2022-08-27

The complete mitochondrial genome of Talpa martinorum (Mammalia: Talpidae), a mole species endemic to Thrace: genome content and phylogenetic considerations

Demirtas, S

http://hdl.handle.net/10026.1/19548

10.1007/s10709-022-00162-w
Genetica
Springer

The complete mitochondrial genome of *Talpa martinorum* (Mammalia: Talpidae), a mole species endemic to Thrace: Genome content and phylogenetic considerations

Sadık Demirtaş[1], Mahir Budak[2], Ertan M. Korkmaz[2], Jeremy B. Searle[3], David T. Bilton[4,5], İslam Gündüz[1*]

[1]Department of Biology, Faculty of Arts and Sciences, Ondokuz Mayis University, Samsun, Turkey

[2]Department of Molecular Biology and Genetics, Faculty of Science, Sivas Cumhuriyet University, Sivas, Turkey

[3]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, 14853-2701, USA

[4]University of Plymouth, School of Biological and Marine Sciences, Plymouth PL4 8AA, Devon, UK

[5]Department of Zoology, University of Johannesburg, PO Box 524, Auckland Park, Johannesburg 2006, South Africa

ORCID IDs: 0000-0003-0859-7887 (S.D.); 0000-0001-5610-486X (M.B.); 0000-0003-0699-1354 (E.M.K.); 0000-0001-7710-5204 (J.B.S.); 0000-0003-1136-0848 (D.T.B.); 0000-0002-6436-5397 (İ.G.).

*Corresponding author. E-mail: gunduzi@omu.edu.tr

Abstract:

The complete mitogenome sequence of *Talpa martinorum*, a recently described Balkan endemic mole, was assembled from next generation sequence data. The mitogenome is similar to that of the three other *Talpa* species sequenced to date, being 16,835 bp in length, and containing 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, an origin of L-strand replication, and a control region or D-loop. Compared to other *Talpa* mitogenomes sequenced to date, that of *T. martinorum* differs in the length of D-loop and stop codon usage. TAG and T-- are the stop codons for the ND1 and ATP8 genes, respectively, in *T. martinorum*, whilst ATG acts as a stop codon for both ND1 and ATP8 in the other three *Talpa* species sequenced. Phylogeny reconstructions based on Maximum Likelihood and Bayesian inference analyses yielded phylogenies with similar topologies, demonstrating that

T. martinorum nests within the western lineage of the genus, being closely related to *T. aquitania* and *T. occidentalis*.

Keywords: *Talpa martinorum*; mitogenome; phylogenetic

Introduction

The subterranean mole genus *Talpa* Linnaeus, 1758 is endemic to the western Palaearctic region, distributed from the Iberian Peninsula to China and Siberia (Hutterer 2005). Only nine species were considered valid in the most recent version of *Mammal Species of the World* (Hutterer 2005): the common mole *T. europaea* Linnaeus, 1758, the blind mole *T. caeca* Savi, 1822, the Roman mole *T. romana* Thomas, 1902, the Levant mole *T. levantis* Thomas, 1906, the Iberian blind mole *T. occidentalis* Cabrera, 1907, the Balkan mole *T. stankovici* Martino and Martino, 1931, the Siberian mole *T. altaica,* Nikolasky, 1883, Père David's mole *T. davidiana* Milne-Edwards, 1884, and the Caucasian mole *T. caucasica* Satunin, 1908. Recent molecular studies, however, have indicated a higher species-level diversity within the group, suggesting that some genetically divergent lineages qualify as cryptic species, which are not readily identified based on morphological characters (Bannikova et al. 2015; Demirtaş et al. 2020). Using a combination of molecular genetics techniques and morphometrics, two new mole species, *T. aquitania* Nicolas et al., 2017 from southern France and northern Spain, and *T. martinorum* Kryštufek et al., 2018 from the south-western Black Sea coast (Thrace), have been described in recent years (Nicolas et al. 2017; Kryštufek et al. 2018). In addition, Bannikova et al. (2015) recently separated two additional, genetically well-defined, lineages in the Caucasus and Anatolia, corresponding to *T. talyschensis* Vereschagin, 1945 and *T. ognevi* Stroganov, 1948. Finally, Demirtaş et al. (2020) have demonstrated that *T. levantis* s.l. in Anatolia is divisible into divergent eastern and western sublineages on both mitochondrial and nuclear markers, and on this basis argued that the eastern sublineage should be considered as a separate species (*T. transcaucasica* Dahl, 1945). As a result of these findings, the number of recognized species in the genus *Talpa* has increased from nine (Hutterer 2005) to 14 (Bannikova et al. 2015; Kryštufek and Motokawa 2018; Kryštufek et al. 2018; Demirtaş et al. 2020).

T. martinorum was originally believed to be restricted to the Thrace region of Bulgaria, along the south-western Black Sea coast. More recently, Kefelioğlu et al. (2020) have demonstrated that *T. martinorum* also occurs in nearby European Turkey. Almost nothing is known about biology of this recently described species, which appears to be restricted to a small area of the southeastern Balkans.

To date, the complete mitogenomes of three species of the genus *Talpa* (*T. aquitania*, *T. europaea* and *T. occidentalis*) have been sequenced (Mouchaty et al. 2000; Gutiérrez et al. 2018; Aleix-Mata et al. 2020), along with those of 11 other species of the Talpidae. Complete

mitochondrial genomes have become much more accessible with the advent of next-generation sequencing (NGS) (Ye et al. 2014), and are very useful for understanding genetic variability at both intra- and interspecific levels, as well as for phylogenetic and phylogeographical reconstruction across a wide range of organisms and taxonomic levels (e.g. Anijalg et al. 2018; Laurimäe et al. 2018; Ding et al. 2019; Nie et al. 2020; Nicolas et al. 2020). In this study, we report the sequencing and characterization (by NGS) of the complete mitogenome of the *Talpa* species *T. martinorum*, and provide additional insights into its evolutionary relationships with other *Talpa* species for which fully described mitogenomes are currently available.

Materials and methods

Specimen Collection and DNA Extraction

A male *T. martinorum* (Kryštufek et al. 2018) was captured at Kağıthane (41° 07′ N 28° 57′ E; Istanbul, Turkey). All capture and sacrifice protocols were approved by the Animal Experiments Local Ethics Committee at Ondokuz Mayis University (code: 2019/28). Total DNA was extracted from muscle tissue using phenol-chloroform (Köchl et al. 2005). The quality of extracted DNA was detected by 1.5% agarose electrophoresis and the DNA was stored at −20 °C until further use.

Preparation of Libraries, Sequencing, Mitogenome Assembly and Gene Annotation

Illumina libraries were generated from total DNA with an Illumina Nextera XT DNA Sample Prep Kit (Illumina, San Diego, CA). DNA quality was assessed with Qubit, final library length distribution and checking for the absence of adapters, performed using Qsep100 (Bioptic, New Taipei City, Taiwan). Normalized and pooled DNA libraries were subjected to *de novo* genome sequencing on an Illumina MiSeq System, using 300-cycle MiSeq Reagent Micro Kit v2 at CUTAM (**http://cutam.cumhuriyet.edu.tr/**). Demultiplexing and adapter trimming were carried out using miseq reporter v2.3.32 (Illumina). FastQC (Andrews 2010) was used for quality control checks on raw sequence data. Raw reads were assembled to a reference complete mitochondrial genome of *Talpa europaea* (NCBI accession Y19192) using Bowtie2 v2.3.5.1 in -very-sensitive mode, equivalent to options -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 (Langmead and Salzberg 2012). The mapped SAM file was processed with Samtools v1.10 (Li et al. 2009) to create a sorted BAM file. The consensus sequence from the .bam file was extracted using vcfutils.pl perl script. Mean coverage was calculated from the bam file using Rsamtools v2.2.3 (Morgan et al. 2020). The resulting *T. martinorum* mitogenome

consensus sequence was annotated using MITOS (http://mitos .bioin f.uni-leipz ig.de) (Bernt et al. 2013) with default settings. Gene boundaries were also checked by alignment against mitogenome sequences of *T. aquitania* (NCBI accession MN443911), *T. europaea* (NCBI accession Y19192) and *T. occidentalis* (NCBI accession NC_039630). Mitochondrial genomes were aligned using MAFFT v7.453 (Katoh and Standley 2013) with --localpair and—maxiterate 1000 options. Translations and codon usage statistics for 13 protein-coding genes were conducted with Geneious Prime 2019.1 (Biomatters Ltd., Auckland, New Zealand). The circularized image of the mitogenome was made using OrganellarGenomeDRAW tools (http://ogdra w.mpimp -golm.mpg.de/) (Lohse et al. 2013). Skewness of nucleotide composition was gauged according to the following formulae: AT skew [(A − T)/(A + T)] and GC skew [(G − C)/(G + C)] (Perna and Kocher 1995). Base composition and skew of the complete mitochondrial genome were calculated using MEGAX v10 (Kumar et al. 2018).

Phylogenetic Analyses

For phylogenetic analyses we used the concatenated sequences of 13 protein-coding genes (PCGs) from other members of the Talpidae available in GenBank (*Condylura cristata* KU144678, *Galemys pyrenaicus* AY833419, *Mogera robusta* KT934322 and MK431828, *Mogera wogura* AB099482, *Parascaptor leucura* MW114662, *Scapanulus oweni* KM506754, *Talpa aquitania* MN443911, *Talpa occidentalis* NC_039630, *Talpa europaea* Y19192, *Uropsilus andersoni* JX945573 and NC_041144, *Uropsilus gracilis* KM379136, *Uropsilus investigator* JX945574, *Uropsilus soricipes* JQ658979 and *Urotrichus talpoides* AB099483). PCG sequences of four species of the family Soricidae (*Crocidura russula* AY769263, *Sorex araneus* KT210896, *Suncus murinus* NC_024604 and *Blarina brevicauda* NC_042734) were used as outgroups. The phylogenetic relationships amongst taxa were reconstructed using the maximum-likelihood (ML) algorithm implemented in PAUP v4.10b (Swofford 2002) and Bayesian inference of phylogeny (BI), as implemented in MRBAYES v3.2.7a (Ronquist et al. 2012). The Akaike information criterion (AIC) implemented in jMODELTEST v1.0 (Posada 2008) was used to establish the optimal model of sequence evolution for our data and this model was subsequently employed in the ML and BI analyses. The ML tree search was conducted using the heuristic search approach, the ‘as is’ addition replicate and node supports were assessed with 1000 bootstrap (BS) replicates. BI analysis involved four Markov chains of one million generations each, with trees being sampled every 100 generations and a burn-in of 25%. The software tool TRACER v1.7.1 (Rambaut et al.

2018) was used to check parameters and to determine the number of trees needed to reach stationarity (burn-in). After discarding burn-in trees and evaluating convergence, remaining samples were retained in order to generate 50% majority rule consensus trees and calculate posterior probabilities (PB). Previous phylogenetic analyses on multiple organisms have suggested that incongruence, the presence of topological conflict, might exist between different tree building approaches (Hess and Goldman, 2011; Song et al. 2012; Steenwyk et al. 2019). Thus, to further evaluate the topological congruence of the ML and BI trees, two main tree topology tests were computed using IQ-TREE web server (Trifinopoulos et al. 2016). We first combined Newick formatted trees (ML and BI) into a single file, and the resulting file was then used as input to IQ-TREE. We used the "GTR+F+I+G4" model and conducted the Shimodaira-Hasegawa (SH; Shimodaira and Hasegawa, 1999) and the approximately unbiased (AU; Shimodaira, 2002) tests. These tests were conducted using 10,000 resamplings using the resampling estimated log-likelihood (RELL) method (Kishino et al. 1990) to evaluate congruence at *p*-values <0.05.

Results and Discussion

The Sequence of Genes

The complete mitochondrial genome of *T. martinorum* is 16,835 bp in length (GenBank: OP082230), shorter than those of *T. europaea* (16,884 bp) and *T. occidentalis* (16,962 bp) and slightly longer than that of *T. aquitania* (16,826 bp) (Mouchaty et al. 2000; Gutiérrez et al. 2018; Aleix-Mata et al. 2020). However, the order and orientation of the *T. martinorum* mitogenome are identical to those of other *Talpa* species and consists of the conserved set of 37 mammal mitochondrial genes, including 13 protein coding genes (PCGs) (*CYTB*, *ND1–6*, *ND4L*, *COX1–3*, *ATP6* and *ATP8*), 22 tRNAs (two for Leu and Ser and one for each of the other amino acids), two rRNAs (*12S* and *16S*), the control region (D-loop) and the origin of the light-strand region (O_L). The PCG region is 11,412 bp long, and 12 of the 13 PCGs are encoded on the heavy (H) strand (*CYTB*, *ND1–5*, *ND4L*, *COX1–3*, *ATP6* and *ATP8*), with the remaining PCG (*ND6*) encoded on the light (L) strand. Eight tRNAs are found on the L-strand while the other 14 tRNAs and the two rRNAs are located on the H-strand. The D-loop is 1375 bp long, located between *tRNA-Pro* and *tRNA-Phe*, as seen in the mitogenomes of other species of *Talpa* (Mouchaty et al. 2000; Gutiérrez et al. 2018; Aleix-Mata et al. 2020) (Table 1; Fig. 1).

.

Table 1 Gene organization of the *Talpa martinorum* mitochondrial genome. H: Heavy strand; L: Light strand.

Gene	**Start position**	**Stop position**	**Length (bp)**	**Intergenic nucleotides (bp)**	**Anticodon**	**Start codon**	**Stop codon**	**Strand**
tRNA-Phe	1	70	70	2	GAA			H
12S rRNA	73	1041	969	0				H
tRNA-Val	1042	1109	68	0	TAC			H
16S rRNA	1110	2681	1572	0				H
tRNA-Leu(UUR)	2682	2756	75	2	TAA			H
ND1	2759	3712	954	2		ATG	TAG	H
tRNA-Ile	3715	3783	69	−3	GAT			H
tRNA-Gln	3781	3853	73	1	TTG			L
tRNA-Met	3855	3923	69	0	CAT			H
ND2	3924	4967	1044	−2		ATA	TAG	H
tRNA-Trp	4966	5033	68	5	TCA			H
tRNA-Ala	5039	5107	69	1	TGC			L
tRNA-Asn	5109	5181	73	0	GTT			L
O_L	5182	5220	39	−3				H
tRNA-Cys	5218	5284	67	0	GCA			L
tRNA-Tyr	5285	5351	67	1	GTA			L

COX1	5353	6897	1545	1		ATG	TAA	H
tRNA-Ser(UCN)	6899	6967	69	7	TGA			L
tRNA-Asp	6975	7043	69	0	GTC			H
COX2	7044	7727	684	3		ATG	TAA	H
tRNA-Lys	7731	7798	68	1	TTT			H
ATP8	7800	8001	202	−41		ATG	T--	H
ATP6	7961	8641	681	−1		ATG	TAA	H
COX3	8641	9424	784	0		ATG	T--	H
tRNA-Gly	9425	9494	70	0	TCC			H
ND3	9495	9840	346	0		ATT	T--	H
tRNA-Arg	9841	9908	68	0	TCG			H
ND4L	9909	10205	297	−7		ATG	TAA	H
ND4	10199	11576	1378	0		ATG	T--	H
tRNA-His	11577	11644	68	0	GTG			H
tRNA-Ser(AGY)	11645	11705	61	2	GCT			H
tRNA-Leu(CUN)	11708	11777	70	0	TAG			H
ND5	11778	13598	1821	− 18		ATT	TAA	H
ND6	13582	14109	528	0		ATG	TAA	L
tRNA-Glu	14110	14178	69	4	TTC			L

CYTB	14183	15322	1140	0		ATG	AGA	H
tRNA-Thr	15323	15391	69	1	TGT			H
tRNA-Pro	15393	15460	68	0	TGG			L
D-loop	15461	16835	1375	0				H

Nucleotide Composition, Degree of Overlap, Intergenic Spacer Regions and Skewness

The mitogenome of *T. martinorum* is AT-biased, with a nucleotide composition of 34.02% A, 24.54% C, 14.32% G, and 27.12% T. The 13 mitochondrial PCGs consist of 33.00% A, 25.40% C, 13.40% G and 28.20% T. The 13 PCGs are AT-biased, with a total AT content of 61.20%, ranging from 57.50% in COX3 to 70.10% in ATP8. Overall, the AT skew (0.11) for the *T. martinorum* mitogenome is positive, reflecting a higher occurrence of As than Ts, and the GC skew (–0.26) is appreciably negative, indicating a higher content of Cs compared to Gs (Supplementary Materials Table S1). There were 7 overlapping regions with a total length of 75 bp (ranging from 1 to 41 bp, with the longest overlapping region located between ATP8 and ATP6 genes) and 14 intergenic spacers with a total length of 33 bp (ranging from 1 to 7 bp) (Table 1).

Protein-Coding Genes and Codon Usage

The 13 mitochondrial PCGs in *T. martinorum* are 11,404 bp in length (11,373 bp in codons and 31 bp in stop codons) and encode 3791 amino acids (Table 1; Supplementary Materials Table S1). There are three start codons used in the *T. martinorum* mtDNA: ATA for *ND2*, ATT for *ND3* and *ND5*, and the most frequent one, ATG, for *ND1*, *COX1*, *COX2*, *ATP8*, *ATP6*, *COX3*, *ND4L*, *ND4*, *ND6*, *CYTB*. Nine genes end on a complete stop codon: TAG (*ND1*, *ND2)*, TAA (*COX1*, *COX2***,** *ATP6*, *ND4L*, *ND5*, *ND6)* and AGA (*CYTB*). The remaining four genes (*ATP8*, *COX3*, *ND3*, *ND4*) end on an abbreviated stop codon (T--)**.** An incomplete stop codon is commonly found in metazoan mitogenomes, and is presumably completed via poly-adenylation of the 3′-end of the mRNA after transcription (Ojala et al. 1981). Accordingly, the most abundant start and stop codons were ATG (76.92%) and TAA (46.15%), respectively, as in other mammal mitogenomes, including mole species (Mouchaty et al. 2000; Cabria et al. 2006; Chen et al. 2015; Kim and Park 2015; Xu et al. 2016; Kim et al. 2017; Gutiérrez et al. 2018; Aleix-Mata et al. 2020; Lamelas et al. 2020). Due to the A+T richness of the mitogenome of *T. martinorum* (Supplementary Materials Table S1), a strong bias toward A+T-rich codons was observed in the PCGs. The AT skew was positive in all but one PCG (*ND4L*), whilst the GC skew was negative in all 13 PCGs (Supplementary Materials Table S1 and Fig. S1). A high proportion of A+T in PCGs is typical in mammalian mitogenomes (Kim et al. 2017; Gutiérrez et al. 2018). As reported by Chen et al. (2014) and Labella et al. (2019), codon usage bias in mitochondrial genomes may be caused by

mutational bias and/or natural selection. In our study, the codon distribution chart (Supplementary Materials Fig. S2) revealed that four amino acids (Leucine 2, 449; Isoleucine, 315; Threonine, 305; and Alanine, 260) were the most common in the mitochondrial PCGs of *T. martinorum*. The five codons with the highest relative synonymous codon usage (RSCU) values described in the PCGs from *T. martinorum* are as follows: CTA (Leucine 2) (2.84), CGA (Arginine) (2.77), TCA (Serine 2) (2.73), CCA (Proline) (2.70), ACA (Threonine) (2.23), GTA (Valine) (2.18), GGA (Glycine) (1.87) and TGA (Tryptophan) (1.81) (see Supplementary Materials Fig. S2).

Transfer RNA and Ribosomal RNA genes, D-loop Sequences and Origin of Replication for the Light Strand (O_L)

The combined size of the 22 tRNA genes (two for Leu and Ser and one for each of the other amino acids) was 1516 bp, varying in size from 61 bp (*tRNA-Ser(AGY)*) to 75 bp (*tRNA-Leu(UUR)*). A total of 14 tRNAs are encoded by the H-strand, and the remaining eight tRNAs are encoded by the L-strand (Table 1). The AT skew of the tRNAs was positive (Supplementary Materials Table S1). All tRNA genes showed a typical clover-leaf structure (Giegé et al. 1993) with the exception of *tRNA-Ser* (GCT) (Supplementary Materials Fig. S3). The *tRNA-Ser* (GCT) gene cannot be folded into typical clover-leaf secondary structures due to the lack of the dihydrouracil (DHU) stem and loop, as in some other mammals (Gissi et al. 1998; Jiang et al. 2012; Ding et al. 2016).

The 12S and 16S rRNAs were 969 bp and 1572 bp in length, respectively. The two rRNA genes are located between the *tRNA-Phe* and *tRNA-Leu(UUR)* genes, and are separated by the *tRNA-Val* gene (Table 1; Fig. 1). The base composition of the two combined rRNA genes was 37.30% A, 20.20% C, %18.20 G and %24.30 T. The AT skew (0.21) for the two combined rRNA genes was appreciably positive, reflecting a higher occurrence of As to Ts, and its GC skew (–0.05) is negative, indicating a slight excess of C over G nucleotides (see Supplementary Materials Table S1).

The most variable region in vertebrate mtDNA, the noncoding control region (or D-loop), was 1375 bp long in *T. martinorum*, located between the *tRNA-Pro* and *tRNA-Phe* genes (Table 1; Fig. 1), as in the mitogenomes of the other three mole species (Mouchaty et al. 2000; Gutiérrez et al. 2018; Aleix-Mata et al. 2020). The nucleotide composition of the D-loop was 35.70% A, 29.40% C, 12.70% G and 22.20% T. This composition is in line with most of the mammals, except Primates, for which A + T > G + C in all the domains of the D-loop region (Sbisà et al. 1997). The D-loop AT skew was

positive, whilst the GC skew was negative (Supplementary Materials Table S1). The origin of light strand synthesis (O_L) in the mtDNA of *T. martinorum* was 39 bp long and located between *tRNA-Asn* and *tRNA-Cys* in the WANCY region, which consists of a cluster of five tRNA genes (*tRNA-Trp*, *tRNA-Ala*, *tRNA-Asn*, *tRNA-Cys*, and *tRNA-Tyr*) (Table 1; Fig. 1), as in other mammals (Kim et al. 2017; Gutiérrez et al. 2018; Aleix-Mata et al. 2020). The stem-loop structure of O_L of *T. martinorum* begins with the conserved motif 5′-CTTCT-3′.

The mitogenome sequence of *T. martinorum* has a similar genome organization and structure as those of other *Talpa* species, but there are differences in the length of D-loop and stop codon usage. *T. occidentalis* has the longest mole D-loop (1504 bp) reported to date, *T. aquitania* having the shortest (1370 bp). TAG and T-- are the stop codons for the *ND1* and *ATP8* genes in *T. martinorum*, respectively. In contrast, ATG is the typical stop codon for both the *ND1* and *ATP8* genes in the other three *Talpa* species sequenced to date (Supplementary Materials Table S2).

Phylogenetic analysis

The best-fit DNA substitution model selected by jMODELTEST under AIC was GTR, with gamma correction (G) of 0.9660 and a proportion of invariable sites (I) of 0.4850; this was then used in phylogenetic analyses. ML and BI analyses yielded similar topologies (*P*-value > 0.05 for both SH and AU tests), differing mainly in relative bootstrap/posterior probability values for some nodes and the position of *Galemys pyrenaicus* on the trees (Fig. 2). In both the ML and BI trees, *Talpa* is located in the clade together with *Mogera* and *Parascaptor leucura* (BS = 100 and PP = 1), and a close association between this group and a smaller one consisting of *Condylura cristata*, *Scapanulus oweni* and *Urotrichus talpoides* was consistently recovered but was not sufficiently supported by bootstrap percentages (PP = 0.90). In the BI tree *Galemys pyrenaicus* grouped with *C. cristata*, *S. oweni* and *U. talpoides*, whilst in the ML tree it was located in a basal position on another early diverging branch that separated from these two groups, as previously reported (Cabria et al. 2006; Tu et al. 2012; Tu et al. 2015; Gutiérrez et al. 2018; Aleix-Mata et al. 2020). Consistent with previous phylogenetic analyses of Talpidae (Shinohara et al. 2003; Tu et al. 2012; Gutiérrez et al. 2018; Aleix-Mata et al. 2020), the inclusion of the four species of *Uropsilus* (subfamily Uropsilinae) in one early diverging well-supported clade was confirmed by both the ML and BI methods (BS = 100 and PP = 1). Relationships within *Talpa* are interesting. Based on partial or complete *cyt b* sequences, it has already been shown that the recently-described *T.*

martinorum belongs to the western group, comprising the common mole *T. europaea*, the blind mole *T. caeca*, the Roman mole *T. romana*, the Levant mole *T. levantis*, the Iberian blind mole *T. occidentalis* and the Balkan mole *T. stankovici* (Kryštufek et al. 2018; Demirtaş et al. 2020; Kefelioğlu et al. 2020). Previous studies also suggested that *T. martinorum* was clustered with *T. aquitania*, *T. europaea* and *T. occidentalis* in the same clade within the western group, but the branching order was not resolved using *cyt b* sequences alone. Based on the 13 mitochondrial PCGs, our ML and BI trees resolved the relationships as (((*T. occidentalis*, *T. aquitania*), *T. martinorum*), *T. europaea*) with strong support values (BS = 100 and PP = 1), our study also suggesting that *T. europaea* forms a basal branch within the western *Talpa* clade (Fig. 2). Sequencing and characterization of the mitogenomes of further species of the genus *Talpa* and the use of nuclear markers will help our understanding the phylogenetic relationships within this genus in the future. This is important because, as with many other genera of small mammals, moles exhibit considerable morphological conservatism, so that molecular data can improve our understanding of their geographical and evolutionary diversity.

Conclusions

The whole mitochondrial genome of *T. martinorum*, a recently described endemic Balkan mole, is sequenced and characterized. The complete mitogenome of *T. martinorum* has a genomic organization and structure similar to those described for other mammal species. It is 16,835 bp in length, consisting of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, and a displacement loop region. It is comparable in size to those from other species of the genus *Talpa* such as *T. aquitania* (16,776-16,826 bp), *T. europaea* (16,884 bp) and *T. occidentalis* (16,962 bp). The mitogenomes of *Talpa* species have highly conserved gene order, and similar to other vertebrate mitogenomes, all of the PCGs in the mitogenome of *T. martinorum* utilize ATN as a start codon, ATG being the most frequent. Twelve PCGs, 14 tRNAs and two rRNAs are located on the heavy strand, while ND6 and eight tRNAs are found on the light strand. All of the tRNAs can be folded into typical clover-leaf secondary structures, with the exception of *tRNA-Ser* (GCT), which lacks a DHU stem and loop. The D-loop is placed between the *tRNA-Pro* and *tRNA-Phe* genes, and the L-strand origin of replication (O_L) is located between *tRNA-Asn* and *tRNA-Cys* in the WANCY region. In this study, phylogenetic reconstructions with other members of the Talpidae based on 12 PCGs using BI and ML methods, resolved phylogenetic relationships in the genus *Talpa*, with

T. martinorum clustering as a monophyletic group with *T. occidentalis*, *T. aquitania*, *T. martinorum*, and *T. europaea*.

Author Contributions: Conceptualization, methodology, validation, formal analysis, investigation, data curation, İ.G, S.D, M.B and M.K; writing—original draft preparation, writing—review and editing, İ.G and D.T.B; project administration, İ.G and S.D; review and editing, J.B.S. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: Permission for specimen capture was kindly granted by the General Directorate of Nature Conservation and National Parks, Republic of Turkey Ministry of Agriculture and Forestry.

Declarations

Conflicts of Interest: The authors declare that they have no conflicts of interest to this work.

References

Aleix-Mata G, Gutiérrez J, Ruiz-Ruano FJ, Lorite P, Marchal JA, Sánchez A (2020) The complete mitochondrial genome of *Talpa aquitania* (Talpidae; Insectivora), a mole species endemic to northern Spain and southern France. Mol Biol Rep 47:2397–2403. https://doi.org/10.1007/s11033-020-05296-8

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Anijalg P, Ho SYW, Davison J, Keis M, Tammeleht E, Bobowik K, Tumanov IL, Saveljev AP, Lyapunova EA, Vorobiev AA, Markov NI, Kryukov AP, Kojola I, Swenson JE, Hagen SB, Eiken HG, Paule L, Saarma U (2018) Large-scale migrations of brown bears in Eurasia and to North America during the Late Pleistocene. J Biogeogr 45:394–405. https://doi.org/10.1111/jbi.13126

Bannikova AA, Zemlemerova ED, Colangelo P, Sözen M, Sevindik M, Kidov AA, Dzuev RI, Kryštufek B, Lebedev VS (2015) An underground burst of diversity—a new look at the phylogeny and taxonomy of the genus *Talpa* Linnaeus, 1758 (Mammalia: Talpidae) as revealed by nuclear and mitochondrial genes. Zool J Linn Soc 175:930–948. https://doi.org/10.1111/zoj.12298

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Cabria MT, Rubines J, Gómez-Moliner B, Zardoya R (2006) On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*). Gene 375:1–13. https://doi.org/10.1016/j.gene.2006.01.038

Chen H, Sun S, Norenburg JL, Sundberg P (2014) Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). PLoS One 9(1):e85631. https://doi.org/10.1371/journal.pone.0085631

Chen S, Tu F, Zhang X, Li W, Chen G, Zong H, Wang Q (2015) The complete mitogenome of stripe-backed shrew, *Sorex cylindricauda* (Soricidae). Mitochondrial DNA 26:477–478. https://doi.org/10.3109/19401736.2013.855756

Demirtaş S, Silsüpür M, Searle JB, Bilton D, Gündüz İ (2020) What should we call the Levant mole? Unravelling the systematics and demography of *Talpa levantis* Thomas, 1906

sensu lato (Mammalia: Talpidae). Mamm Biol 100:1–18. https://doi.org/10.1007/s42991-020-00010-4

Ding L, Li W, Liao J (2016) Mitochondrial genome of *Cricetulus migratorius* (Rodentia: Cricetidae): insights into the characteristics of the mitochondrial genome and the phylogenetic relationships of *Cricetulus* species. Gene 595:121–129. https://doi.org/10.1016/j.gene.2016.10.003

Ding S, Li W, Wang Y, Cameron SL, Muranyi D, Yang D (2019) The phylogeny and evolutionary timescale of stoneflies (Insecta: Plecoptera) inferred from mitochondrial genomes. Mol Phylogenet Evol 135:123–135. https://doi.org/10.1016/j.ympev.2019.03.005

Giegé R, Puglisi JD, Florentz C (1993) tRNA structure and aminoacylation efficiency. Prog Nucleic Acid Res Mol Biol 45:129–206. https://doi.org/10.1016/s0079-6603(08)60869-7

Gissi C, Gullberg A, Arnason U (1998) The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculus*. Genomics 50:161–169. https://doi.org/10.1006/geno.1998.5282

Gutiérrez J, Lamelas L, Aleix-Mata G, Arroyo M, Marchal JA, Palomeque T, Lorite P, Sánchez A (2018) Complete mitochondrial genome of the Iberian mole *Talpa occidentalis* (Talpidae, Insectivora) and comparison with *Talpa europaea*. Genetica 146:415–423. https://doi.org/10.1007/s10709-018-0033-z

Hess J, Goldman N (2011) Addressing inter-gene heterogeneity in maximum likelihood phylogenomic analysis: yeasts revisited. PLoS One 6:e22783. https://doi.org/10.1371/journal.pone.0022783

Hutterer R (2005) Order Soricomorpha. In: Wilson DE, Reeder DM (ed) Mammal species of the world, Vol. 1, 3rd edn, Johns Hopkins University Press, Baltimore, pp 220–311.

Jiang X, Gao J, Ni L, Hu J, Li K, Sun F, Xie J, Bo X, Gao C, Xiao J, Zhou Y (2012) The complete mitochondrial genome of *Microtus fortis calamorum* (Arvicolinae, Rodentia) and its phylogenetic analysis. Gene 498:288–295. https://doi.org/10.1016/j.gene.2012.02.022

Katoh K, Standley, DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. https://doi.org/10.1093/molbev/mst010

Kefelioğlu H, Kryštufek B, Selçuk AY, Hutterer R, Astrin JJ (2020) Taxonomic revision of the Levant moles of Turkey (Mammalia: Talpidae). Bonn Zool Bull 69:275–291. https://doi.org/10.20363/BZB-2020.69.2.275

Kim JY, Park YC (2015) Gene organization and characterization of the complete mitogenome of *Hypsugo alaschanicus* (Chiroptera: Vespertilionidae). Genet Mol Res 14:16325–16331. https://doi.org/10.4238/2015.December.8.24

Kim NH, Lim SJ, Chae HM, Park YC (2017) Complete mitochondrial genome of the Amur hedgehog *Erinaceus amurensis* (Erinaceidae) and higher phylogeny of the family Erinaceidae. Genet Mol Res 16(1). https://doi.org/10.4238/gmr16019300

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 31:151–160. https://doi.org/10.1007/BF02109483

Köchl S, Niederstätter H, Parson W (2005) DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. In: Carracedo A (ed), Forensic DNA Typing Protocols. Meth Mol Biol 297:13–29. https://doi.org/10.1385/1-59259-867-6:013

Kryštufek B. Motokawa M (2018) Family Talpidae (moles, desmans, star-nosed moles and shrew moles). In: Wilson DE, Lacher TA, Mittermeier RA (ed), Handbook of the Mammals of the World, Vol. 8, Insectivores, sloths and colugos, Lynx Edicions, Barcelona, pp 552–619.

Kryštufek B, Nedyalkov N, Astrin JJ, Hutterer R (2018) News from the Balkan refugium: Thrace has an endemic mole species (Mammalia: Talpidae). Bonn Zool Bull. 67:41–57.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A (2019) Variation and selection on codon usage bias across an entire subphylum. PLoS Genet 15(7):e1008304. https://doi.org/10.1371/journal.pgen.1008304

Lamelas L, Aleix-Mata G, Rovatsos M, Marchal JA, Palomeque T, Lorite P, Sánchez A (2020) Complete mitochondrial genome of three species of the genus *Microtus* (Arvicolinae, Rodentia). Animals 10:2130. https://doi.org/10.3390/ani10112130

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Laurimäe T, Kinkar L, Romig T, Omer RA, Casulli A, Umhang G, Gasser RB, Jabbar A, Sharbatkhori M, Mirhendi H, Ponce-Gordo F, Lazzarini LE, Soriano SV, Varcasia A, Rostami Nejad M, Andresiuk V, Maravilla P, González LM, Dybicz M, Gawor J, Šarkūnas

M, Šnábel V, Kuzmina T, Saarma U (2018) The benefits of analysing complete mitochondrial genomes: Deep insights into the phylogeny and population structure of *Echinococcus granulosus* sensu lato genotypes G6 and G7. Infect Genet Evol 64:85–94. https://doi.org/10.1016/j.meegid.2018.06.016

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW-A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41:W575–W581. https://doi.org/10.1093/nar/gkt289

Morgan M, Pagès H, Obenchain V, Hayden N (2020) Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.6.0, https://bioconductor.org/packages/Rsamtools

Mouchaty SK, Gullberg A, Janke A, Arnason U (2000) The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences, Mol Biol Evol 17:60–067. https://doi.org/10.1093/oxfordjournals.molbev.a026238

Nicolas V, Fabre P-H, Bryja J, Denys C, Verheyen E, Missoup AD, Olayemi A, Katuala P, Dudu A, Colyn M, Peterhans JK, Demos T (2020). The phylogeny of the African wood mice (Muridae, *Hylomyscus*) based on complete mitochondrial genomes and five nuclear genes reveals their evolutionary history and undescribed diversity. Mol Phylogenet Evol 144:106703. https://doi.org/10.1016/j.ympev.2019.106703

Nicolas V, Martínez-Vargas J, Hugot J-P (2017) *Talpa aquitania* sp. nov. (Talpidae, Soricomorpha), a new mole species from SW France and N Spain. Mammalia 81:641–642. https://doi.org/10.1515/mammalia-2017-0057

Nie R-E, Andújar C, Gómez-Rodríguez C, Bai M, Xue H-J, Tang M, Yang C-T, Tang, P, Yang X-K, Vogler AP (2020) The phylogeny of leaf beetles (Chrysomelidae) inferred from mitochondrial genomes. Syst Entomol 45:188–204. https://doi.org/10.1111/syen.12387

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474. https://doi.org/10.1038/290470a0

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358. https://doi.org/10.1007/BF00186547

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256. https://doi.org/10.1093/molbev/msn083

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125–140. https://doi.org/10.1016/S0378-1119(97)00404-6

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114. https://doi.org/10.1093/oxfordjournals.molbev.a026201

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492–508. https://doi.org/10.1080/10635150290069913

Shinohara A, Campbell KL, Suzuki H (2003) Molecular phylogenetic relationships of moles, shrew moles, and desmans from the New and Old Worlds. Mol Phylogenet Evol 27:247–258. https://doi.org/10.1016/s1055-7903(02)00416-5

Song S, Liu L, Edwards SV, Wu S (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc Natl Acad Sci USA 109:14942–14947. https://doi.org/10.1073/pnas.1211733109

Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A (2019) A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. mBio 10:e00925-19. https://doi.org/10.1128/mBio.00925-19

Swofford DL (2002) PAUP*. phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland: Sinauer Associates.

Trifinopoulos J, Nguyen T-L, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44:W232–235. https://doi.org/10.1093/nar/gkw256

Tu F, Fan Z, Chen S, Yin Y, Li P, Zhang X, Liu S, Yue B (2012) The complete mitochondrial genome sequence of the gracile shrew mole, *Uropsilus gracilis* (Soricomorpha: Talpidae). Mitochondrial DNA. 23:382–384. https://doi.org/10.3109/19401736.2012.696634

Tu F, Fan Z, Murphy RW, Chen S, Zhang X, Yan C, Liu Y, Sun Z, Fu J, Liu S, Yue B (2015) Molecular phylogenetic relationships among Asiatic shrew moles inferred from the complete mitogenomes. J Zoolog Syst Evol Res 53:155–160. https://doi.org/10.1111/jzs.12081

Xu Y, Huang X, Hu Y, Tu F (2016) Description of the mitogenome of Gansu mole (*Scapanulus oweni*). Mitochondrial DNA Part A. 27:2083–2084. https://doi.org/10.3109/19401736.2014.982567

Ye F, Samuels DC, Clark T, Guo Y (2014) High-throughput sequencing in mitochondrial DNA research. Mitochondrion 17:157–163. https://doi.org/10.1016/j.mito.2014.05.004

Figure captions

Figure 1. Circular map of the mitogenome of *Talpa martinorum*. The outside and inside of the ring indicate the heavy (H) and light (L) strands, respectively. The inner ring shows the GC content of the genome.

Figure 2. Results of BI and ML analyses combined on a ML tree based on concatenated PCG sequences of *T. martinorum* and other members of Talpidae for which mitogenomes are available. Numbers at nodes indicate bootstrap support values (ML)/posterior probabilities (BI). Bootstrap values ≥90% and Bayesian posterior probabilities ≥0.90 are shown.

Supplementary Materials

Figure S1. Graphical illustration showing the AT and GC skew in the protein-coding genes (PCGs) of *Talpa martinorum*.

Figure S2. Relative synonymous codon usage (RSCU) and codon composition counts in the mitogenome of *Talpa martinorum*.

Figure S3. Predicted secondary structures for the transfer RNA (tRNA) genes in *Talpa martinorum*. Structural features are listed on *tRNA-Val* at the bottom right. tRNAs are labelled with the abbreviations of their corresponding amino acids.

Table S1 Mitogenome nucleotide composition of *Talpa martinorum*.

Table S2 Sequence comparisons among the mitogenomes of *Talpa* species (*T. aquitania*, *T. europaea*, *T. martinorum* and *T. occidentalis*)

Figure 1

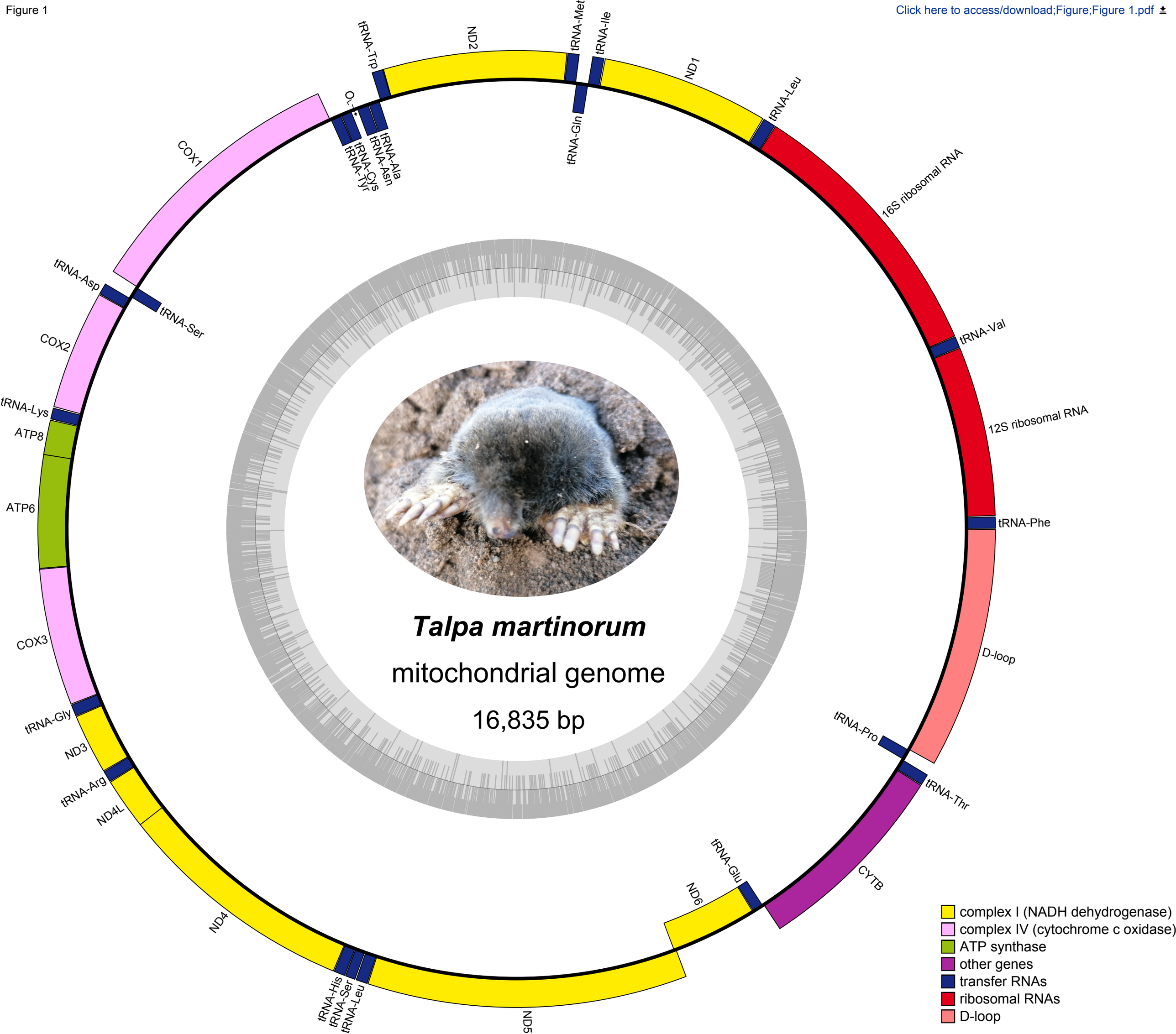
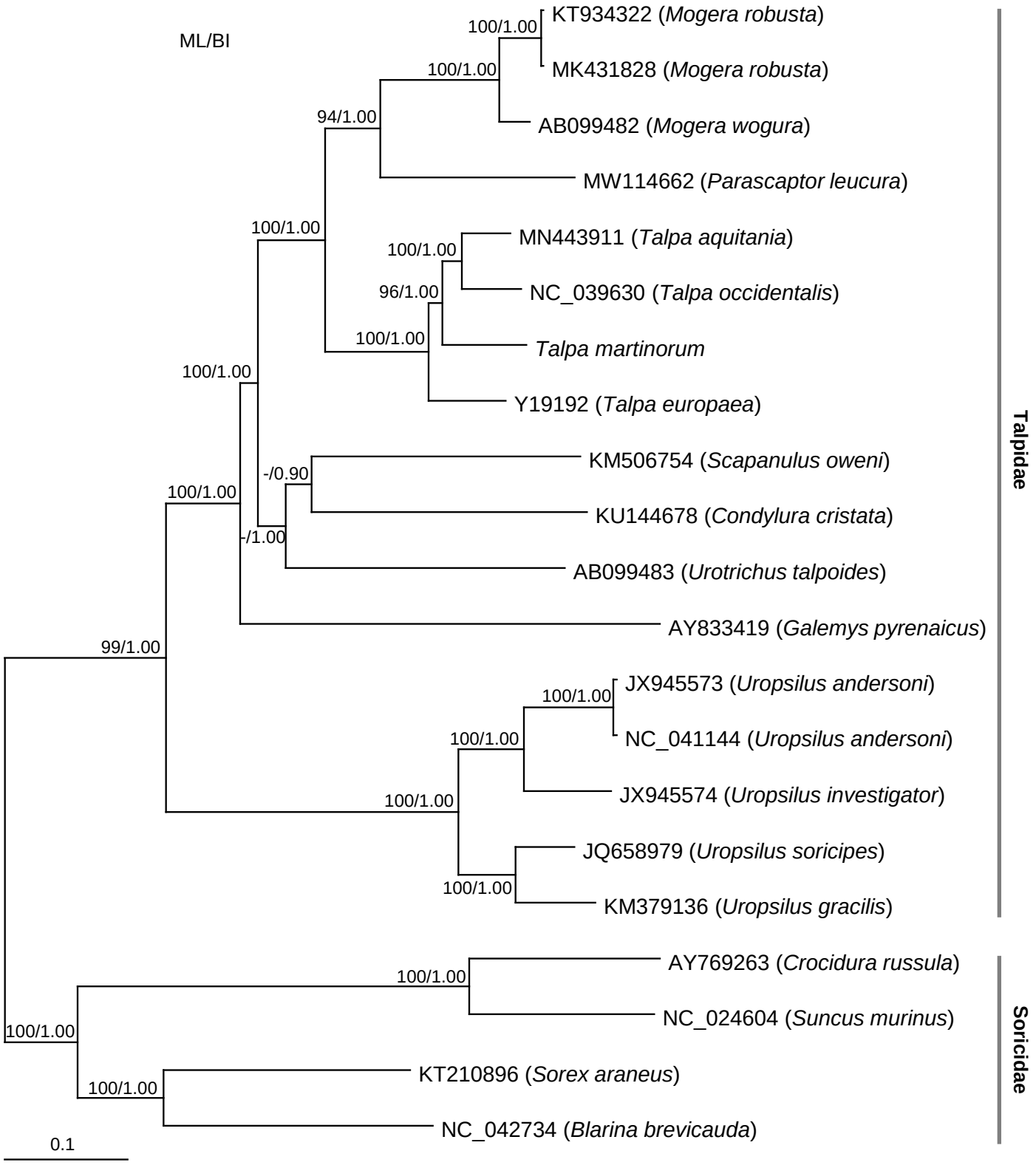
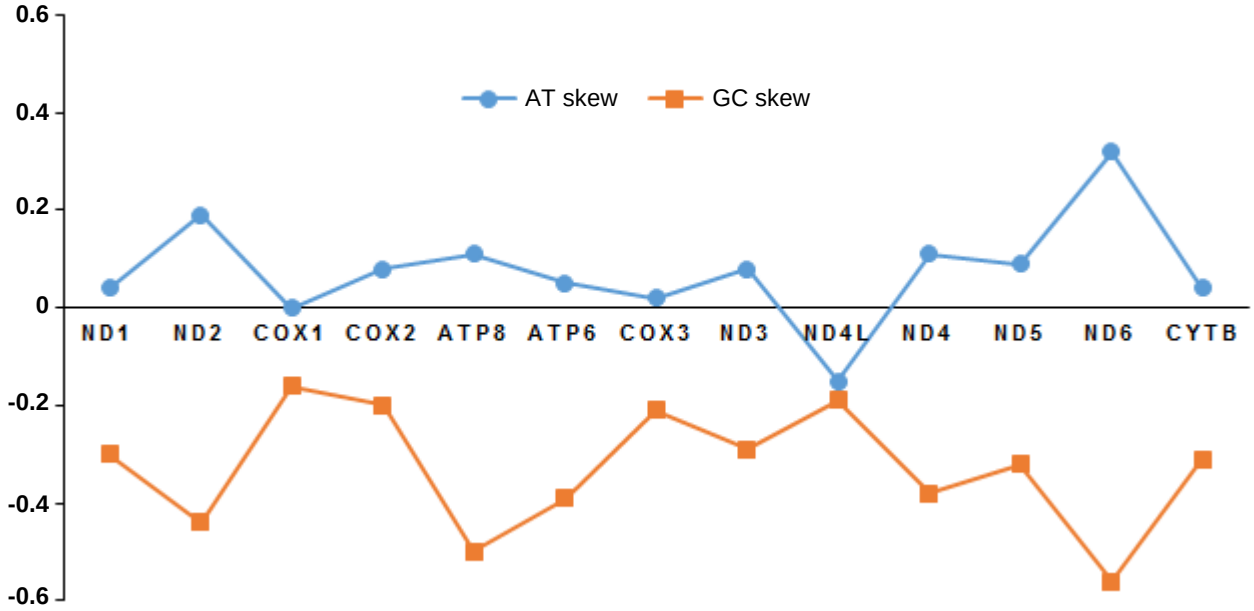
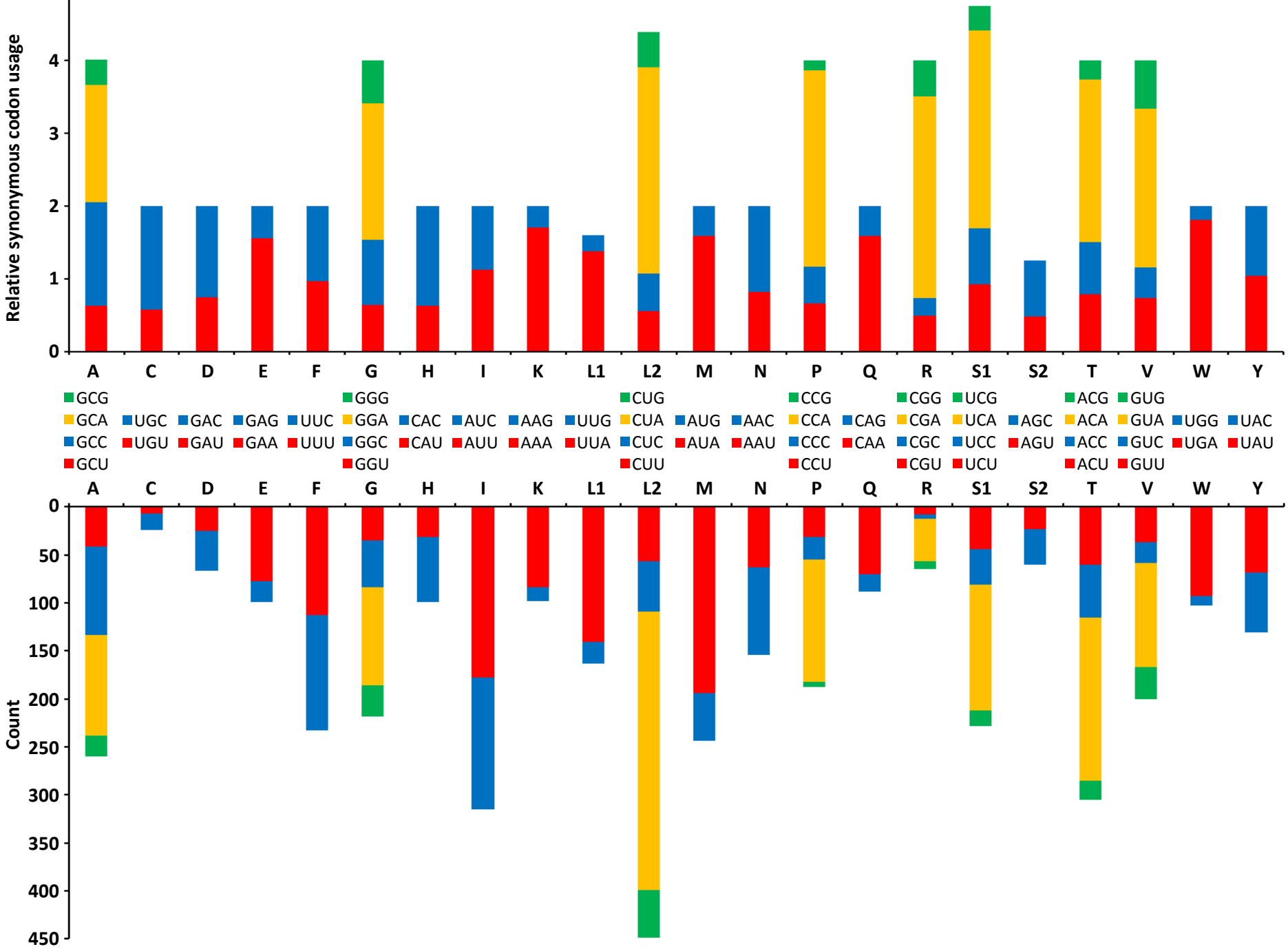


Figure 2




0.6
0.4
0.2
0
-0.2
-0.4
-0.6
AT skew
GC skew
ND1
ND2
COX1
COX2
ATP8
ATP6
COX3
ND3
ND4L
ND4
ND5
ND6
CYTB

Supplementary Materials Figure S2



Supplementary Materials Figure S3

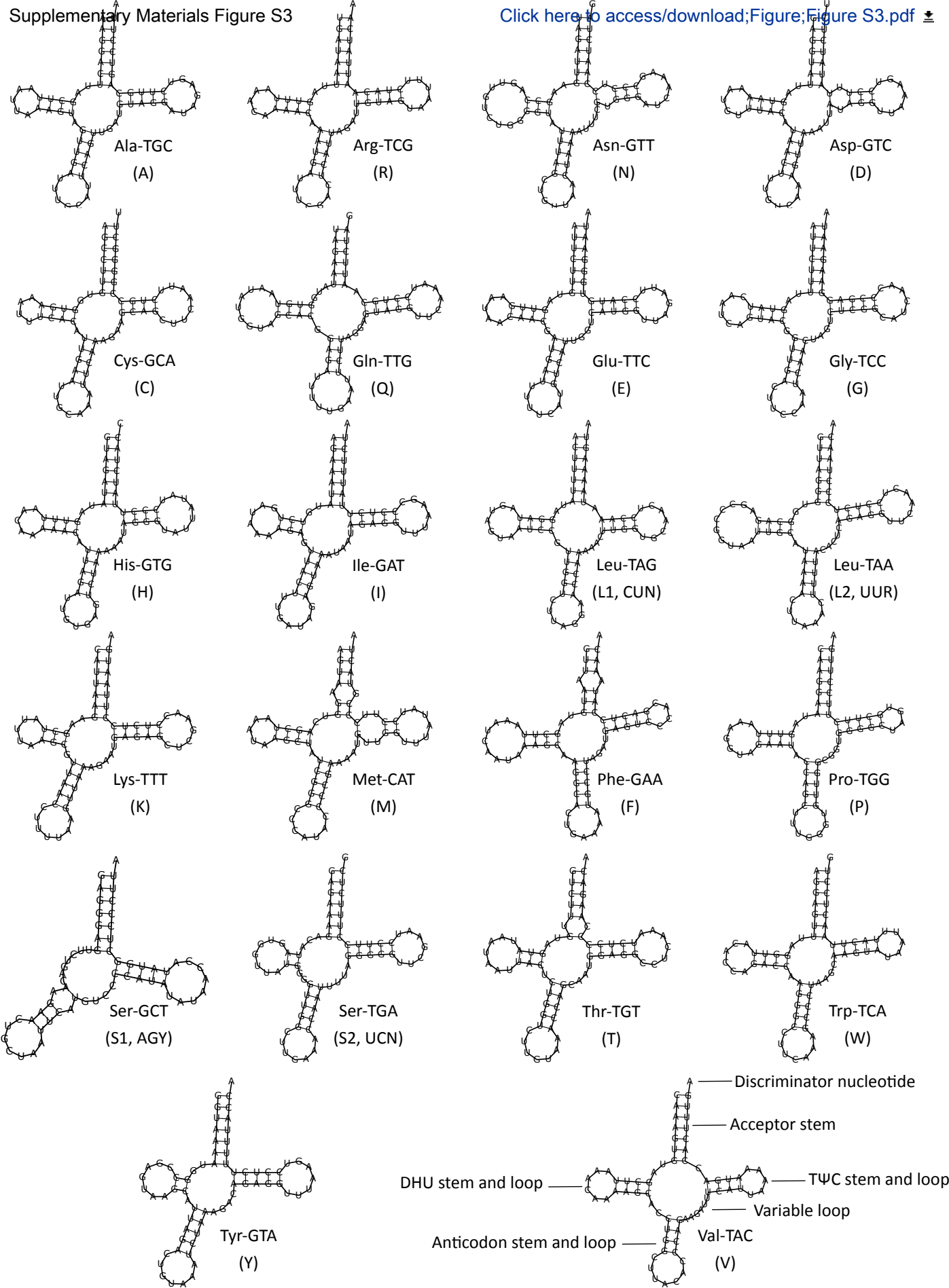


Table S1 Mitogenome nucleotide composition of *Talpa martinorum*.

Gene		Length (bp)	A (%)	C (%)	G (%)	T (%)	A + T (%)	G + C (%)	AT Skew	GC Skew
Whole genome		16,835	34.02	24.54	14.32	27.12	61.14	38.86	0.11	–0.26
PCGs		11,404	33.00	25.40	13.40	28.20	61.20	38.80	0.08	–0.31
	ND1	954	31.50	25.90	13.80	28.80	60.30	39.70	0.04	–0.30
	ND2	1044	39.50	24.30	9.50	26.70	66.20	33.80	0.19	–0.44
	COX1	1545	29.40	23.80	17.40	29.40	58.80	41.20	0.00	–0.16
	COX2	684	33.50	22.60	15.10	28.80	62.30	37.70	0.08	–0.20
	ATP8	202	38.80	22.40	7.50	31.30	70.10	29.90	0.11	–0.50
	ATP6	681	31.30	27.90	12.20	28.60	59.90	40.10	0.05	–0.39
	COX3	784	29.40	25.70	16.90	28.10	57.50	42.60	0.02	–0.21
	ND3	346	32.80	25.50	13.90	27.80	60.60	39.40	0.08	–0.29
	ND4L	297	27.20	21.40	14.60	36.70	63.90	36.00	–0.15	–0.19
	ND4	1378	34.70	25.60	11.50	28.10	62.80	37.10	0.11	–0.38
	ND5	1821	33.50	25.70	13.10	27.70	61.20	38.80	0.09	–0.32
	ND6	528	41.10	29.50	8.40	21.00	62.10	37.90	0.32	–0.56
	CYTB	1140	30.50	27.10	14.30	28.10	58.60	41.40	0.04	–0.31
rRNAs		2541	37.30	20.20	18.20	24.30	61.60	38.40	0.21	–0.05
tRNAs		1516	34.03	21.04	16.10	28.83	62.86	37.14	0.08	–0.13
D-loop		1375	35.70	29.40	12.70	22.20	57.90	42.10	0.23	–0.40

Table S2 Sequence comparisons among the mitogenomes of *Talpa* species (*T. aquitania*, *T. europaea*, *T. martinorum* and *T. occidentalis*)

Gene region	Gene length (bp)				Start/stop codons				Strand			
	T. aqu.	*T. eur.*	*T. mar.*	*T. occ.*	*T. aqu.*	*T. eur.*	*T. mar.*	*T. occ.*	*T. aqu.*	*T. eur.*	*T. mar.*	*T. occ.*
Protein coding genes (PCGs)	11,406	11,406	11,404	11,406								
ND1	954	954	954	954	ATG/TAA	ATG/TAA	ATG/ TAG	ATG/TAA	H	H	H	H
ND2	1044	1044	1044	1044	ATA/TAG	ATA/TAG	ATA/ TAG	ATA/TAG	H	H	H	H
COX1	1545	1545	1545	1545	ATG/TAA	ATG/TAA	ATG/ TAA	ATG/TAA	H	H	H	H
COX2	684	684	684	684	ATG/TAA	ATG/TAA	ATG/ TAA	ATG/TAA	H	H	H	H
ATP8	204	204	202	204	ATG/TAA	ATG/TAA	ATG/ T--	ATG/TAA	H	H	H	H
ATP6	681	681	681	681	ATG/TAA	ATG/TAA	ATG/ TAA	ATG/TAA	H	H	H	H
COX3	784	784	784	784	ATG/T--	ATG/T--	ATG/ T--	ATG/T--	H	H	H	H
ND3	346	346	346	346	ATT/T--	ATT/T--	ATT/ T--	ATT/T--	H	H	H	H
ND4L	297	297	297	297	ATG/TAA	ATG/TAA	ATG/ TAA	ATG/TAA	H	H	H	H
ND4	1378	1378	1378	1378	ATG/T--	ATG/T--	ATG/ T--	ATG/T--	H	H	H	H
ND5	1821	1821	1821	1821	ATT/TAA	ATT/TAA	ATT/ TAA	ATT/TAA	H	H	H	H
ND6	528	528	528	528	ATG/TAA	ATG/TAA	ATG/ TAA	ATG/TAA	L	L	L	L
CYTB	1140	1140	1140	1140	ATG/AGA	ATG/AGA	ATG/ AGA	ATG/AGA	H	H	H	H
12S rRNA and 16S rRNA	2540	2545	2541	2542								
22 tRNAs	1517	1517	1516	1518								
Total length excluding the D-loop	15,456	15,462	15,460	15,458								
D-loop	1370	1422	1375	1504								
Total	16,826	16,884	16,835	16,962								