project data processing the sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
83. Quinlan, A.R.; Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842, doi:10.1093/bioinformatics/btq033.