ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

HERMES: an Improved Method to Test Mitochondrial Genome Molecular Synapomorphies among Clades.

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Federico Plazzi, G.P. (2021). HERMES: an Improved Method to Test Mitochondrial Genome Molecular Synapomorphies among Clades. MITOCHONDRION, 58, 285-295 [10.1016/j.mito.2021.02.009].

Availability:
This version is available at: https://hdl.handle.net/11585/832445 since: 2021-09-16

Published:

DOI: http://doi.org/10.1016/j.mito.2021.02.009

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

07 November 2026

This is the final peer-reviewed accepted manuscript of:

Federico Plazzi, Guglielmo Puccio, Marco Passamonti, HERMES: An improved method to test mitochondrial genome molecular synapomorphies among clades, Mitochondrion, Volume 58, 2021, Pages 285-295

The final published version is available online at: ***https://dx.doi.org/10.1016/j.mito.2021.02.009***

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

HERMES: an Improved Method to Test Mitochondrial Genome Molecular Synapomorphies among Clades[1]

Federico Plazzi[a]*

Guglielmo Puccio[a1]

Marco Passamonti[a]

[a] Department of Biological, Geological and Environmental Sciences, University of Bologna, via Selmi, 3 – 40126 Bologna (Italy). E-mail addresses: federico.plazzi@unibo.it (FP); guglielmo.puccio@unipa.it (GP); marco.passamonti@unibo.it (MP).

[1] Present address: Department of Agricultural, Food and Forestry Sciences, University of Palermo, Palermo (Italy).

* Author for Correspondence: Federico Plazzi, Department of Biological, Geological and Environmental Sciences, University of Bologna, via Selmi, 3 – 40126 Bologna (Italy), telephone number +39 051 20 94 172, fax number +39 051 20 94 173, email address federico.plazzi@unibo.it.

[1] Abbreviations: Amount of Mitochondrial Identical Gene Arrangement, AMIGA; Doubly Uniparental Inheritance, DUI; mitochondrial genome, mtDNA; Open Reading Frame, ORF; Strand Usage, SU; Unassigned Region, UR.

Abstract

Mitochondrial chromosomes have diversified among eukaryotes and many different architectures and features are now acknowledged for this genome. Here we present the improved HERMES index, which can measure and quantify the amount of molecular change experienced by mitochondrial genomes. We test the improved approach with ten molecular phylogenetic studies based on complete mitochondrial genomes, representing six bilaterian Phyla. In most cases, HERMES analysis spotted out clades or single species with peculiar molecular synapomorphies, allowing to identify phylogenetic and ecological patterns. The software presented herein handles linear, circular, and multi-chromosome genomes, thus widening the HERMES scope to the complete eukaryotic domain.

Keywords

Mitochondrial genome; genome architecture; molecular synapomorphy; Bilateria; HERMES index.

1 Introduction

1.1 The myth of a conserved genome

As remarked in Lavrov and Pett (2016), the myth of a "typical" and conserved metazoan mitochondrial genome (mtDNA) arose when the complete mtDNA of *Drosophila yakuba* was obtained (Clary and Wolstenholme 1984) and found to share most features with the already available human mtDNA (Anderson et al. 1981): most notably, a size of approximately 16 kb, a circular topology, the same 37 genes (*atp6*, *atp8*, *cox1-3*, *cytb*, *nad1-6*, *nad4l*, 2 rRNAs, and 22 tRNAs), a single large noncoding region, incomplete stop codons, absence of introns. Given the huge evolutionary distance between humans and fruit flies, it was assumed these to be the common (and basal) features of all animal mtDNAs (Clary and Wolstenholme 1985; see also Saccone et al. 2002).

A comparably small size of mtDNA was measured in representatives of several animal groups (Borst and Kroon 1969). In the same years, mtDNAs of chickens, cows and mice turned out to be circular (Sinclair and Stevens 1966; Van Bruggen et al. 1966). As reconstructed in Williamson (2002), this led to a widespread belief in a small, circular molecule to be present in all mitochondria. Consequently, the claims for linear mtDNA examples that did arise in the same years (Suyama and Miura 1968) were discarded as naturalistic oddities, and the linear molecules that were repeatedly observed within yeast mitochondria were interpreted as broken circles for approximately 30 years (Williamson 2002).

Currently, the query "mitochondrion, complete genome" yields 95,406 results from the NCBI Nucleotide database (accessed on the 9th of January, 2021), yet the common ideas about mitochondrial genomes are still grounded on the first results and are spread in many textbooks and lectures. Nonetheless, there is overwhelming evidence that, while a small, circular, intronless, somewhat conserved mtDNA is in fact common among bilaterians, many eukaryotic

lineages show genome structures and organizations that are contrasting to various extents with this oversimplified conception.

1.2 Variability in mtDNA architecture

The mitochondrial proteome size is estimated between 1,000 and 1,500 proteins (e.g., Muthye and Lavrov 2020; Becker et al. 2012; Chacinska et al. 2009; Meisinger et al. 2008; Gabaldón and Martijn 2004); however, only a small cluster of protein coding genes (PCGs) has been retained on mtDNA: typically, 13 PCGs in metazoans (e.g., Breton et al. 2014; Boore 1999), and even as few as 3 in apicomplexans (Rehkopf et al. 2000; Feagin 1994). Most genes of the original endosymbiont were therefore lost or have been transferred to the nucleus – a process termed Genome Reductive Evolution (GRE; Kannan et al. 2014; Ghiselli et al. 2013; Khachane et al. 2007; Andersson and Kurland 1998).

Among Eukaryotes, jakobids (order Jakobida) comprise the supergroup Discoba along with Tsukubamonadida, Heterolobosea, and Euglenozoa (see Gray et al. 2020; Hampl et al. 2009; Derelle et al. 2015; and references therein). Jakobids show the most gene-rich mtDNAs of all eukaryotes, with many bacterial features, and the mtDNA of the species *Andalucia godoyi* was found to encode the most genes to our knowledge (Burger et al. 2013; Lang et al. 1997). Notably, all eukaryotic mitochondria harbor a single-subunit, bacteriophage-like, nucleus-encoded RNA polymerase, while jakobids' mtDNA encodes subunits of a bacterial-like RNA polymerase (Yabuki et al. 2018; Burger et al. 2013; Shutt and Gray 2006; Gray and Lang 1998) and no evidence of a nucleus-encoded RNA polymerase supporting the mitochondrion-encoded one was found in a recent genomic draft of *A. godoyi* (Gray et al. 2020). Moreover, 21% (31 on at least 145) proteins involved in the energetic metabolism and 19% (29 on at least 149) proteins involved in the translation apparatus are mitochondrially encoded in *A. godoyi* (Gray et al. 2020). These and many other observations point to the fact that mtDNA in jakobids retains most

ancestral features of LECA, the Last Eukaryotic Common Ancestor (Gray et al. 2020; Bullerwell and Gray 2004; Lang et al. 1997).

Contrastingly, in the very same eukaryotic supergroup Discoba, the phylum Euglenozoa displays completely different mtDNAs. The phylum Euglenozoa is comprised by Symbiontida, Euglenida, Kinetoplastea, and Diplonemea (Kaur et al. 2020; Adl et al. 2019; Burki et al. 2019). The mtDNA of the ill-famed pathogen *Trypanosoma brucei* (Kinetoplastea) is organized as a kinetoplast body (see Jensen and Englund 2012), a compact network of DNA maxicircles (~25 kb) and thousands of minicircles (~1 kb), carrying the common mitochondrial genes and small RNAs involved in gene editing, respectively (Burger and Valach 2018; Lukeš et al. 2018). Even weirder is the organization of mtDNA in Diplonemea: in *Diplonema papillatum* the mtDNA is composed by 81 circular chromosomes that are not interwoven in a kinetoplast body (Burger and Valach 2018; Maslov et al. 1999). Chromosomes can be subdivided into two size classes (6 and 7 kb-long); chromosomes of the same class share approximately 95% of the sequence, while the remainder is unique for each chromosome and carries the only coding region of the chromosome, a fragment of a gene ranging 40 to 540 bp in length (Kaur et al. 2020). In other diplonemids, some chromosomes bear multiple (sometimes overlapping) cassettes (Burger and Valach 2018). Each chromosome is transcribed independently and a complex machinery of *trans*-splicing and post-transcriptional modifications leads to the mature mRNA (Kaur et al. 2020; Burger and Valach 2018). This unconventional architecture makes diplonemid mtDNA a highly disperse genome, and gigantic as well: in the genus *Perkinsela*, the mitochondrion harbors more than six times DNA than the nucleus (Lukeš et al. 2018).

Among Viridiplantae, mtDNAs from green algae (Chlorophyta) range from >200-kb long genomes of Bryopsidales (Repetti et al. 2020) to 95.9-kb long genome of *Pseudendoclonium akinetum* to the very small, highly derived mtDNAs of the genus *Chlamydomonas*, with a reduced set of genes and fragmented rRNAs (Bullerwell and Gray 2004; Denovan-Wright et al.

1994; Michaelis et al. 1990; Boer and Gray 1988). Charophytes are the closest algal relatives to land plants (together comprising Streptophyta) and show mtDNAs smaller than 70 kb (Bullerwell and Gray 2004 – but up to 106 kb in *Klebsormidium nitens*; Hori et al. 2014), while larger genomes are found in land plants, up to 570 kb in *Zea mays*, 2,400 kb in some Cucurbitaceae, >6 Mb in some conifers and 11.3 Mb in some species of the genus *Silene* (Guo et al. 2020; Sloan et al. 2012; Bullerwell and Gray 2004; Ward et al. 1981). The enormous size of mtDNA in some chlorophytes and vascular plants is typically due to expansions of noncoding regions and to the proliferation of introns (Repetti et al. 2020; Bullerwell and Gray 2004), and gene content is somewhat conserved among green plants (Wu et al. 2020; Sloan 2013; Sloan et al. 2012; but see Skippington et al. 2015; Mower et al. 2012). Moreover, although plant mtDNA is conventionally annotated as a single circular molecule, the real structure of plant mtDNA involves entangled branching patterns and alternative molecules that coexist within cells (Wu et al. 2020; Kozik et al. 2019; Gualberto and Newton 2017; Sloan 2013; Bendich 1993; Palmer and Shields 1984), and some mitochondria may contain only partial or no genome at all (Preuten et al. 2010). Inter- and intra-mitochondrial recombination is repeatedly reported (Tsujimura et al. 2019; McCauley 2013; Abdelnoor et al., 2003), a process which can generate multiple subgenomes (Barr et al. 2005). In other cases, however, the mtDNA appears to be subdivided into autonomous circles, as is the case in onion (Tsujimura et al. 2019), in some *Silene* species (Sloan et al. 2012) and in the cucumber (Alverson et al. 2011). Finally, extensive posttranscriptional modifications are required, including *cis*- and *trans*-intron splicing and RNA editing (Ichinose and Sugita 2016; Bonen 2012; Mower et al. 2012).

The large supergroup of Opisthokonta encompasses Metazoa, Fungi and a wide array of eukaryotes, showing different mtDNA structures. While basal chytridiomycetes typically show circular mtDNAs coding for most genes present in the opisthokont common ancestor, including in some cases a full complement of tRNAs, other mtDNAs from Fungi lost many genes

(Bullerwell and Gray 2004). We mentioned above the broken-circle theory: now it is known that mtDNA of yeast is composed by a large population of interspersed linear chromosomes (Dujon 2020; Freel et al. 2015; Williamson 2002).

At the animal-fungal boundary, among Ichthyosporea (Adl et al. 2019), *Amoebidium parasiticum* (order Ichthyophonida) has a large (>200 kb) mtDNA bloated with introns, consists of hundreds of linear chromosomes, and only the 20% of the molecule is coding (Burger et al. 2003); conversely, *Sphaerotecum destruens* (order Dermocystida) has a compact (~24 kb), intronless mtDNA, where more than the 96% of the molecule is coding (Sana et al. 2020). Within Holozoa, the only (to our knowledge) choanoflagellate whose complete mitochondrial DNA has been sequenced is *Monosiga brevicollis*, that exhibits a 76-kb long mtDNA; though coding for 55 genes, the coding region covers the 47% of the genome (Sana et al. 2020; Osigus et al. 2013; Bullerwell and Gray 2004).

Non-bilaterian animals show a wide array of different mtDNA structures, organizations, and gene contents. The mitochondrial genome of placozoans is probably the closest to the ancestral animal mtDNA (Bernt et al. 2013): it is a large, circular molecule with a full complement of tRNAs (Signorovitch et al. 2007); however, the *atp8* gene is lacking and introns are commonly found, including a one-base long micro-exon (Osigus et al. 2017, 2013). Sponges have a variable number of tRNAs, from 2 to 27, and some PCGs are uniquely found, like *tatC* and *atp9*; notably, whereas most sponges have a single, circular mtDNA (Lavrov and Pett 2016), the calcareous sponge *Clathrina clathrus* has six linear mitochondrial chromosomes (Osigus et al. 2013), but other calcaronean sponges have hundreds (Lavrov and Pett 2016). Introns and very few tRNAs have been found in cnidarians (see, f.i., Chi et al. 2019; Zhang et al. 2017; Osigus et al. 2013); while anthozoans display single, circular mitochondrial chromosomes, other cnidarians display linear mitochondrial chromosomes, ranging from one in scyphozoans to one or two in hydrozoans to eight in cubozoans (Lavrov and Pett 2016; Osigus et al. 2013).

Ctenophores harbor single, circular, very small mtDNAs, with rapid evolutionary rates and without many genes that are common in bilaterians (tRNAs, *atp6*, *atp8*; Schultz et al. 2020; Lavrov and Pett 2016; Bernt et al. 2013; Kohn et al. 2012; Pett et al. 2011) – on the other hand, unidentified open reading frames (URFs), not connectible to each other, are described in ctenophore mtDNA (Schultz et al. 2020; Arafat et al. 2018).

Most bilaterians animals do share a single, circular mitochondrial chromosome, with a relatively stable complement of genes (e.g., Boore 1999; Breton et al. 2014); however, many exceptions to that are actually known. Linear molecules and concatenamers have been reported from mammals (Pohjoismäki and Goffart 2011; and references therein). Species of the genus *Globodera* (Nematoda) have at least six recombining circular chromosomes (Gibson et al. 2007a,b), and many insect genera show a mtDNA structured into several minicircles, up to 20 in mammalian sucking lice and 17 in avian feather lice (Sweet et al. 2020; Shao et al. 2017; Shi et al. 2016; Dickey et al. 2015; Herd et al. 2015; Shao et al. 2015; Dong et al. 2014a; Dong et al. 2014b; Jiang et al. 2013; Shao et al. 2012; Wei et al. 2012; Shao et al. 2009). While in *Globodera* the minicircles harbor multiple copies of the mitochondrial genes, as well as pseudogenes, in the parasitic louse *Columbicula* each minicircle harbors a single PCG and a total of two genes on average, including minicircles with tRNAs only (Sweet et al. 2020; Gibson et al. 2007a).

Finally, even the "canonical" gene content shows some peculiarities. Open Reading Frames (ORFs) with no known homology to other genes (termed ORFans or URFs) are known in some bilaterian mtDNAs, including brachiopods, bivalves, and humans (Plazzi et al. 2016, 2013; Breton et al. 2014; Cohen 2014; Lee et al. 2013; Endo et al. 2005). Mitochondrial introns are almost completely absent, yet they have been signaled in a few annelid species (Vallès et al. 2008; Richter et al. 2015). Long and small noncoding RNAs have been signaled to transcribe from the mtDNA (Larriba et al. 2018; Riggs et al. 2018; Dong et al. 2017; Vendramin et al.

2017; Ro et al. 2013; Landerer et al. 2011; Mercer et al. 2011; Rackham et al. 2011), and some of these elements play a role even in nuclear gene regulation (smithRNAs; Passamonti et al. 2020; Pozzi et al. 2017). Bilaterian tRNAs lost many identity elements and this must have entailed a complex coevolution of nucleus-encoded aminoacyl-tRNA synthetases (Kuhle et al. 2020).

1.3. Genome molecular synapomorphies

Not only is the mitochondrial genome highly plastic in terms of structure and organization: the genome plasticity often displays phylogenetic patterns, with peculiar mtDNA features associated to precise clades (e.g., Gray et al. 2020; Guo et al. 2020; Schultz et al 2020; Sweet el at. 2020; Žihala and Eliáš 2019; Li et al. 2018; Lavrov and Pett 2016; Plazzi et al. 2016; Freel et al. 2015; Bernt et al. 2013; Osigus et al. 2013): indeed, Song and colleagues (2019) proposed to erect a clade for the monophyletic cluster comprised by the three mammal-associated louse lineages, naming it Mitodivisia, which in fact underlines a peculiar mitochondrial feature to distinguish the clade itself. In this paper, we present the updated and improved version of the Hyper-Empirical Relative Mitochondrial Evolutionary Speed index (HERMES; Plazzi et al. 2016), which has been developed in order to quantify the amount of mitochondrial evolution that led to mitochondrial molecular synapomorphies in the group of interest, detecting taxa or clades that underwent peculiar processes of mitochondrial evolution. The HERMES approach can be used to trace the evolution of mtDNA in a focal group of organisms.

2 Material and Methods

2.1 Global overview

We presented the basics of the HERMES index approach in a previous paper (Plazzi et al. 2016), in the context of bivalve mitogenomics. Even if the method was developed working on bilaterian mitogenomes, it has now been improved and it can be virtually applied to all eukaryotes. In this section, we describe the main outlines of the method. (i) Hyper-Empirical. The basic idea of this method is to compute from annotated complete mitochondrial genomes many different variables, related to nucleotide composition, gene content, phylogeny, and more (see below). Therefore, the analysis is carried out after phylogenetic and genomic analyses. These empirical measurements are then summarized in a single number (the HERMES index of a given mtDNA) using a maximum likelihood factor analysis. (ii) Relative. The amount of molecular evolution of focal mtDNAs is quantified using a user-selected taxon as a benchmark. This taxon should be chosen since it retains most genomic plesiomorphies of the group, given the state-of-art knowledge about mitogenomics of the group of interest. (iii) Mitochondrial. The method can cope with single-chromosome, as well as multiple-chromosome mtDNAs, be they linear or circular; moreover, simple adjustments may account for plastid genomes (basically, a new gene nomenclature dictionary would be needed; see below). (iv) Evolutionary Speed. We term “evolutionary speed” the amount of genomic change, at different molecular organization levels (i.e., from nucleotide composition to gene arrangement), that each mtDNA underwent with respect to the benchmark taxon. This can be useful to trace single taxa, as well as groups of taxa, that experienced characteristic patterns of molecular evolution.

We developed a tool to compute the HERMES index, providing GenBank annotations or a list of GenBank Accession Numbers, gene alignments and a phylogenetic tree. The tool was written in Python3 and R (R Development Core Team 2008). Moreover, the software RAxML 8.2.12

(Stamatakis 2014) is needed to run HERMES and is called internally. The tool is freely available for download at https://github.com/federicoplazzi/HERMES.

2.2 Genomic variables

At first, the following 14 variables are computed for each provided mtDNA.

(i) Unassigned Region (UR) content. The percentage of URs within the mtDNA is computed, i.e. the amount of genomic positions not annotated as genes.

(ii) Amount of Mitochondrial Identical Gene Arrangement (AMIGA). Following Plazzi et al. (2016), AMIGA is defined as follows.

$$AMIGA = \frac{N_{IGA} - 1}{N - 1}$$

where N_{IGA} is the number of taxa in the group of interest that share an Identical Gene Arrangement (IGA) on the molecule with the focal taxon (including the focal taxon itself), and N is the number of taxa. Taxa with unique gene arrangements will therefore score an AMIGA equal to 0; on the other hand, if all taxa share the same gene arrangement, every taxon will score an AMIGA equal to 1. Values between 0 and 1 are associated to an intermediate degree of conservation of the gene arrangement. Because of the variability of rRNA and tRNA positions in many taxa, they are excluded from the analysis, and the AMIGA index relies on PCGs only.

(iii) Strand Usage (SU) skew. Following Plazzi et al. (2016), SU skew is defined as follows.

$$SU\ skew = \frac{H - L}{H + L}$$

where H is the number of genes annotated on (what is regarded as) the H strand and L is the number of genes annotated on (what is regarded as) the L strand. In case of a perfectly balanced SU, $H = L$, and the SU skew is equal to 0. Conversely, a negative SU skew indicates a bias towards the L strand, and a positive SU skew indicates a bias towards the H strand.

(iv) Root-to-tip distance. It computed for each OTU on the user-supplied best-known likelihood (BKL) tree using the Python package ETE (Huerta-Cepas et al. 2010).

(v) Maximum Likelihood (ML) distance from an outgroup. The ML distance is computed between the focal taxon and a benchmark outgroup taxon selected by the user; the ML estimate is carried out with RAxML 8.2.12 specifying the same model that was used to compute the BKL tree.

(vi) AT content.

(vii) AT skew. Following Reyes et al. (1998), AT skew is defined as follows.

$$AT\ skew = \frac{A - T}{A + T}$$

where A is the percentage of A on the reference sequence and T is the percentage of T on the reference sequence.

(viii) GC skew. Following Reyes et al. (1998), GC skew is defined as follows.

$$GC\ skew = \frac{G - C}{G + C}$$

where G is the percentage of G on the reference sequence and C is the percentage of C on the reference sequence.

(ix) Number of (annotated) genes.

(x) Length of the molecule (bp).

(xi) Codon Adaptation Index (CAI). Following Sharp and Li (1987) and recommendations by Xia (2007), CAI is defined as follows.

$$CAI = \left(\prod_{l=1}^{L} w_l\right)^{\frac{1}{L}} = \left(\prod_{l=1}^{L} \frac{N_{ij}}{N_{imax}}\right)^{\frac{1}{L}}$$

or, to overcome real number underflow problems in computer calculations,

$$CAI = e^{\left(\frac{1}{L}\sum_{l=1}^{L} \ln w_l\right)}$$

which is the geometric mean of the relative adaptiveness w of the L codons of the mtDNA, excluding initiation and termination codons. In turn, the relative adaptiveness of the l-th codon of the genome (w_l) is defined as follows. If this codon is the j-th codon of a codon family for the amino acid i, its relative adaptiveness is its number of occurrences N_{ij} divided by the number of occurrences N_{imax} of the most frequently used codon (within its codon family) for the amino acid i. All the codons that enter the geometric mean are obviously used at least one time, so that $N_{ij} > 0$ in our case. Finally, all codons in the standard mitochondrial genetic code table have at least one synonym, therefore no codon should be subtracted from L.

(xii) Tolopogy of the molecule. This is set to 1 for linear chromosome and 0 for circular chromosome. When the mitochondrial genome is split into several chromosomes, a different topology can be handled for each of them.

(xiii) AT content of URs.

(xiv) Median length of URs.

2.3 Factor analysis

The factor analyses are carried out using the psych (Revelle 2014) package of R; plots are prepared using the ggplot2 package. The factor analyses use normalization and varimax rotation; correlation preserving is used to get factor scores; correlations are found following the Pearson method; given the possible presence of a missing value, missing data are imputed using the median. As a major improvement with respect to the original version of HERMES (Plazzi et al. 2016), the R script which is called internally by HERMES tests all the possible combinations of at least two variables from the list above, therefore there is no need for the user to look for the best-performing combinations, which is automatically selected by the software. Each factor analysis is evaluated using different goodness-of-fit tests, following the recommendations proposed by Hu and Bentler (1999): Tucker-Lewis Index (TLI; Tucker and

Lewis 1973) > 0.95; root mean square of the residuals (SRMR) < 0.08; root mean squared error of approximation (RMSEA) and relative confidence interval < 0.06; moreover, the Kaiser-Meyer-Olkin index (KMO; Kaiser 1970) is also computed. The best-performing factor analysis, as well as the source combination of variables, is automatically selected and output to the user. We define the single loading obtained for each OTU in the best-performing factor analysis the HERMES score of that OTU.

2.4 Test of the method

In our previous paper, we applied a basic HERMES approach to the mitogenomics of bivalve molluscs (Plazzi et al. 2016). In the present paper, we test the improved and updated HERMES approach to ten datasets that were selected among recent phylogenetic studies based on complete mitochondrial genomes. Datasets were selected in order to maximize the phylogenetic representativeness among bilaterians, i.e. spanning across Deuterostomia, Ecdysozoa, and Spiralia. Namely, these datasets were produced for Asteroidea (Echinodermata; Quek et al. 2019), Paridae (Aves; Li et al. 2017, 2016), Phasianidae (Aves; Wang et al. 2017), Trematoda (Platyhelminthes; Semyenova et al. 2017), Acariformes (Arachnida; Xue et al. 2017), Pseudoniphargidae (Amphipoda; Stokkan et al. 2018), Ensifera (Orthoptera; Zhou et al. 2017), Conoidea (Gastropoda; Uribe et al. 2018), Mytilidae (Bivalvia; Lee et al. 2019), and Aphroditiformia (Annelida; Zhang et al. 2018). Outgroups were selected following the phylogenetic setup of the original publication and are listed in Table 1.

3 Results

Best-performing combinations are listed in Table 1 for each dataset. As shown in Figure 1, the AMIGA index, accounting for the uniquity or sharedness of the gene arrangement of a given mtDNA, and the ML distance to the outgroup are the most commonly used variables, entering six out of ten best-performing combinations. All 'best-performing' combinations are based on five or six genomic variables, and up to seven for Paridae (Table 2); the mean total communality is 0.4980. TLI, SRMR, and RMSEA confidence interval passed the goodness-of-fit test for all the best-performing combinations, with the exception of Paridae dataset SRMR (0.0880), which is slightly greater than the selected threshold (0.08); moreover, KMO was greater than 0.6 in all cases (Table 2). Below, we briefly present HERMES results for each dataset.

3.1 Asteroidea

A single combination of parameters maximized the number of acceptable goodness-of-fit test statistics. The variable set was AMIGA + root-to-tip distance + ML distance + AT + GC skew + length (TLI=0.9818; KMO=0.7100), for a total communality of 0.6641 (Table 1 and 2). In the work of Quek et al. (2019) the complete mitochondrial genome of *Archaster typicus* is presented. Its HERMES score is similar to that of the strictly related *Acanthaster* species (*Ac. brevispinus* and *Ac. planci*; Quek et al. 2019; Mah and Blake 2012; Knott and Wray 2000); conversely, the sea star *Echinaster brasiliensis* (family Echinasteridae) stands out for its higher HERMES score among the remaining asteroideans (Figure 2A). Finally, it is worth noting that the two sea urchins included in the analysis (genus *Strongylocentrotus*) have significantly lower HERMES scores, which suggests a lower rate in mitochondrial molecular change with respect to the common ancestor of non-crinozoan echinoderms.

3.2 Paridae

Three combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=0.9820; KMO=0.7056) used seven variables: URs, root-to-tip distance, GC skew, length, CAI, UR AT content, and UR median length, for a total communality of 0.5360 (Table 1 and 2). The work of Li and colleagues (2017, 2016) focuses on the genus *Parus* and on the Tibetan ground tit *Pseudopodoces humilis*, which was recently moved to the family Paridae (James et al. 2003) and which did nest within the *Parus* clade (Li et al. 2017, 2016). The parid species *Sylviparus modestus* is consistently recovered as the sister species of this clade: indeed, the HERMES analysis uncovered a similar evolutionary pace for all the *Parus*/*Pseudopodoces* species, while evidencing a more derived status of *S. modestus* (Figure 2B). Moreover, *Ps. humilis* scores the lowest HERMES index within the *Parus*/*Pseudopodoces* clade, possibly evidencing a less derived mtDNA for this genus.

3.3 Phasianidae

Eighteen combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=1.0289; KMO=0.7742) used five variables: AMIGA, SU skew, AT skew, genes, and length, for a total communality of 0.6427 (Table 1 and 2). The HERMES score is somewhat conserved among Phasianidae, with the exception of *Gallus gallus*, which shows a slightly higher score, and of a small group of species which conversely show lower scores: *Argusianus argus*, *Rhizothera longirostris*, *Caloperdix oculeus*, *Lerwa lerwa*, *Ithaginis cruentus*, and *Tetrastes bonasia* (Figure 2C). Interestingly, three of these species are either sister group or basal within the so-called “erectile” clade (Kimball and Brown 2008): namely, *Ithaginis* and *Lerwa* comprise a monophyletic clade, which is the sister group of the “erectile” clade (Wang et al. 2017; Meiklejohn et al. 2014; Shen et al. 2014; Wang et al. 2013). In fact, the two mtDNAs analyzed for *A. argus* showed two very different HERMES scores; the same holds for *R. longirostris*. While Wang and colleagues (2017) notice the

possibility of cryptic species diversity in this group of Galliformes, it has to be noted that a handful of mtDNAs used in their study, including those of *A. argus* and *R. longirostris*, have not been completely sequenced and annotated, which leads to misleading results about genomic variables like the number of annotated genes or the percentage of URs.

3.4 Trematoda

A single combination of parameters maximized the number of acceptable goodness-of-fit test statistics. The variable set was AMIGA + root-to-tip distance + ML distance + AT skew + genes + UR AT content (TLI=1.0906; KMO=0.6925), for a total communality of 0.6839 (Table 1 and 2). The resulting HERMES analysis evidences four groups of species (Figure 2D): the two *Fasciola* species exhibit the lowest scores, followed by the two *Trichobilharzia* species. The genus *Schistosoma* is associated to the highest HERMES values, which are in turn lower for *S. mekongi* and *S. japonicum*, and higher for *S. mansoni* and *S. haematobium*. Finally, the species *Paragonimus westermani* has a HERMES score similar to those of *Trichobilharzia*. HERMES analysis appears to be highly driven by the phylogenetic relationships of the species, included as root-to-tip and ML distances: the longer the branch in the tree presented by Semyenova et al. (2017), the higher the HERMES score. The HERMES analysis thus unveils a higher rate of genomic evolution for the mtDNA of Schistosomatidae with respect to Fasciolidae and Paragonimidae among Trematoda.

3.5 Acariformes

Twenty-seven combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=0.9873; KMO=0.8086) used six variables: AMIGA, ML distance, AT skew, length, CAI, and UR AT content, for a total communality of 0.6130 (Table 1 and 2). The HERMES index seems to precisely adhere to the systematics of

Acariformes (Figure 2E) and to the phylogenetic reconstruction proposed by Xue and colleagues (2017): family Tetranychidae (here represented by genera *Tetranychus* and *Panonychus*) show the highest scores, followed by the superfamily Eriophyoidea *sensu lato*, which was recovered as the sister group to the order Sarcoptiformes (but see Klimov et al. 2018), and the genus *Demodex* (family Democidae), which was recovered as the sister group to Tetranychidae. Finally, with the exception of *Steganacarus magnus*, Sarcoptiformes yielded higher HERMES scores than the remaining Trombidiformes.

3.6 Pseudoniphargidae

Twenty-two combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=1.0000; KMO=0.6759) used five variables: AMIGA, root-to-tip distance, GC skew, CAI, and UR median length, for a total communality of 0.4537 (Table 1 and 2). The study of Stokkan and colleagues (2018) focuses on species of the genus *Pseudoniphargus* (family Pseudoniphargidae): these consistently scored the lowest HERMES values (with the only exception of the hadziid *Bahadzia jaraguensis*; Figure 2F). Moreover, all the included isopod mtDNAs and the three metacrangonyctid amphipod mtDNAs showed higher degrees of molecular mitochondrial change with respect to remaining amphipods, thus unveiling a different evolutionary pace between isopods and amphipods in mtDNA.

3.7 Ensifera

Two combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=1.3759; KMO=0.6294) used five variables: ML distance, AT, AT skew, genes, and CAI, for a total communality of 0.2869 (Table 1 and 2). HERMES scores seem quite unrelated to systematics and/or ecology in this case (Figure 2G):

among the top five HERMES scores, however, all the representatives of the subfamily Aemodogryllinae were found (genera *Diestrammena* and *Diestramima*), which suggests a different mitochondrial evolutionary pace for this subfamily within the family Raphidophoridae.

3.8 Conoidea

Seven combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=1.2869; KMO=0.6294) used six variables: SU skew, root-to-tip distance, AT, GC skew, genes, and length, for a total communality of 0.3422 (Table 1 and 2). The seminal work of Puillandre et al. (2011) subdivided the sixteen conoidean families (Bouchet et al. 2011) into two sister clades; the same clades were recovered as strongly supported by Uribe et al. (2018). With the exception of *Glyphostoma* sp. (Clathurellidae) and *Lilliconus sagei* (Conidae), the HERMES analysis did result into two clear groups as well, yet with no apparent phylogenetic meaning (Figure 2H).

3.9 Mytilidae

A single combination of parameters maximized the number of acceptable goodness-of-fit test statistics. The variable set was URs + AMIGA + ML distance + AT + AT skew (TLI=1.3618; KMO=0.7328), for a total communality of 0.4149 (Table 1 and 2). The phylogenetic study by Lee et al. (2019) aims to unravel the evolutionary relationships between the subfamilies of the family Mytilidae. The subfamily Mytilinae scored the highest HERMES values within genera *Mytilus* and *Crenomytilus*, while the genus *Perna* showed lower amount of mitochondrial evolution (Figure 2I). With the exception of *Bathymodiolus thermophilus* and *Modiolus modiolus*, members of the subfamilies Bathymodiolinae and Modiolinae (Mytilidae Clade 1 *sensu* Liu et al. 2018) show intermediate HERMES index values; the other subfamily largely

sampled in the study of Lee et al. (2019), Brachidontinae, resulted in a low amount of mitochondrial evolution for *Mytilisepta keenae* and *M. virgata* and in a low-to-intermediate one for the genus *Brachidontes*.

3.10 Aphroditiformia

Seven combinations of variables maximized the number of acceptable goodness-of-fit test statistics. The best-performing one (TLI=1.0489; KMO=0.6031) used six variables: ML distance, AT, genes, CAI, UR AT content, and UR median length, for a total communality of 0.3425 (Table 1 and 2). The HERMES scores computed for the species of the suborder Aphroditiformia included in the study by Zhang and colleagues (2018) appear somewhat related to the phylogenetic reconstruction of the suborder, for example by recovering similar HERMES values for sigalionids polychaetes (Figure 2H). Moreover, we evidenced an increasing HERMES trend in the family Polynoidae. Specifically, lower HERMES scores were observed in *Melaenis* sp. (subfamily Polynoinae; sampled one meter below the surface), *Halosydna* sp., and *Lepidonotus* sp. (subfamily Lepidonotinae; both sampled one meter below the surface), followed by increasing scores for Lepidonotopodinae: *Levensteiniella iris* (sampled at a depth of 2,400 m) and *Lepidonotopodium okinawae* (sampled at a depth of 1,555 m), up to *Branchipolynoe longqiensis* (~2,800 m), *Branchinotogluma japonicus* (1,555 m) and *Branchipolynoe pettiboneae* (1,122 m), thus also uncovering a trend of higher amount of molecular evolution of the mtDNA for deep sea polynoid species.

4 Discussion

4.1 Case studies

The HERMES analysis allowed to immediately and straightforwardly identify phylogenetic clusters (i.e., clades) or single species with marked differences in the amount of molecular change in mtDNAs, most likely mirroring a more derived or a more ancestral status of the genome with respect to the supposed common ancestor. These can be considered as molecular synapomorphies specific of the mitochondrial genome as a whole: it is the case, for example, of the sea urchins (Fig. 2A); of *Pseudopodoces humilis* and *Sylviparus modestus* (Fig. 2B); of the genera *Fasciola*, *Trichobilharzia*, and *Schistosoma* (Fig. 2D); of the family Tetranychidae and of the superfamily Eriophyoidea (Fig. 2E); of isopods (Fig. 2F); of the subfamilies Bathymodiolinae and Modiolinae (Fig. 2I).

The HERMES analysis also allowed to identify species whose marked difference in mitochondrial evolutionary speed is most likely associated to peculiar ecological or physiological conditions. It is tempting, for example, to associate the HERMES scores of the genus *Mytilus* (which are the highest among Mytilidae; Fig. 2I) to the Doubly Uniparental Inheritance (DUI) of mtDNA, a phenomenon widespread in this genus (see, f.i., Zouros and Rodakis 2019; Gusman et al. 2016; Zouros, 2013; Passamonti and Ghiselli, 2009; Breton et al. 2007), which was previously demonstrated to be connected with selective pressures on mtDNA (Plazzi and Passamonti 2018). A further example of HERMES scores associated to a peculiar condition is given by Lepidonotopodinae annelids, that showed increasing amount of molecular evolution for deep sea species, a condition which was expectedly shown to result in selective pressures on energetic metabolic pathways and, thus, on mitochondria (see, f.i., Shen et al. 2019; Yang et al. 2019; Mu et al. 2018; Zhang et al. 2018, 2017).

In few cases, the HERMES analysis did not add anything significant to the phylogenetic analysis. This may well be expected, in that not all the mtDNA features are under direct

phylogenetic control: adaptations to specific environments and local modifications occurring in single lineages lead to fluctuations and molecular novelties. Whenever this holds for many different taxa in a given clade, we expect the HERMES signal to be blurred by such interfering contingencies.

For example, HERMES scores seem somewhat scattered around among Ensifera and the total communality is low (Tab. 2). The family Raphidophoridae globally exhibits high HERMES values, but this does not hold for *Troglophilus neglectus* (Fig. 2G). Raphidophorids are basal in the "tettigonioid" clade (Gwynne 1995; Ander 1939) and have often been regarded as somewhat "transitional" between "grylloids" and the other Ensifera (Zhou et al. 2017; Desutter-Grandcloas 2003; Ander 1939). It is possible that the results of the HERMES analysis have a meaning in this context, but it is also possible that different species underwent different patterns of mtDNA molecular evolution, so that even the best-performing combination of variables yields a very low signal, as measured by the total communality. Similarly, we detected two groups of HERMES values for conoideans (Fig. 2H), but a close look to the best combination's goodness-of-fit tests suggests that a significant discussion of the HERMES scores is not sound. In fact, these clusters of mtDNAs may associate to other features, like ecology, bathymetry, energy metabolism; however, the HERMES scores is particularly dependent upon the SU skew (which is slightly higher for taxa with higher HERMES scores) and to the number of annotated genes (which is 37 instead of 36 for taxa with higher HERMES scores): these genomic features are clearly connected to each other and this simple difference may be sufficient to result in the observed HERMES dichotomy, recall the low total communality (~34%; Tab. 2).

4.2 The application of the HERMES index

The HERMES approach, originally presented in the context of an appraisal of bivalve mitogenomics (Plazzi et al. 2016), has proved to be suitable in the wider context of bilaterians.

In the present paper, the HERMES approach is evaluated using ten case studies. In few cases, the best-performing combination did appear to only weakly describe mtDNA features: in the Aphroditiformia, Conoidea, and Ensifera datasets, the total communality is smaller than 40. Nonetheless, in the vast majority of cases, the best-performing combination yielded a loading set that robustly summarized many genomic features, up to 7 for Paridae (see Tab. 2).

It is well beyond the scope of the present paper to provide a thorough evaluation of each case study, and evolutionary and ecological implications of our results have been summarized elsewhere (see sections 3 and 4.1). Generally speaking, from the present work it is evident that HERMES contributes with further pieces of information to the available knowledge about the evolution of a focal group in many cases, unveiling molecular evolutionary dynamics that are not explicitly addressed in phylogenetic analyses, like CAI or strand usage: see, f.i., the case of Trematoda, Acariformes, Mytilidae, or Aphroditiformia; see also the case of bivalves in Plazzi et al. (2016). Therefore, we regard to HERMES as a cost-effective approach to be applied to phylogenetic reconstructions and genomic surveys, looking for significant insights on mtDNA evolution.

4.3 Future perspectives

The present version of HERMES (3.0) is well suited to analyze circular and linear, single-chromosome and multi-chromosome mitochondrial genomes; in case of multi-chromosome mitochondrial genomes, however, they will be assigned the same topology. Given the state-of-art of research on mitochondrial genome architectures, this makes the HERMES approach suitable for most eukaryotic mtDNAs. Moreover, it is also possible to provide custom-annotated genomes; however, in this case, a single chromosome is allowed and the circular topology will be assumed.

However, many different architectures and organizations are currently known for mtDNAs, and possibly more are to be discovered (see section 1). Extending the scope of the HERMES approach is possible as well. It is conceivable, for example, to insert a descriptor for the presence of introns, or, better, for their extent in terms of base pairs.

Concluding, future improvements of HERMES, which are currently under development, aim to handle custom-annotated, multi-chromosomes mtDNAs; the presence of introns; multiple genetic codes in the same dataset. This will hopefully allow to further extend the HERMES approach to all eukaryotes.

Data Availability Statement

The version 3.0 of HERMES is freely available for download from https://github.com/federicoplazzi/HERMES. The package performs all the analyses described in the present paper and directly provide the final HERMES scores.

5 Acknowledgements

We are deeply indebted to Elsa Spagnol for her invaluable help with data analysis, and to Thomas Bourguignon, Andrei Guliaev, Yuan Huang, Alexandra Lisenkova, Kieren J. Mitchell, Joong-Ki Park, Joan Pons, Nicolas Puillandre, Zheng B. R. Quek, Jian-Wen Qiu, Renfu Shao, Ning Wang, and Zhijun Zhou, who kindly accepted to share raw data from their phylogenetic studies, so that we could test the HERMES approach on many different case studies. Moreover, we wish to thank Liliana Silva, who first drove our attention to the potential of factor analysis; eventually, insightful comments of two anonymous reviewers led to several improvements of the 2.2 version, leading to the present 3.0 release. This work was supported by the “Canziani Bequest” fund (University of Bologna, grant number A.31.CANZELSEW). Declaration of interest: none.

6 References

Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA. 2003. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. Proc Natl Acad Sci USA. 100:5968-5973.

Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol. 66:4-119.

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell. 23:2499-2513.

Ander K. 1939. Vergleichend-Anatomische und Phylogenetische Studien über die Ensifera (Saltatoria). Opusc Entomol Suppl. 2:1-306.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature. 290:457-465.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. Trends Microbiol. 6:263-268.

Arafat H, Alamaru A, Gissi C, Huchon D. 2018. Extensive mitochondrial gene rearrangements in ctenophora: insights from benthic platyctenida. BMC Evol Biol. 18:65.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist. 168:39-50.

Becker T, Böttinger L, Pfanner N. 2012. Mitochondrial protein import: from transport pathways to an integrated network. Trends Biochem Sci. 37:85-91.

Bendich AJ. 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr Genet. 24:279-290.

Bernt M, Braband A, Schierwater B, Stadler PF. 2013. Genetic aspects of mitochondrial genome evolution. Mol Phylogenet Evol. 69:328-338.

Boer PH, Gray MW. 1988. Transfer RNA genes and the genetic code in *Chlamydomonas reinhardtii* mitochondria. Curr Genet. 14:583-590.

Bonen L. 2012. Evolution of mitochondrial introns in plants and photosynthetic microbes. Adv Bot Res. 63:155-186.

Boore JL. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27:1767-1780.

Borst P, Kroon A. 1969. Mitochondrial DNA: physicochemical properties, replication, and genetic function. Int Rev Cytol. 26:107-190.

Bouchet P, Kantor YI, Sysoev A, Puillandre N. 2011. A new operational classification of the Conoidea (Gastropoda). J Mollus Stud. 77:273-308.

Breton S, Doucet-Beaupré H, Stewart DT, Hoeh WR, Blier PU. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 23:465-474.

Breton S, Milani L, Ghiselli F, Guerra D, Stewart DT, Passamonti M. 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends Genet. 30:555-564.

Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. Curr Opin Microbiol. 7:528-534.

Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. Proc Natl Acad Sci USA. 100:892-897.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol Evol. 5:418-438.

Burger G, Valach M. 2018. Perfection of Eccentricity: Mitochondrial Genomes of Diplonemids. IUBMB Life. 70:1197-1206.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2019. The new tree of eukaryotes. Trends Ecol Evol. 35:43-55.

Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N. 2009. Importing mitochondrial proteins: machineries and mechanisms. Cell. 138:628-644.

Chi SI, Dahl M, Emblem Å, Johansen SD. 2019. Giant group I intron in a mitochondrial genome is removed by RNA back-splicing. BMC Mol Biol. 20:16.

Clary DO, Wolstenholme DR. 1984. The *Drosophila* mitochondrial genome. Oxf Surv Eukaryot Genes. 1:1-35.

Clary DO, Wolstenholme DR. 1985. The mitochondrial DNA molecule of Drosophila yakuba: nucleotide sequence, gene organization, and genetic code. J Mol Evol. 22:252-271.

Cohen P. 2014. New role for the mitochondrial peptide humanin: protective agent against chemotherapy-induced side effects. J Natl Cancer Inst. 106:dju006.

Denovan-Wright EM, Lee RW. 1994. Comparative structure and genomic organization of the discontinuous mitochondrial ribosomal RNA genes of *Chlamydomonas eugametos* and *Chlamydomonas reinhardtii*. J Mol Biol. 241:298-311.

Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. 2015. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci USA. 112:E693-E699.

Desutter-Grandcolas L. 2003. Phylogeny and the evolution of acoustic communication in extant Ensifera (Insecta, Orthoptera). Zoolog Scr. 6:525-561.

Dickey AM, Kumar V, Morgan JK, Jara-Cavieres A, Shatters Jr. RG, McKenzie CL, Osborne LS. 2015. A novel mitochondrial genome architecture in thrips (Insecta: Thysanoptera): extreme size asymmetry among chromosomes and possible recent control region duplication. BMC Genomics. 16:439.

Dong W-G, Song S, Guo X-G, Jin D-C, Yang Q, Barker SC, Shao R. 2014a. Fragmented mitochondrial genomes are present in both major clades of the blood-sucking lice (suborder Anoplura): evidence from two *Hoplopleura* rodent lice (family Hoplopleuridae). BMC Genomics. 15:751.

Dong W-G, Song S, Jin D-C, Guo X-G, Shao R. 2014b. Fragmented mitochondrial genomes of the rat lice, *Polyplax asiatica* and *Polyplax spinulosa*: intra-genus variation in fragmentation pattern and a possible link between the extent of fragmentation and the length of life cycle. BMC Genomics. 15:44.

Dong Y, Yoshitomi T, Hu J-F, Cui J. 2017. Long noncoding RNAs coordinate functions between mitochondria and the nucleus. Epigenet Chromatin. 10:41.

Dujon B. 2020. Mitochondrial genetics revisited. Yeast. 37:191-205.

Endo K, Noguchi Y, Ueshima R, Jacobs HT. 2005. Novel repetitive structures, deviant protein-encoding sequences and unidentified ORFs in the mitochondrial genome of the brachiopod *Lingula anatina*. J Mol Evol. 61:36-53.

Feagin JE. 1994. The extrachromosomal DNAs of apicomplexan parasites. Annu Rev Microbiol. 48:81-104.

Freel KC, Friedrich A, Schacherer J. 2015. Mitochondrial genome evolution in yeasts: an all-encompassing view. FEMS Yeast Research. 15:fov023.

Gabaldón T, Huynen MA. 2004. Shaping the mitochondrial proteome. Biochim Biophys Acta. 1659:212-220.

Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin SV, Passamonti M. 2013. Structure, Transcription, and Variability of Metazoan Mitochondrial Genome: Perspectives from an Unusual Mitochondrial Inheritance System. Genome Biol Evol. 5:1535-1554.

Gibson T, Blok VC, Dowton M. 2007a. Sequence and characterization of six mitochondrial subgenomes from *Globodera rostochiensis*: multipartite structure is conserved among close nematode relatives. J Mol Evol. 65:308-315.

Gibson T, Blok VC, Phillips MS, Hong G, Kumarasinghe D, Riley IT, Dowton M. 2007b. The mitochondrial subgenomes of the nematode *Globodera* pallida are mosaics: evidence of recombination in an animal mitochondrial genome. J Mol Evol. 64:463-471.

Gray MW, Burger G, Derelle R, Klimeš V, Leger MM, Sarrasin M, Vlček Č, Roger AJ, Eliáš M, Lang BF. 2020. The draft nuclear genome sequence and predicted mitochondrial proteome of *Andalucia godoyi*, a protist with the most gene-rich and bacteria-like mitochondrial genome. BMC Biol. 18:22.

Gray MW, Lang BF. 1998. Transcription in chloroplasts and mitochondria: a tale of two polymerases. Trends Microbiol. 6:1-3.

Gualberto JM, Newton KJ. 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 68:225-252.

Guo W, Zhu A, Fan W, Adams RP, Mower JP. 2020. Extensive Shifts from *Cis*- to *Trans*-splicing of Gymnosperm Mitochondrial Introns. Mol Biol Evol. 37:1615-1620.

Gusman A, Lecomte S, Stewart DT, Passamonti M, Breton S. 2016. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtDNA. PeerJ. 4:e2760.

Gwynne DT. 1995. Phylogeny of the Ensifera (Orthoptera): a hypothesis supporting multiple origins of acoustical signalling, complex spermatophores and maternal care in crickets, katydids, and weta. J Orth Res. 4:203-218.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. Proc Natl Acad Sci USA. 106:3859-3864.

Herd KE, Barker SC, Shao R. 2015. The mitochondrial genome of the chimpanzee louse, *Pediculus schaeffi*: insights into the process of mitochondrial genome fragmentation in the blood-sucking lice of great apes. BMC Genomics. 16:661.

Hill GE. 2019. Mitonuclear Ecology. Oxford: Oxford University Press.

Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N, Moriyama T, Ikeuchi M, Watanabe M, Wada H, Kobayashi K, Saito M, Masuda T, Sasaki-Sekimoto Y, Mashiguchi K, Awai K, Shimojima M, Masuda S, Iwai M, Nobusawa T, Narise T, Kondo S, Saito H, Sato R, Murakawa M, Ihara Y, Oshima-Yamada Y, Ohtaka K, Satoh M, Sonobe K, Ishii M, Ohtani R, Kanamori-Sato M, Honoki R, Miyazaki D, Mochizuki H, Umetsu J, Higashi K, Shibata D, Kamiya Y, Sato N, Nakamura Y, Tabata S, Ida S, Kurokawa K, Ohta H. 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. Nat Commun. 5:3978.

Hu L-T, Bentler PM. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 6:1-55.

Huerta-Cepas J, Dopazo J, Gabaldón T. 2010. ETE: a python Environment for Tree Exploration. BMC Bioinformatics. 11:24.

Ichinose M, Sugita M. 2016. RNA editing and its molecular mechanism in plant organelles. Genes. 8:5.

James HF, Ericson PGP, Slikas B, Lei F-M, Gill FB, Olson SL. 2003. *Pseudopodoces humilis*, a misclassified terrestrial tit (Paridae) of the Tibetan Plateau: evolutionary consequences of shifting adaptive zones. Ibis. 145:185-202.

Jensen RE, Englund PT. 2012. Network news: the replication of kinetoplast DNA. Annu Rev Microbiol. 66:473-491.

Jiang H, Barker SC, Shao R. 2013. Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. Genome Biol Evol. 5:1298-1308.

Kaiser HF. 1970. A second generation little jiffy. Psychometrika. 35:401-415.

Kannan S, Rogozin IB, Koonin EV. 2014. MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. BMC Evol Biol. 14:237.

Kaur B, Záhonová K, Valach M, Faktorová D, Prokopchuk G, Burger G, Lukeš J. 2020. Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. Nucleic Acids Res. 48:2694-2708.

Khachane AN, Timmis KN, Martins dos Santos VA. 2007. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. Mol Biol Evol. 24:449-456.

Kimball RT, Braun EL. 2014. Does more sequence data improve estimates of galliform phylogeny? Analyses of a rapid radiation using a complete data matrix. PeerJ. 2:e361.

Klimov PB, OConnor BM, Chetverikov PE, Bolton SJ, Pepato AR, Mortazavi AL, Tolstikov AV, Bauchan GR, Ochoa R. 2018. Comprehensive phylogeny of acariform mites (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a long branch. Mol Phylogenet Evol. 119:105-117.

Knott KE, Wray GA. 2000. Controversy and consensus in asteroid systematics: new insights to ordinal and familial relationships. Am Zool. 40:382-392.

Kohn AB, Citarella MR, Kocot KM, Bobkova YV, Halanych KM, Moroz LL. 2012. Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. Mol Phylogenet Evol. 63:203-207.

Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. 2019. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. PLoS Genet. 15:e1008373.

Kuhle B, Chihade J, Schimmel P. 2020. Relaxed sequence constraints favor mutational freedom in idiosyncratic metazoan mitochondrial tRNAs. Nat Commun. 11:969.

Landerer E, Villegas J, Burzio VA, Oliveira L, Villota C, Lopez C, Restovic F, Martinez R, Castillo O, Burzio LO. 2011. Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. Cell Oncol (Dordrecht). 34:297-305.

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature. 387:493-497.

Larriba E, Rial E, del Mazo J. 2018. The landscape of mitochondrial small non-coding RNAs in the PGCs of male mice, spermatogonia, gametes and in zygotes. BMC Genomics. 19:634.

Lavrov DV, Pett W. 2016. Animal Mitochondrial DNA as We Do Not Know It: mt-Genome Organization and Evolution in Nonbilaterian Lineages. Genome Biol Evol. 8:2896-2913.

Lee C, Yen K, Cohen P. 2013. Humanin: a harbinger of mitochondrial-derived peptides? Trends Endocrinol Metab. 24:222-228.

Lee Y, Kwak H, Shin J, Kim S-C, Kim T, Park J-K. 2019. A mitochondrial genome phylogeny of Mytilidae (Bivalvia: Mytilida). Mol Phylogenet Evol. 139:106533.

Li X-J, Lin L-L, Cui A-M, Bai J, Wang X-Y, Xin C, Zhang Z, Yang C, Gao R-R, Huang Y, Lei F-M. 2016. Taxonomic status and phylogenetic relationship of tits based on mitogenomes and nuclear segments. Mol Phylogenet Evol. 104:14-20.

Li X-J, Lin L-L, Cui A-M, Bai J, Wang X-Y, Xin C, Zhang Z, Yang C, Gao R-R, Huang Y, Lei F-M. 2017. Data on taxonomic status and phylogenetic relationship of tits. Data in Brief. 10:390-397.

Li Y, Kocot KM, Tassia MG, Cannon JT, Bernt M, Halanych KM. 2018. Mitogenomics Reveals a Novel Genetic Code in Hemichordata. Genome Biol Evol. 11:29-40.

Liu J, Liu H, Zhang H. 2018. Phylogeny and evolutionary radiation of the marine mussels (Bivalvia: Mytilidae) based on mitochondrial and nuclear genes. Mol Phylogenet Evol. 126:233-240.

Lukeš J, Wheeler R, Jirsová D, David V, Archibald JM. 2018. Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. IUBMB Life. 70:1267-1274.

Mah CL, Blake DB. 2012. Global diversity and phylogeny of the Asteroidea (Echinodermata). PLoS One. 7:e35644.

Maslov DA, Yasuhira S, Simpson L. 1999. Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. Protist. 150:33-42.

McCauley DE. 2013. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytologist. 200:966-977.

Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2014. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. Mol Phylogenet Evol. 78:314-323.

Meisinger C, Sickmann A, Pfanner N. 2008. The mitochondrial proteome: from inventory to function. Cell. 134:22-24.

Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood A-MJ, Haugen E, Bracken CP, Rackham O, Stamatoyannopoulos JA, Filipovska A, Mattick JS. 2011. The human mitochondrial transcriptome. Cell. 146:645-658.

Michaelis G, Vahrenholz C, Pratje E. 1990. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the gene for apocytochrome b and the complete functional map of the 15.8 kb DNA. Mol Gen Genet. 223:211-216.

Mower JP. Sloan DB, Alverson AJ 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ (eds). Plant genome diversity volume 1: plant genomes, their residents, and their evolutionary dynamics. Vienna: Springer. pp. 123-144.

Mu W, Liu J, Zhang H. 2018. The first complete mitochondrial genome of the Mariana Trench *Freyastera benthophila* (Asteroidea: Brisingida: Brisingidae) allows insights into the deep-sea adaptive evolution of Brisingida. Ecol Evol. 8:10673-10686.

Muthye V, Lavrov DV. 2020. Causes and consequences of mitochondrial proteome size variation in animals, Mitochondrion. 52:100-107.

Osigus H-J, Eitel M, Bernt M, Donath A, Schierwater B. 2013. Mitogenomics at the base of Metazoa. Mol Phylogenet Evol. 69:339-351.

Osigus H-J, Eitel M, Schierwater B. 2017. Deep RNA sequencing reveals the smallest known mitochondrial micro exon in animals: The placozoan *cox1* single base pair exon. PLoS One. 12:e0177959.

Palmer JD, Shields CR. 1984. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature. 307:437-440.

Passamonti M, Calderone M, Delpero M, Plazzi F. 2020. Clues of *in vivo* nuclear gene regulation by mitochondrial short non-coding RNAs. Sci Rep. 10:8219.

Passamonti M, Ghiselli F. 2009. Doubly Uniparental Inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA Cell Biol. 28:79-89.

Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD, Lavrov DV. 2011. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insight from mtDNA and the nuclear genome. Mitochondrial DNA. 22:130-142.

Plazzi F, Passamonti M. 2019. Footprints of unconventional mitochondrial inheritance in bivalve phylogeny: Signatures of positive selection on clades with doubly uniparental inheritance. J Zool Syst Evol Res. 57:258-271.

Plazzi F, Puccio G, Passamonti M. 2016. Comparative Large-Scale Mitogenomics Evidences Clade-Specific Evolutionary Trends in Mitochondrial DNAs of Bivalvia. Genome Biol Evol. 8:2544-2564.

Plazzi F, Ribani A, Passamonti M. 2013. The complete mitochondrial genome of Solemya velum (Mollusca: Bivalvia) and its relationships with Conchifera. BMC Genomics. 14:409.

Pohjoismäki JLO, Goffart S. 2011. Of circles, forks and humanity: topological organisation and replication of mammalian mitochondrial DNA. BioEssays. 33:290-299.

Pozzi A, Plazzi F, Milani L, Ghiselli F, Passamonti M. 2017. SmithRNAs: could mitochondria "bend" nuclear regulation? Mol Biol Evol. 34:1960-1973.

Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Börner T. 2010. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. Plant J. 64:948-959.

Puillandre N, Kantor YI, Sysoev A, Couloux A, Meyer C, Rawlings T, Todd J, Bouchet P. 2011. The dragon tamed? A molecular phylogeny of the Conoidea (Gastropoda). J Mollus Stud. 77:259-272.

Quek ZBR, Chang JJM, Ip YCA, Huang D. 2019. Complete mitochondrial genome of the sea star *Archaster typicus* (Asteroidea: Archasteridae). Mitochondrial DNA part B. 4:3130-3132.

R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rackham O, Shearwood A-MJ, Mercer TR, Davies SMK, Mattick JS, Filipovska A. 2011. Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. RNA 17:2085-2093.

Rehkopf DH, Gillespie DE, Harrell MI, Feagin JE. 2000. Transcriptional mapping and RNA processing of the *Plasmodium falciparum* mitochondrial mRNAs. Mol Biochem Parasitol. 105:91-103.

Repetti SI, Jackson CJ, Judd LM, Wick RR, Holt KE, Verbruggen H. 2020. The inflated mitochondrial genomes of siphonous green algae reflect processes driving expansion of noncoding DNA and proliferation of introns. PeerJ. 8:e8273.

Revelle W. 2014. psych: Procedures for Personality and Psychological Research. Evanston: Northwestern University. R package version 1.4.12.

Reyes A, Gissi C, Pesole G, Saccone C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol. 15:957-966.

Richter S, Schwarz F, Hering L, Boggemann M, Bleidorn C. 2015. The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). Genome Biol Evol. 7:3443-3462.

Riggs CL, Summers A, Warren DE, Nilsson GE, Lefevre S, Dowd WW, Milton S, Podrabsky JE. 2018. Small non-coding RNA expression and vertebrate anoxia tolerance. Front Genet. 9:230.

Ro S, Ma H-Y, Park C, Ortogero N, Song R, Hennig GW, Zheng H, Lin Y-M, Moro L, Hsieh J-T, Yan W. 2013. The mitochondrial genome encodes abundant small noncoding RNAs. Cell Res. 23:759-774.

Saccone C, Gissi C, Reyes A, Larizza A, Sbisà E, Pesole G. 2002. Mitochondrial DNA in metazoa: degree of freedom in a frozen event. Gene. 286:3-12.

Sana S, Hardouin EA, Paley R, Zhang T, Andreou D. 2020. The complete mitochondrial genome of a parasite at the animal-fungal boundary. Parasites Vectors. 13:81.

Schultz DT, Eizenga JM, Corbett-Detig RB, Francis WR, Christianson LM, Haddock SHD. 2020. Conserved novel ORFs in the mitochondrial genome of the ctenophore *Beroe forskalii*. PeerJ. 8:e8356.

Semyenova S, Chrisanfova G, Mozharovskaya L, Guliaev A, Ryskov A. 2017. The complete mitochondrial genome of the causative agent of the human cercarial dermatitis, the visceral bird schistosome species *Trichobilharzia szidati* (Platyhelminthes: Trematoda: Schistosomatidae). Mitochondrial DNA Part B. 2: 469-470.

Shao R, Barker SC, Li H, Song S, Poudel S, Su Y. 2015. Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). Sci Rep. 5:17389.

Shao R, Kirkness EF, Barker SC. 2009. The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. Genome Res. 19:904-912.

Shao R, Li H, Barker SC, Song S. 2017. The mitochondrial genome of the guanaco louse, *Microthoracius praelongiceps*: insights into the ancestral mitochondrial karyotype of sucking lice (Anoplura, Insecta). Genome Biol Evol. 9:431-445.

Shao R, Zhu X-Q, Barker SC, Herd K. 2012. Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome Biol Evol. 4:1088-1101.

Sharp PM, Li WH. 1987. The Codon Adaptation Index--a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications. Nucleic Acids Res. 15:1281-1295.

Shen X, Pu Z, Chen X, Murphy RW, Shen Y. 2019. Convergent Evolution of Mitochondrial Genes in Deep-Sea Fishes. Front Genet. 10:925.

Shen YY, Dai K, Cao X, Murphy RW, Shen XJ, Zhang YP. 2014. The updated phylogenies of the Phasianidae based on combined data of nuclear and mitochondrial DNA. PLoS ONE. 9:e95786.

Shi Y, Chu Q, Wei D-D, Qiu Y-J, Shang F, Dou W, Wang J-J. 2016. The mitochondrial genome of booklouse, *Liposcelis sculptilis* (Psocoptera: Liposcelididae) and the evolutionary timescale of *Liposcelis*. Sci Rep. 6:30660.

Shutt TE, Gray MW. 2006. Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. 22:90-95.

Signorovitch A, Buss L, Dellaporta S. 2007. Comparative genomics of large mitochondria in Placozoans. PLoS Genet. 3:e13.

Sinclair JH, Stevens BJ. 1966. Circular DNA filaments from mouse mitochondria. Proc Natl Acad Sci USA. 56:508-514.

Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. Proc Natl Acad Sci USA. 112:E3515-E3524.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. PLoS Biol. 10:e1001241.

Sloan DB. 2013. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytologist. 200:978-985.

Song F, Li H, Liu G-H, Wang W, James P, Colwell DD, Tran A, Gong S, Cai W, Shao R. 2019. Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals. Syst. Biol. 68:430-440.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312-1313.

Stokkan M, Jurado-Rivera JA, Oromí P, Juan C, Jaume D, Pons J. 2018. Species delimitation and mitogenome phylogenetics in the subterranean genus *Pseudoniphargus* (Crustacea: Amphipoda). Mol Phylogenet Evol. 127:988-999.

Suyama Y, Miura, K. 1968. Size and structural variations of mitochondrial DNA. Proc Natl Acad Sci USA. 60:235-242.

Sweet AD, Johnson KP, Cameron SL. 2020. Mitochondrial genomes of *Columbicola* feather lice are highly fragmented, indicating repeated evolution of minicircle-type genomes in parasitic lice. PeerJ. 8:e8759.

Tsujimura M, Kaneko T, Sakamoto T, Kimura S, Shigyo M, Yamagishi H, Terachi T. 2019. Multichromosomal structure of the onion mitochondrial genome and a transcript analysis. Mitochondrion. 46:179-186.

Tucker LR, Lewis C. 1973. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 38:1-10.

Uribe JE, Zardoya R, Puillandre N. 2018. Phylogenetic relationships of the conoidean snails (Gastropoda: Caenogastropoda) based on mitochondrial genomes. Mol Phylogenet Evol. 127:898-906.

Vallès Y, Halanych KM, Boore JL. 2008. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS One. 3:e1488.

Van Bruggen EFJ, Borst P, Ruttenberg GJCM, Gruber M, Kroon AM. 1966. Circular mitochondrial DNA. Biochim Biophys Acta. 119:437-439.

Vendramin R, Marine JC, Leucci E. 2017. Non-coding RNAs: the dark side of nuclear-mitochondrial communication. EMBO J. 36:1123-1133.

Wang N, Hosner PA, Liang B, Braun EL, Kimball RT. 2017. Historical relationships of three enigmatic phasianid genera (Aves: Galliformes) inferred using phylogenomic and mitogenomic data. Mol Phylogenet Evol. 109:217-225.

Wang N, Kimball RT, Braun EL, Liang B, Zhang Z. 2013. Assessing phylogenetic relationships among Galliformes: a multigene phylogeny with expanded taxon sampling in Phasianidae. PLoS ONE. 8:e64312.

Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell. 25:793-803.

Wei D-D, Shao R, Yuan M-L, Dou W, Barker SC, Wang J-J. 2012. The multipartite mitochondrial genome of *Liposcelis bostrychophila*: insights into the evolution of mitochondrial genomes in bilateral animals. PLoS One. 7:e33973.

Williamson D. 2002. The curious history of yeast mitochondrial DNA. Nat Rev Genetics. 3:1-7.

Wu Z, Waneka G, Sloan DB. 2020. The Tempo and Mode of Angiosperm Mitochondrial Genome Divergence Inferred from Intraspecific Variation in *Arabidopsis thaliana*. G3 (Bethesda). 10:1077-1086.

Xia X. 2007. An Improved Implementation of Codon Adaptation Index. Evol Bioinform. 3:53-58.

Xue X-F, Dong Y, Deng W, Hong X-Y, Shao R. 2017. The phylogenetic position of eriophyoid mites (superfamily Eriophyoidea) in Acariformes inferred from the sequences of mitochondrial genomes and nuclear small subunit (18S) rRNA gene. Mol Phylogenet Evol. 109:271-282.

Yabuki A, Gyaltshen Y, Heiss AA, Fujikura K, Kim E. 2018. *Ophirina amphinema* n. gen., n. sp., a new deeply branching discobid with phylogenetic affinity to jakobids. Sci Rep. 8:16219.

Yang M, Gong L, Sui J, Li X. 2019. The complete mitochondrial genome of *Calyptogena marissinica* (Heterodonta: Veneroida: Vesicomyidae): Insight into the deep-sea adaptive evolution of vesicomyids. PLoS One. 14:e0217952.

Zhang B, Zhang, Y-H, Wang X, Zhang H-X, Lin Q. 2017. The mitochondrial genome of a sea anemone *Bolocera* sp. exhibits novel genetic structures potentially involved in adaptation to the deep-sea environment. Ecol Evol. 7:4951-4962.

Zhang Y, Sun J, Rouse GW, Wiklund H, Pleijel F, Watanabe HK, Chen C, Qian P-Y, Qiu J-W. 2018. Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). Mol Phylogenet Evol. 125:220-231.

Zhou Z, Zhao L, Liu N, Guo H, Guan B, Di J, Shi F. 2017. Towards a higher-level Ensifera phylogeny inferred from mitogenome sequences. Mol Phylogenet Evol. 108:22-33.

Žihala D, Eliáš M. 2019. Evolution and Unprecedented Variants of the Mitochondrial Genetic Code in a Lineage of Green Algae. Genome Biol Evol. 11:2992-3007.

Zouros E, Rodakis GC. 2019. Doubly Uniparental Inheritance of mtDNA: An Unappreciated Defiance of a General Rule. Adv Anat Embryol Cell Biol. 231:25-49.

Zouros E. 2013. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. Evol Biol. 40:1-31.

Table 1. Outgroups and best-performing combinations.

Dataset	Outgroup	Best-performing combination
Asteroidea	*Glyptocidaris crenularis*	AMIGA, RtoTdist, MLdist, AT, GCskew, length
Paridae	*Remiz consobrinus*	URs, RtoTdist, GCskew, length, CAI, URs_AT, URs_MedLen
Phasianidae	*Rhynchortyx cinctus*	AMIGA, SUskew, ATskew, genes, length
Trematoda	*Diphyllobothrium latum*	AMIGA, RtoTdist, MLdist, ATskew, genes, URs_AT
Acariformes	*Paratemnoides elegantus*	AMIGA, MLdist, ATskew, length, CAI, URs_AT
Pseudoniphargidae	*Neomysis orientalis*	AMIGA, RtoTdist, GCskew, CAI, URs_MedLen
Ensifera	*Locusta migratoria*	MLdist, AT, ATskew, genes, CAI
Conoidea	*Galeodea echinophora*	SUskew, RtoTdist, AT, GCskew, genes, length
Mytilidae	*Anadara sativa*	URs, AMIGA, MLdist, AT, ATskew
Aphroditiformia	*Nephtys* sp.	MLdist, AT, genes, CAI, URs_AT, URs_MedLen

Table 2. HERMES goodness-of-fit tests for the best-performing combination.

Dataset	Variables[a]	KMO	TLI	SRMR	Lower RMSEA[b]	RMSEA	Upper RMSEA[c]	Total communality
Asteroidea	6	0.7100	0.9818	0.0762	0.0000	0.0000	0.3892	0.6641
Paridae	6	0.7056	0.9820	0.0880	0.0000	0.0000	0.3107	0.5360
Phasianidae	5	0.7742	1.0289	0.0756	0.0000	0.0000	0.2824	0.6427
Trematoda	6	0.6925	1.0906	0.0756	0.0000	0.0000	0.4049	0.6839
Acariformes	6	0.8086	0.9873	0.0699	0.0000	0.0429	0.2482	0.6130
Pseudoniphargidae	5	0.6759	1.0000	0.0589	0.0000	0.0000	0.2214	0.4537
Ensifera	5	0.6294	1.3759	0.0194	0.0000	0.0000	0.0000	0.2869
Conoidea	6	0.6294	1.2869	0.0556	0.0000	0.0000	0.0000	0.3422
Mytilidae	5	0.7328	1.3618	0.0297	0.0000	0.0000	0.0000	0.4149
Aphroditiformia	6	0.7108	4.8472	0.0687	0.0000	0.0000	0.0000	0.3425

[a] Number of variables in the best-performing combination (see Tab. 1 for details).

[b] Lower bound of the 95% RMSEA confidence interval.

[c] Upper bound of the 95% RMSEA confidence interval.

Figure Captions

Figure 1. Selected genomic variables. An overall count of genomic features that entered best-performing combinations in the ten presented HERMES factor analyses.

Figure 2. HERMES index for ten case studies. A, Asteroidea; B, Paridae; C, Phasianidae; D, Trematoda; E, Acariformes; F, Pseudoniphargidae; G, Ensifera; H, Conoidea; I, Mytilidae; J, Aphroditiformia.