New mitochondrial genomes of 39 soil dwelling Coleoptera from metagenome sequencing

Carmelo Andujar, Paula Arribas, Michal Motyka, Mathew Bocek, Ladislav Bocak, Benjamin Linard, Alfried Vogler

To cite this version:

Carmelo Andujar, Paula Arribas, Michal Motyka, Mathew Bocek, Ladislav Bocak, et al.. New mitochondrial genomes of 39 soil dwelling Coleoptera from metagenome sequencing. Mitochondrial DNA Part B Resources, Taylor & Francis Online, 2019, 4 (2), pp.2447-2450. 10.1080/23802359.2019.1637289 . hal-02922399

HAL Id: hal-02922399

https://hal.archives-ouvertes.fr/hal-02922399

Submitted on 26 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
New mitochondrial genomes of 39 soil dwelling Coleoptera from metagenome sequencing

Carmelo Andújar, Paula Arribasa, Michal Motyka, Mathew Bocek, Ladislav Bocak, Benjamin Linard and Alfried P. Vogler

ABSTRACT

High-throughput DNA methods hold great promise for the study of the hyperdiverse arthropod fauna of the soil. We used the mitochondrial metagenomic approach to generate 39 mitochondrial genomes from adult and larval specimens of Coleoptera collected from soil samples. The mitogenomes correspond to species from the families Carabidae (6), Chrysomelidae (1), Curculionidae (9), Dermentidae (1), Elateridae (1), Latridiidae (1), Scarabaeidae (3), Silvanidae (1), Staphylinidae (12), and Tenebrionidae (4). All the mitogenomes followed the putative ancestral gene order for Coleoptera. We provide the first available mitogenome for 30 genera of Coleoptera, including endemoge representatives of the genera Torneuma, Coiffaitiella, Otiorynchus, Oligotyphlopsis, and Typhlocharis.

The mitochondrial metagenomics approach (MMG) provides a cost-effective method for sequencing mitochondrial genomes from numerous species (Andújar et al. 2015; Crampton-Platt et al. 2015). Total genomic DNA from multiple specimens, either extracted individually or in bulk, is shotgun sequenced in a metagenomics mixture, followed by assembly with standard genomic assemblers, from which whole mitochondrial genomes emerge preferentially due to their high copy number relative to most of the nuclear genome. This ‘genome skimming’ approach was used to sequence the mitogenomes of beetle specimens collected from soil samples of the southern Iberian Peninsula at Sierra de Grazalema (36.7N, −5.4W), Sierra de Cabra (37.4N, −4.3W) and Sierra Madrona (38.4N, −4.3W) (see Andújar et al. 2015) by following the ‘Floatation-Berlese-Floation’ (FBF) protocol (Arribas et al. 2016). Briefly, aliquots of the DNA extracts from 1494 specimens (vouchered at the Natural History Museum London) were pooled to generate 3 pools with roughly equal molar DNA concentration per specimen, after the dsDNA concentration of extracts was measured (Qubit 2.0 Fluorometer, Life Technologies Corp., Carlsbad, CA). Further, TruSeq DNA libraries were constructed and sequenced in the Illumina MiSeq platform (Illumina Inc., San Diego, CA) (2 × 300 bp; 800–950 bp insert size).

The output was processed and assembled in three assemblers as indicated in Andújar et al. (2017). The resulting contigs were subjected to super-assembly in Geneious 7.1.9 (http://www.geneious.com) using the de novo assembly function and showed wide overlap. The procedure resulted in more than 200 mitogenomes longer than 5000 bp, of which 39 were selected for further annotation and identification to species or genus level. Thirty-four of these include the full set of protein-coding, rRNA and tRNA genes (>15,000 bp), of which 17 were complete circular mitogenomes. The remaining 17 were not circularized due to difficulties with the assembly of the control region. Five additional mitogenomes were incomplete by the lack of one or two genes (sequence length between 12,221 and 14,453 bp).

The mitogenomes were annotated using gene predictions with MITOS (Bernt et al. 2013) and manually refined in Geneious. All mitogenomes were structured following the putatively ancestral gene order for the Coleoptera. Mitogenomes assembled from the shotgun mixture were linked to particular specimens using the cox1 barcode sequences obtained from the same specimens with PCR-Sanger sequencing. For those cases where Sanger sequencing failed (5/39), validation was performed by unambiguous match to the species level on BOLD Public Data Portal (Ratnasingham and Hebert 2007, accessed on 20th March 2019). The mitogenomes correspond to 39 different species from the families Carabidae (6), Chrysomelidae (1), Curculionidae (9), Dermentidae (1), Elateridae (1), Latridiidae (1), Scarabaeidae (3), Silvanidae (1), Staphylinidae (12) and Tenebrionidae (4), and include representatives from 37...
Figure 1. Phylogenetic tree from maximum-likelihood analysis showing the relationships of the 39 newly generated mitogenomes. Circles in branch tips indicate the locality where each specimen was collected (Sierra de Grazalema: black; Sierra de Cabra: grey and Sierra Madrona: white). Shaded frames according with beetle families. GenBank accession numbers are in brackets.

Table 1. Additional data for the 39 mitogenomes of Coleoptera in this study.

| Genbank Accession Numbers | Voucher code | Family | Species | FG** | FSP*** | Life stage | Identification |
|---------------------------|--------------|--------|---------|------|--------|------------|----------------|
| MK692552                  | BMNH 1041149| Carabidae | Microlestes mauntianicus | x    | x      | Adult      | J.L. Lencina det. |
| MK692553                  | BMNH 1042258| Tenebrionidae | Ochotrus unicolor | x    | x      | Adult      | J.L. Lencina det. |
| MK692554                  | BMNH 1041892| Curculionidae | Teneuma sp. | x    | x      | Adult      | C. Hernando det. |
| MK692556                  | BMNH 1044019| Staphylinidae | Achenium seditiosum | x    | x      | Adult      | V. Assing det. |
| MK692557                  | BMNH 1041971| Carabidae | Typhlocharis sp. | x    | x      | Adult      | C. Andújar det. |
| MK692559                  | BMNH 1041157| Tenebrionidae | Microlestes reitteri | x    | x      | Adult      | J.L. Lencina det. |
| MK692560                  | BMNH 1042672| Staphylinidae | Othius myrmecophilus | x    | x      | Larva      | BOLD match > 99% |
| MK692567                  | BMNH 1042021| Staphylinidae | Oligotyplopsis sp. | x    | x      | Adult      | C. Hernando det. |
| MK692568                  | BMNH 1042062| Curculionidae | Tychius pusillus | x    | x      | Adult      | BOLD match > 99% |
| MK692574                  | BMNH 1041943| Staphylinidae | Tachyporus nitidulus | x    | x      | Adult      | V. Assing det. |
| MK692579                  | BMNH 1043732| Staphylinidae | Trechus Obtusus | x    | x      | BOLD match > 99% |
| MK692585                  | BMNH 1041967| Elateridae | Hypera postica | x    | x      | Adult      | BOLD match > 99% |
| MK692586                  | BMNH 1041911| Curculionidae | Centopus elongatus | x    | x      | Adult      | C. Hernando det. |
| MK692587                  | BMNH 1042074| Curculionidae | Olyrhina longiseta | x    | x      | Adult      | BOLD match > 99% |
| MK692591                  | BMNH 1041150| Carabidae | Trechus obtusus | x    | x      | Adult      | J.L. Lencina det. |
| MK692592                  | BMNH 1042238| Tenebrionidae | Scaurus uncinus | x    | x      | Adult      | J.L. Lencina det. |
| MK692593                  | BMNH 1043977| Staphylinidae | Atheta sp. | x    | x      | Adult      | V. Assing det. |
| MK692599                  | NA           | Staphylinidae | Ocyopus serthios | x    | NA      | BOLD match > 99% |
| MK692601                  | BMNH 1042249| Staphylinidae | Medon sp. | x    | x      | Adult      | V. Assing det. |
| MK692603                  | BMNH 1042190| Staphylinidae | Micusculus testaceus | x    | x      | Adult      | V. Assing det. |
| MK692605                  | BMNH 1042074| Curculionidae | Hypera postica | x    | x      | Adult      | BOLD match > 99% |
| MK692606                  | BMNH 1042031| Tenebrionidae | Cinemoplastia atropis | x    | x      | Adult      | J.L. Lencina det. |
| MK692607                  | BMNH 1042209| Scabaeidae | Phaenops caesi | x    | x      | Adult      | J.L. Lencina det. |
| MK692609                  | BMNH 1041982| Curculionidae | Protapion trifoli | x    | x      | Adult      | J.L. Lencina det. |
| MK692616                  | BMNH 1041162| Staphylinidae | Geotrupidae | x    | x      | Adult      | V. Assing det. |
| MK692625                  | NA           | Chrysomelidae | Cryptoglyphus pilosus | x    | NA      | BOLD match > 99% |
| MK692626                  | BMNH 1042569| Carabidae | Calathus granatensis | x    | x      | Larva      | BOLD match > 99% |
| MK692638                  | NA           | Staphylinidae | Lomechusa pubicollis | x    | x      | NA         | BOLD match > 99% |
| MK692642                  | BMNH 1041893| Silvanidae | Oryzaephilus surinamensis | x    | x      | NA         | J.L. Lencina det. |
| MK692645                  | NA           | Curculionidae | Echinidea andalusiensis | x    | x      | NA         | BOLD match > 99% |
| MK692646                  | BMNH 1042067| Curculionidae | Otionyphus hyalinus | x    | x      | Adult      | C. Hernando det. |
| MK692648                  | BMNH 1042068| Scabaeidae | Eysmus pusillus | x    | x      | Adult      | J.L. Lencina det. |
| MK692661                  | BMNH 1042524| Staphylinidae | Anthaxius mustus | x    | x      | Larva      | BOLD match > 99% |
| MK692677                  | BMNH 1041924| Latridiidae | Carpicura sp. | x    | x      | Adult      | J.L. Lencina det. |
| MK692678                  | BMNH 1042175| Dermestidae | Thorius sp. | x    | x      | Adult      | J.L. Lencina det. |
| MK692681                  | NA           | Carabidae | Sycius flavus | x    | x      | NA         | BOLD match > 99% |
| MK692702                  | BMNH 1042255| Staphylinidae | Mecynothorax fuscus | x    | x      | Adult      | BOLD match > 99% |
| MK692707                  | BMNH 1042182| Tenebrionidae | Ctenorius elongatus | x    | x      | Adult      | J.L. Lencina det. |

Genbank Accession Numbers, voucher codes, taxonomic identification, life stage, and information on whether the provided mitogenomes are the first available for the genus (FG column) and for the species (FSP column).

**All mitogenomes were obtained by bulk sequencing of a mix of specimens. Voucher code refers to the specimen with PCR-Sanger sequencing that matches (100% similarity) the obtained mitogenomes. Vouchers are not available for mitogenomes if PCR-Sanger sequencing failed for a particular specimen.

***FG: Marked with ‘x’ if the mitogenome is the first available for the genus.

**FSP: Marked with ‘x’ if the mitogenome is the first available for the species.
genera. For 30 of these genera, we provide the first available mitogenome and only two species (Oryzaephilus surinamensis and Hypera postica) have an available mitogenome. The new mitogenomes include endogean representatives of the genera Toreuma, Coiffaitiella, Otiorynchus, Oligotyphlopsis, and Typhlocharis. For further details on specimens and mitogenomes see Figure 1, Tables 1 and 2, and GenBank Accession Numbers.

For the 39 newly generated mitogenomes, the 13 protein-coding genes (PCGs) were extracted using Geneious and individually aligned using the FFT-NS-i-x2 algorithm of MAFFT (Katoh et al. 2002). Individual gene alignments were trimmed and concatenated to get a final dataset of 39 taxa and 12,940 bp. This alignment was used for phylogenetic inference using Maximum-likelihood in IQ-TREE (Nguyen et al. 2015), performed through the IQ-TREE web server (Trifinopoulos et al. 2016) without data partitioning, allowing the software to determine the best-fit substitution model and estimating an ultrafast bootstrap approximation with 10,000 replicates. The obtained tree showed the expected relationships among the families within Coleoptera, including the monophyly of the suborders Adephaga and Polyphaga and the monophyly of all families (with more than one mitogenome) (Figure 1).

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
The study was funded by the NHM Biodiversity Initiative (UK) and the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad) under a project with the number CGL2015-74178-JIN.

References
Andújar C, Arribas P, Linard B, Kundrata R, Bocak L, Vogler AP. 2017. The mitochondrial genome of Iberobaenia (Coleoptera: Iberobaenidae): first rearrangement of protein-coding genes in the beetles. Mitochondrial DNA Part A. 28:156–158.
Andújar C, Arribas P, Ruzicka F, Platt AC, Timmermans MJTN, Vogler AP. 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. Mol Ecol. 24:3603–3617.

Table 2. Sampling localities for the 39 mitogenomes of Coleoptera in this study.

| GB accession | Locality* | Latitude (N) | Longitude (W) | Altitude (Meters) | Habitat |
|---------------|-----------|--------------|---------------|-------------------|---------|
| MK692552      | La Dehesilla, Benaocaz, Cádiz, Spain | 36.7074 | −5.4570 | 480 | Olea europaea field |
| MK692553      | N-420 km 105, Fuencaliente, Ciudad Real, Spain | 38.4445 | −4.3247 | 730 | Grassland-riverside |
| MK692554      | Arroyo del Espino, El Bosque, Cádiz, Spain | 36.7613 | −5.5069 | 275 | Riverside |
| MK692556      | Nava de Cabra, Cabra, Córdoba, Spain | 37.4856 | −4.3634 | 995 | Grassland |
| MK692557      | La Dehesilla, Benaocaz, Cádiz, Spain | 36.7074 | −5.4567 | 470 | Grassland |
| MK692559      | La Dehesilla, Benaocaz, Cádiz, Spain | 36.7074 | −5.4570 | 480 | Olea europaea field |
| MK692560      | Robledo de las Hoyas, Fuencaliente, Ciudad Real, Spain | 38.4371 | −4.3413 | 950 | Quercus faginea forest |
| MK692567      | Arroyo del Bosque, El Bosque, Cádiz, Spain | 36.7667 | −5.5011 | 290 | Riverside |
| MK692568      | Llanos del Repúblicano, Villaluenga del Rosario, Cádiz, Spain | 36.6817 | −5.3574 | 810 | Grassland |
| MK692574      | Llanos del Repúblicano, Villaluenga del Rosario, Cádiz, Spain | 36.6907 | −5.3639 | 925 | Quercus suber forest |
| MK692579      | Huertediondia, Tavizna, Benaocaz, Cádiz, Spain | 36.7192 | −5.4850 | 360 | Olea europaea field |
| MK692585      | Colada de la Brena, Benaocaz, Cádiz, Spain | 36.7070 | −5.4704 | 430 | Quercus suber forest |
| MK692586      | El Pinsapar, Puerto del Pinar, Grazalema, Cádiz, Spain | 36.7726 | −5.4240 | 1115 | Abies pinsapo forest |
| MK692587      | Puerto del Boyar, Grazalema, Cádiz, Spain | 36.7536 | −5.3939 | 1120 | Grassland |
| MK692591      | La Dehesilla, Benaocaz, Cádiz, Spain | 36.7074 | −5.4570 | 480 | Olea europaea field |
| MK692592      | Arroyo del Bosque, El Bosque, Cádiz, Spain | 36.7667 | −5.5011 | 290 | Riverside |
| MK692593      | Cortijo del Nazauzelo, Carcabuye, Córdoba, Spain | 37.4852 | −4.3412 | 1035 | Grassland |
| MK692597      | Ermita Nta. Sra. de la Sierra, Cabra, Córdoba, Spain | 37.4905 | −5.4813 | 1145 | Pinus halepensis |
| MK692599      | Sierra de Cabra, Córdoba, Spain | NA | NA | NA | NA |
| MK692601      | Ladera de la Casa de Cipriano, Fuencaliente, Ciudad Real, Spain | 38.4190 | −4.3138 | 765 | Quercus suber forest |
| MK692603      | Nava de Cabra, Cortijo de los Benitez, Cabra, Córdoba, Spain | 37.4856 | −4.3634 | 995 | Grassland |
| MK692605      | Casa de la Vinuela, Cabra, Córdoba, Spain | 37.4852 | −4.3861 | 1020 | Quercus faginea forest |
| MK692606      | Llanos del Repúblicano, Villaluenga del Rosario, Cádiz, Spain | 36.6857 | −5.3648 | 910 | Quercus ilic forest |
| MK692607      | Arroyo del Palancar, Carcabuye, Córdoba, Spain | 37.4827 | −4.3676 | 525 | Riverside |
| MK692609      | Arroyo del Bosque, El Bosque, Cádiz, Spain | 36.7667 | −5.5011 | 290 | Riverside |
| MK692616      | Llanos del Campo, Benamahoma, Cádiz, Spain | 36.7556 | −5.4556 | 642 | Quercus ilic forest |
| MK692625      | Sierra de Grazalema, Cádiz, Spain | NA | NA | NA | NA |
| MK692626      | Llanos del Repúblicano, Villaluenga del Rosario, Cádiz, Spain | 36.6907 | −5.3639 | 925 | Quercus suber forest |
| MK692638      | Sierra Madrona, Ciudad Real, Spain | NA | NA | NA | N.A. |
| MK692642      | El Boyar, Cortijo del Santo, Grazalema, Cádiz, Spain | 36.7549 | −5.4194 | 920 | Quercus ilic forest |
| MK692645      | Sierra de Grazalema, Cádiz, Spain | NA | NA | NA | NA |
| MK692646      | Ermita Nta. Sra. de la Sierra, Cabra, Córdoba, Spain | 37.4905 | −5.4813 | 1145 | Pinus halepensis |
| MK692648      | Nava de Cabra, Cabra, Córdoba, Spain | 37.5067 | −5.4671 | 968 | Quercus ilic forest |
| MK692661      | Camino Viejo a la Ermita, Cabra, Córdoba, Spain | 37.4811 | −4.3885 | 970 | Grassland |
| MK692677      | Arroyo del Espino, El Bosque, Cádiz, Spain | 36.7613 | −5.5069 | 275 | Riverside |
| MK692678      | Cortijo del Nazauzelo, Carcabuye, Córdoba, Spain | 37.4852 | −4.3412 | 1035 | Grassland |
| MK692701      | Sierra de Grazalema, Cádiz, Spain | NA | NA | NA | NA |
| MK692702      | Collado del Nazauzelo, Carcabuye, Córdoba, Spain | 37.4801 | −4.3347 | 995 | Olea europaea field |
| MK692707      | Nava de Cabra, Cortijo de los Benitez, Cabra, Córdoba, Spain | 37.4856 | −4.3634 | 995 | Grassland |

*All mitogenomes were obtained by bulk sequencing of a mix of specimens, and voucher codes to particular specimens assigned by the PCR-Sanger sequencing that matches (100% similarity) the obtained mitogenomes (see Table 1). Metagenomes not linked to a particular vouchered specimen are assigned to a locality but lack detailed information (precise coordinates, altitude, and habitat).
Arribas P, Andújar C, Hopkins K, Shepherd M, Vogler APAP, Andújar C, Hopkins K, Shepherd M, Vogler APAP. 2016. Metabarcoding and mitochondrial metagenomics of endogeans arthropods to unveil the mesofauna of the soil. Methods Ecol Evol. 7:1071–1081.
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.
Crampton-Platt A, Timmermans MJTN, Gimmel ML, Kutty SN, Cockerill TD, Vun Khen C, Vogler AP. 2015. Soup to tree: The phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean Rainforest sample. Mol Biol Evol. 32:2302–2316.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res. 30:3059–3066.
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–274.
Ratnasingham S, Hebert PDN. 2007. BARCODING, BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes. 7:355–364.
Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44:W232–W235.