

MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa

Paolo D'Onorio de Meo¹, Mattia D'Antonio², Francesca Griggio³, Renato Lupi³, Massimiliano Borsani³, Giulio Pavesi³, Tiziana Castrignanò¹, Graziano Pesole^{2,4} and Carmela Gissi^{3,*}

¹CASPUR, Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, Rome,

²Dipartimento di Biochimica e Biologia Molecolare 'E. Quagliariello', Università degli Studi di Bari, Bari,

³Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Milan and

⁴Istituto di Biomembrane e Bioenergetica, Consiglio Nazionale delle Ricerche, Bari, Italy

Received August 10, 2011; Revised September 26, 2011; Accepted November 9, 2011

ABSTRACT

The MITOchondrial genome database of metaZOAns (MitoZoa) is a public resource for comparative analyses of metazoan mitochondrial genomes (mtDNA) at both the sequence and genomic organizational levels. The main characteristics of the MitoZoa database are the careful revision of mtDNA entry annotations and the possibility of retrieving gene order and non-coding region (NCR) data in appropriate formats. The MitoZoa retrieval system enables basic and complex queries at various taxonomic levels using different search menus. MitoZoa 2.0 has been enhanced in several aspects, including: a re-annotation pipeline to check the correctness of protein-coding gene predictions; a standardized annotation of introns and of precursor ORFs whose functionality is post-transcriptionally recovered by RNA editing or programmed translational frameshifting; updates of taxon-related fields and a BLAST sequence similarity search tool. Database novelties and the definition of standard mtDNA annotation rules, together with the user-friendly retrieval system and the BLAST service, make MitoZoa a valuable resource for comparative and evolutionary analyses as well as a reference database to assist in the annotation of novel mtDNA sequences. MitoZoa is freely accessible at <http://www.caspur.it/mitozoa>.

INTRODUCTION

The mitochondrial genome (mtDNA) of Metazoa is a major target of studies focused on phylogenetic reconstructions, population genetics and molecular evolution (1). Whole-genome sequencing projects of this relatively small and mostly circular molecule have been undertaken since the development of the Sanger sequencing method (2,3) and have seen an explosive increase with the establishment of next-generation sequencing technologies (4–8). To date, over 4000 entries described as complete mitochondrial genomes are collected in the EMBL nucleotide database (release 108), with about 10 000 additional entries corresponding to human mt genome variants.

The MITOchondrial genome database of metaZOAns (MitoZoa; MZ; <http://www.caspur.it/mitozoa>) is a unique resource that provides manually curated data on gene annotation, gene order, gene content and non-coding regions (NCR) of complete and nearly-complete (≥ 7 kb) mtDNA entries of all available metazoan species. One representative entry is present for those metazoan species/subspecies for which the mtDNA has been sequenced in several individuals (9).

Most mtDNA databases focus only on metazoan subgroups. For example, AMiGA collects only arthropod mtDNA sequences (10); MamMiBase focuses on mammals (11); HmtDB and Human mtDB on human (12,13); MitoFish on fishes (<http://mitofish.aori.u-tokyo.ac.jp/>). Only the no longer updated OGRE (14) and the currently non-functional Mitome (15) databases collected complete mtDNAs of all metazoans. In addition, the NCBI Organelle Genome Resource (16,17) and GOBASE (18) databases contain all mitochondrial and

*To whom correspondence should be addressed. Tel: +39 02 50314918; Fax: +39 02 50314912; Email: carmela.gissi@unimi.it

chloroplastic genomes from all taxonomic groups. However, GOBASE and the Organelle Resource do not attempt to address, or fail in the correction of the large number of misannotations present in mtDNA entries (1,9,14,19). On the contrary, MitoZoa collects sequences from all metazoan species, and systematically identifies and resolves gene misannotations. It also offers several additional types of information and search options absent in other available mtDNA databases (9). Indeed, an associative retrieval system provides a set of tools to carry out basic and complex queries. Thus, MitoZoa users can easily retrieve gene order, NCR sequences, NCR location data, gene/genome sequences, reannotation information and other mito-genomic characteristics, for a given metazoan taxon or for congeneric species.

MitoZoa has already proved to be a useful tool for the scientific community, particularly for studies using mtDNA as a phylogenetic marker (20–23), but also for molecular evolutionary (24,25) and evolutionary ecology analyses (26) including studies on the parallel evolution of minimal mt rRNA secondary structures in metazoans, and on the development of software for environmental metagenomics analyses.

MitoZoa presents several innovative features compared to other mtDNA databases, including a user-friendly retrieval system with one general and three specialized search menus (9). Innovative features of MitoZoa, already described in (9), include:

- (1) Extensive controls and correction of gene annotations using a mtDNA-specific re-annotation pipeline.
- (2) Standard messages and new entry fields, unambiguously reporting all modifications and data enrichments of the original entries, and making these changes easily searchable by MitoZoa users. The 'MitoZoa Reannotation Summary' (MRS) is one of the main novelties of the EMBL-like MitoZoa entry format.
- (3) NCRs of any size are annotated under the new 'NCR' FTkey, thus they can be retrieved with the specialized 'NCR Menu' using several selection criteria.
- (4) Gene names are standardized using hidden aliases, thus all sequences of a given gene can be simply retrieved using the 'Gene Content Menu'.
- (5) The mtDNA gene order is stored as a string of standardized gene names using a FASTA-like format. Thus, entries sharing a given gene order can be retrieved with the 'Gene Order Menu'.
- (6) mtDNAs of congeneric species can be easily selected by the 'General Search Menu', thanks to the creation of the new 'ConGeneric' field.

Several new features have been introduced in MitoZoa 2.0, including: (i) the implementation of a sequence similarity search service by BLAST; (ii) the improvement of the gene re-annotation strategy and of the related pipeline; (iii) the inspection of protein-coding genes; (iv) the systematic and standardized annotation of introns and 'precursor ORFs' post-transcriptionally

restored by RNA editing or programmed translational frameshifting (PTF) (27,28); and (v) updating of entries.

NEW FEATURES IN MITOZOA 2.0

BLAST service

The MitoZoa web resource now includes a dedicated BLAST page. The BLAST service allows sequence similarity searches not only against the MitoZoa database (i.e. the full 'mtDNA' sequence of each MitoZoa entry) but also against five additional MitoZoa-derived data sets (Table 1). Each of these additional data sets contains functionally homogeneous mitogenomic 'sub-sequences', such as NCRs or gene categories. Moreover, each sequence of these five additional data sets is described in the header by the entry Accession number, the species name and also the MitoZoa-defined standardized gene name or NCR code (Table 1). These gene names/NCR codes will greatly help the use of BLAST results for annotation of newly produced mt sequences, and for re-annotation of existing mtDNA sequences.

It should be emphasized that all BLAST data sets derived from MitoZoa are automatically updated in concert with MitoZoa. As an example, Table 1 reports the size of the BLAST data sets built from MitoZoa release 9.1. The BLAST service uses the most recent version (2.2.25) of the BLAST+ package (29,30).

Quality checks of protein-coding gene annotation

Unlike the previous MitoZoa reannotation pipeline (9), MitoZoa 2.0 now includes specific checks that verify the correctness of protein-coding gene (CDS) annotations. As a result, possible CDS name errors are fixed and CDS boundaries are also significantly improved.

The quality check pipeline involves both automatic and manual steps, described in detail in [Supplementary Data](#). In particular, examination of CDS multi-alignments allows the detection of two types of CDS inconsistencies resolved in MitoZoa in the following ways:

- Modification of the CDS boundaries: by shifting the annotated start/stop codon, we can recover highly conserved N/C-terminal protein regions identified in the CDS multi-alignment of a given large taxon. Similarly, we can also eliminate extra N/C-terminal protein regions not present in all other multi-aligned CDS. Thus, the encoded protein is accordingly lengthened or shortened.
- Warning message on 'loss of highly conserved aminoacidic regions(s) that can be recovered by frame-shift(s)': highly conserved protein region(s) identified in certain multi-alignments are lost in some CDS but can be easily recovered by CDS frameshift(s). Most of such CDS frameshifts are likely due to inaccurate sequencing, as they are located close to sequencing error hot spots (i.e. long homopolymers >8 nt). However, other frameshift cases cannot be easily explained and could represent real losses of functional regions. Thus, we have not modified the boundaries of these CDS but have highlighted them in the MRS

Table 1. Mitochondrial data sets searchable with BLAST, together with the data set size in MitoZoa Release 9.1

Data set name	FTkey used as data set source	Additional data to the sequence header	No. of sequences
mtDNA	Full entry	mtDNA	2894
CDS_nt	CDS	Standard gene name	37 022
tRNA	tRNA	Standard gene name	61 228
rRNA	rRNA	Standard gene name	5699
NCR ≥ 25 nt	NCR ≥ 25 nt	NCR code ^a	8761
Protein	CDS translation, excluding pseudogenes	Standard gene name	37 016

^aThe NCR code defined by MitoZoa relates to species, flanking genes and NCR length (in bp). See also the online MitoZoa Help.

Table 2. Inconsistencies of protein-coding genes (CDS) corrected or pointed out with a warning message in MitoZoa Release 9.1

CDS inconsistency	No. of CDS	No. of entries
Modification of name	2 ^a	1 ^a
Modification of strand and boundaries	2 ^b	1 ^b
Modification of boundaries	203	184
Internal stop codons resolved by adding a 'join' ^c	9 ^d	8
Unusual start codon resolved by deleting a 'join' ^c	2 ^e	2 ^e
Warning on 'loss of highly conserved regions'	107	84
MitoZoa Release 9.1	27 022	2894

^aExchanged annotation between *atp8* and *atp6* in the snake *Anilius scytale* (FJ755180, v2 EMBL entry).

^b*atp8* and *nad3* of the gastropod *Platelix mortoni* (GU475132).

^cSpecial cases of the category 'modification of boundaries'. The 'join' operator, defined by GenEMBL, is used to exclude internal positions from CDS or other FTkeys.

^dIn *nad2* of the gastropod *Ilyanassa obsoleta* (NC_007781), the addition of the 'join' operator is also accompanied by modification of the start codon position. In all remaining cases, the CDS boundary modification consists of only the addition of the 'join' operator.

^eIn both cases (DQ340844 and NC_000844), the presence of the join operator was due to the hypothesis of the existence of a four-base start codon in *cox1*, recently rejected by experimental data (32).

('MitoZoa Reannotation Summary') field using a specific warning message (see figure 1 of the online MitoZoa Help). Consequently, MitoZoa users can easily select these CDS, and are warned to pay special attention to the analyses of these CDS and their possible flanking NCRs.

Our CDS quality check strategy identified a total of 207 CDSs that need 'modifications of name/boundaries', and 107 CDS that invoke a warning on the 'loss of highly conserved aminoacidic regions' (Table 2). We emphasize that most CDS modifications and warning notes cause the disappearance of flanking NCRs or gene overlaps. In addition, 4 CDS errors have effects on the determination of gene order ('gene name' and 'gene strand' modifications in Table 2). Finally, 9 CDSs were likely incorrect because they showed multiple internal stop codons (Table 2). Therefore, the CDS re-annotation process has significant consequences on the CDSs themselves (and their use in phylogenetic reconstruction), the determination of flanking NCRs, and even on the overall gene order.

As a final point, we would emphasize that CDS re-annotation has required the definition of specific criteria for mt CDS determination based on the

peculiarities of the mt transcriptional and maturation processes (31–33). These criteria can be also regarded as tentative rules for the standardization of mt CDS annotation and are detailed in the [Supplementary Data](#).

Standardized annotation of introns and frameshifts

Group I and II self-splicing introns as well as frameshift sites post-transcriptionally resolved by RNA editing or programmed translational frameshifting (PTF) (27,28,34) occur in some protein-coding genes of few metazoan taxa. However, original entries often contain non-standard annotations of these phenomena, rendering automated parsing difficult. In MitoZoa 2.0, we have implemented a specific pipeline, detailed in the [Supplementary Data](#), to identify and standardize such annotations.

These CDS peculiarities are now clearly recorded in the MRS field with appropriate standardized messages (see figure 1 of the Online MitoZoa Help), thus they can be easily retrieved by MitoZoa users. Moreover, we have created a new FTkey 'prec_ORF' in order to annotate all 'precursor ORFs' with frameshift site(s) corrected by RNA editing or PTF. This new FTkey allows the automatic retrieval and analysis of these 'precursor ORF' sequences. As discussed in the [Supplementary Data](#), we have used the 'prec_ORF' annotation to study the reliability of the currently hypothesised RNA editing/PTF cases. Thus, we are confident that this MitoZoa novelty will help the correct annotation of future cases of RNA editing/PTF.

In the current MitoZoa release, we have identified and annotated 40 CDS with introns and 198 CDS with frameshift sites (see [Supplementary Tables S1–S3](#)).

MitoZoa format novelties

For each MitoZoa entry, the gene order is reported in a FASTA-like format as a string of standardized gene names (9). In MitoZoa 2.0, the gene order format has been improved adding to the header a token that indicates the linear topology (L) or the partial status (P) of the entry. This novelty helps to identify linear and partial mtDNAs from the inspection of gene order header. It can be advantageous to users interested in extensive analyses of the gene order in large taxonomic groups.

MitoZoa entry updates

Pre-existing MZ entries are now updated at each new MZ release. This update is essential to allow reliable entry

selections with the Taxonomy, the Organism Species (OS) and the ConGeneric (CG) fields of the 'General Search Menu'.

In particular, the update of the Taxonomy field is indispensable because it comes from the Taxonomy database (<http://www.ncbi.nlm.nih.gov/taxonomy>), where even high taxonomic levels are frequently reorganized by NCBI curators. Furthermore, the OS field of existing entries are sometimes modified by the authors of entries owing to revised taxonomic assignment of the biological sample used for sequence production. Specific standardized messages are added to the MRS field to track these changes and allow easily retrieval (see figure 1 of the online MitoZoa Help).

As an example of the extent of MZ entry update, the migration of the 2633 pre-existing entries from MitoZoa Rel. 7 to Rel. 8 involved changes of 300 entries (11.4%) in the OC field, and 65 entries (2.5%) in the OS field (plus OC, if necessary).

Miscellanea

The MZ re-annotation pipeline includes some completely manual steps involving literature check, evaluation of unusual mtDNA characteristics, and *de novo* annotation of interesting entries. All these steps depend on curator expertise and are time-consuming. Thus, we have set up specific file formats and scripts to assist curators. Some examples of manually revised entries are reported in [Supplementary Table S4](#).

The previous MitoZoa list of the mt genetic codes has been updated adding a new genetic code absent in the translation table list compiled by the NCBI (<http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>). This code, named '5bis', has been recently identified in the nematode *Radopholus similis* by Jacob *et al.* (35).

SUMMARY AND FUTURE DIRECTIONS

MitoZoa provides carefully revised annotations of all mt gene categories, thus it ensures high accuracy of gene sequences, NCRs and gene order data extracted from MitoZoa. Moreover, all corrections and improvements of the entries are indicated by standardized messages (mainly located in the MRS field), further assisting MitoZoa users in the analysis of the revised elements.

The MitoZoa retrieval system permits the easy selection both of highly studied mt protein-coding genes and some often overlooked mt features such as NCR sequences and gene order, even for large taxonomic data sets. Among these features, NCR sequences and gene order data are difficult or impossible to retrieve from other mt databases. Indeed, MitoZoa permits flexible queries not feasible by any other system. For example, the selection of the teleost L-strand replication origin sequences can be achieved through the 'NCR Menu' searching for all NCRs longer than 20 bp, located between *trnN* and *trnC*, and belonging to the taxon Teleostei. Likewise, all metazoan mtDNAs having the mammalian-distinctive 'WANCY' region can be simply extracted through the 'Gene Order Menu'

searching for entries having the '*trnW* -*trnA* -*trnN* -*trnC* -*trnY*' gene string.

We believe that both the correction of annotation inconsistencies and the user-friendly retrieval system makes MitoZoa a valuable resource for researchers interested in phylogenetic reconstructions and also in peculiar aspects of mtDNA evolution. MitoZoa could also direct the mitochondrial community to new investigations, thanks to the emphasis on taxa/genes characterized by problematic annotations or unusual features. Finally, the implementation of the BLAST sequence similarity search could make MitoZoa a reference database for the annotation of novel mt genomes, and the definition of widely shared mt annotation rules whose requirement has been often invoked in the past (19). Indeed, as stressed in the section on CDS quality check, the correction of gene boundaries requires the definition of general annotation rules based on the knowledge of the mt transcription and translation processes.

In the future, we plan to develop new tools for the examination of gene order and to implement services for the analyses of retrieved sequences (programs for sequence multi-alignment, prediction of secondary structures, etc). Suggestions from MitoZoa users on new options for data visualization and extraction will be also taken into account.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online: [Supplementary Tables S1–S4](#).

FUNDING

Ministero dell'Istruzione, dell'Università e della Ricerca, Italy: Programma di Ricerca Scientifica di Rilevante Interesse Nazionale 2009; Progetto DM19410. Università degli Studi di Milano, Italy: Programma dell'Università per la Ricerca FIRST 2007. Consiglio Nazionale delle Ricerche, Italy: LifeWatch. Funding for open access charge: Università degli Studi di Milano, Italy: Programma dell'Università per la Ricerca (to C.G.).

Conflict of interest statement. None declared.

REFERENCES

1. Gissi, C., Iannelli, F. and Pesole, G. (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, **101**, 301–320.
2. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F. *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.
3. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, **26**, 167–180.
4. Jex, A.R., Hall, R.S., Littlewood, D.T. and Gasser, R.B. (2010) An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Res.*, **38**, 522–533.
5. Jex, A.R., Littlewood, D.T. and Gasser, R.B. (2010) Toward next-generation sequencing of mitochondrial genomes—focus

- on parasitic worms of animals and biotechnological implications. *Biotechnol Adv.*, **28**, 151–159.
6. McComish, B.J., Hills, S.F., Biggs, P.J. and Penny, D. (2010) Index-free de novo assembly and deconvolution of mixed mitochondrial genomes. *Genome Biol. Evol.*, **2**, 410–424.
7. Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L. *et al.* (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.*, **20**, 908–916.
8. Timmermans, M.J., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T., Pons, J. and Vogler, A.P. (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Res.*, **38**, e197.
9. Lupi, R., D'Onorio de Meo, P., Picardi, E., D'Antonio, M., Paoletti, D., Castrignano, T., Pesole, G. and Gissi, C. (2010) MitoZoa: a curated mitochondrial genome database of metazoans for comparative genomics studies. *Mitochondrion*, **10**, 192–199.
10. Feijao, P.C., Neiva, L.S., de Azeredo-Espin, A.M. and Lessinger, A.C. (2006) AMiGA: the arthropodan mitochondrial genomes accessible database. *Bioinformatics.*, **22**, 902–903.
11. Vasconcelos, A.T., Guimaraes, A.C., Castelletti, C.H., Caruso, C.S., Ribeiro, C., Yokaichiya, F., Armoa, G.R., Pereira, G.A.S., da Silva, I.T., Schrago, C.G. *et al.* (2005) MamMiBase: a mitochondrial genome database for mammalian phylogenetic studies. *Bioinformatics.*, **21**, 2566–2567.
12. Attimonelli, M., Accetturo, M., Santamaria, M., Lascaro, D., Scioscia, G., Pappada, G., Russo, L., Zanchetta, L. and Tommaso-Ponzetta, M. (2005) HmtDB, a human mitochondrial genomic resource based on variability studies supporting population genetics and biomedical research. *BMC Bioinformatics.*, **6**, S4.
13. Ingman, M. and Gyllenstein, U. (2006) mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res.*, **34**, D749–751.
14. Jameson, D., Gibson, A.P., Hudelot, C. and Higgs, P.G. (2003) OGRE: a relational database for comparative analysis of mitochondrial genomes. *Nucleic Acids Res.*, **31**, 202–206.
15. Lee, Y.S., Oh, J., Kim, Y.U., Kim, N., Yang, S. and Hwang, U.W. (2008) Mitome: dynamic and interactive database for comparative mitochondrial genomics in metazoan animals. *Nucleic Acids Res.*, **36**, D938–D942.
16. Wolfsberg, T.G., Schafer, S., Tatusov, R.L. and Tatusov, T.A. (2001) Organelle genome resource at NCBI. *Trends Biochem Sci.*, **26**, 199–203.
17. Pruitt, K.D., Tatusova, T., Klimke, W. and Maglott, D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
18. O'Brien, E.A., Zhang, Y., Wang, E., Marie, V., Badejoko, W., Lang, B.F. and Burger, G. (2009) GOBASE: an organelle genome database. *Nucleic Acids Res.*, **37**, D946–D950.
19. Boore, J. (2006) Requirements and standards for organelle genome databases. *OMICS*, **10**, 119–126.
20. Irisarri, I., San Mauro, D., Green, D.M. and Zardoya, R. (2010) The complete mitochondrial genome of the relict frog *Leiopelma archeyi*: insights into the root of the frog Tree of Life. *Mitochondrial DNA*, **21**, 173–182.
21. Irisarri, I., Vences, M., San Mauro, D., Glaw, F. and Zardoya, R. (2011) Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. *BMC Evolutionary Biol.*, **11**, 114.
22. Jia, W.Z., Yan, H.B., Guo, A.J., Zhu, X.Q., Wang, Y.C., Shi, W.G., Chen, H.T., Zhan, F., Zhang, S.H., Fu, B.Q. *et al.* (2011) Complete mitochondrial genomes of *Taenia multiceps*, *T. hydatigena* and *T. pisiformis*: additional molecular markers for a tapeworm genus of human and animal health significance. *BMC Genomics*, **11**, 447.
23. Rawlings, T.A., MacInnis, M.J., Bieler, R., Boore, J.L. and Collins, T.M. (2010) Sessile snails, dynamic genomes: gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. *BMC Genomics*, **11**, 440.
24. Castellana, S., Vicario, S. and Saccone, C. (2011) Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. *Genome Biol. Evol.*, **3**, 1067–1079.
25. Klimov, P.B. and Knowles, L.L. (2011) Repeated parallel evolution of minimal rRNAs revealed from detailed comparative analysis. *J. Heredity*, **102**, 283–293.
26. Bengtsson, J., Eriksson, K.M., Hartmann, M., Wang, Z., Shenoy, B.D., Grelet, G.A., Abarenkov, K., Petri, A., Alm Rosenblad, M. and Nilsson, R.H. (2011) Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie Van Leeuwenhoek*, **100**, 471–475.
27. Russell, R.D. and Beckenbach, A.T. (2008) Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. *J. Mol. Evol.*, **67**, 682–695.
28. Rosengarten, R.D., Sperling, E.A., Moreno, M.A., Leys, S.P. and Dellaporta, S.L. (2008) The mitochondrial genome of the hexactinellid sponge *Aphrocactites vastus*: evidence for programmed translational frameshifting. *BMC Genomics.*, **9**, 33.
29. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
30. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics.*, **10**, 421.
31. Montoya, J., Lopez-Perez, M.J. and Ruiz-Pesini, E. (2006) Mitochondrial DNA transcription and diseases: past, present and future. *Biochim Biophys Acta.*, **1757**, 1179–1189.
32. Stewart, J.B. and Beckenbach, A.T. (2009) Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. *Gene.*, **445**, 49–57.
33. Temperley, R.J., Wydro, M., Lightowlers, R.N. and Chrzanowska-Lightowlers, Z.M. (2010) Human mitochondrial mRNAs—like members of all families, similar but different. *Biochim Biophys Acta.*, **1797**, 1081–1085.
34. Mindell, D.P., Sorenson, M.D. and Dimcheff, D.E. (1998) An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol. Biol. Evol.*, **15**, 1568–1571.
35. Jacob, J.E., Vanholme, B., Van Leeuwen, T. and Gheysen, G. (2009) A unique genetic code change in the mitochondrial genome of the parasitic nematode *Radopholus similis*. *BMC Res Notes.*, **2**, 192.