# Vertebrate Mitochondrial Genome Organization: Evidence of Convergent Evolution

**Maria Paula Montaña Lozano**
Universidad del Tolima
**Manuela Alejandra MorenoCarmona**
Universidad del Tolima
**Jesus Mauricio Ochoa Capera**
Universidad del Tolima
**Natalia Sofía Medina Camacho**
Universidad del Tolima
**Jeffrey L. Boore**
Providence St. Joseph Health and Institute for Systems Biology
**Carlos Fernando Prada Quiroga** ( cfpradaq@ut.edu.co )
Universidad del Tolima

**Research Article**

**Keywords:** Comparative genomics, Bioinformatics analyses, Mitochondrial genomes, Gene order

**Posted Date:** April 29th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-456479/v1

# Vertebrate Mitochondrial genome organization: Evidence of convergent evolution

**Running head: Vertebrate Mitochondrial genome evolution**

Maria Paula Montaña Lozano[1], Manuela Alejandra Moreno Carmona[1], Jesus Mauricio Ochoa Capera[1], Natalia Sofía Medina Camacho[1], Jeffrey L. Boore[2] and Carlos Fernando Prada Quiroga[1,*]

[1]Grupo de investigación de Biología y ecología de artrópodos. Facultad de Ciencias. Universidad del Tolima. Colombia.

[2]Providence St. Joseph Health and Institute for Systems Biology, 401 Terry Avenue N, Seattle, WA 98109, USA.

*Correspondence: cfpradaq@ut.edu.co

## Abstract

The evolution of the vertebrate mitochondrial genome has been the focus of numerous genetic and evolutionary studies over the last several decades. Initially, sampling was heavily biased toward taxonomic orders of greatest economic or health importance, but recent advances in DNA sequencing technology have facilitated a much broader phylogenetic sampling from which we can clarify general evolutionary trends such as patterns of gene rearrangement. Toward this end, we performed a comparative genomic analysis of the 2,831 vertebrate mitochondrial genomes representing 12 classes that are available in the NCBI

database. Using a combination of bioinformatics methods, we determined that there is a great variation in the proportion of rearrangement by gene and by taxonomic class, with higher rates being observed in Reptilia, Amphibia, Petromyzonti, Mammalia, and Actinopteri. Further, within each class, there is large variation in proportion of reorganization among different orders or even taxonomic families. Eleven events of convergence in the genic order among different taxonomic orders were determined, most of them not previously reported.

**Keywords**

Comparative genomics, Bioinformatics analyses, Mitochondrial genomes, Gene order

## Introduction

Vertebrate mitochondrial genomes (mtDNAs) are circular, typically 14–20 kbp, and contain genes for 13 proteins (*atp6, atp8, cob, cox1–3, nad1–6, nad4L*), 2 ribosomal RNA (*rRNAs; rrnS, rrnL*), and 22 transfers RNA (tRNAs; *trnX*, where X is the one letter code for the corresponding amino acid [1-3].

Gene rearrangement is one of the most studied features for animal mitochondrial genomes [4]. In vertebrates, the order of genes on the mitochondrial genomes is generally considered to be conserved [5]. However, it has been reported some rearrangement include gene transposition, gene loss and gene duplication [6]. These events have often been modeled by a process of tandem duplication followed by random gene losses (TDRL) which is the most frequently invoked model to explain the diversity of gene rearrangements in metazoan mitogenomes [6-8]. Due to advances in DNA sequencing technology, it has become possible to compare

hundreds of vertebrate mitochondrial genomes from various taxonomic lineages, showing that gene order can vary far more than previously recognized [2,9,10]. Although the ancestral vertebrate mitochondrial gene arrangement is mostly found in vertebrates [2], some rearrangements have long been noted among birds, some species of lizards, crocodilians, marsupial mammals, snakes, tuatara, lamprey, amphibians, and some species of fish; most of these rearrangements involve genes flanking one or both of the two origins of replication, sites where gene duplications that have been proposed to mediate translocations may be especially common [2,8,10-14].

Differences in size in some vertebrate mitochondrial genomes are due to deletion or duplication events, mainly in the D-loop region and/or tRNAs, and as an artifact in the databases where genomes reported as complete are actually missing some portion, often parts of the D-loop region [3,12]. Rearrangements of mitochondrial genes can have profound functional implications on gene expression and genome replication, can be correlated with genomic variation, aspects of physiology, molecular mechanism, life history, or genomic evolutionary processes [1,15].

The aim of this study is to systematically compare the rate of mitochondrial gene rearrangements in each of the taxonomic orders of vertebrates and to identify possible convergence events that occurred in the evolution of this important animal group.

## Methods

*Mitochondrial Genome Sequences of Vertebrate Species*

We retrieved the sequences and gene annotations of the 2,831 complete vertebrate mitochondrial genomes, representing 143 taxonomic orders organized into 12 taxonomic classes (Myxini, Petromyzonti, Elasmobranchii, Holocephali, Cladistii, Actinopteri, Coelacanthi, Dipneusti, Amphibia, Reptilia, Aves, and Mammalia) that are available at the organelle genome resources database from NCBI (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/ ) as of December 20, 2019. Mitogenomes representing strains within the same species were not included (as in the case of the mouse, *Mus muscullus*, for which there are now mitogenome sequences for at least 20 strains). A list of these species, sorted taxonomically, with the GenBank Reference IDs and reported gene rearrangements is provided in the Supplementary Table S1.

*Gene order and rearrangements rate analysis*

We attempted to verify the correctness of reported gene annotations based on the methodology previously used by Prada and Boore [12]. The ancestral gene order of the vertebrate mitochondrial genome postulated by different authors [1,2,16], was used to identify the possible reorganizations (duplication, deletion and inversion-translocation) by observational analysis of the genes and their annotations using the program Geneious version 4.8.5 [17]. For this aim, a numerical gene order was made (from 1 to 37, considering 1 as the *trnI* and 2 as *trnQ* gene; and so on) and their gene orientation according to the position of the gene in the heavy (plus/plus as +) and light strand (plus/minus as -). For this analysis, the D-loop region was not taken into account due to the absence of its annotation or, indeed, even of the sequence (complete or partial) likely to contain this feature, even those listed as being complete, in a significant number of mt genomes.

The NCBI-BLAST2 sequence comparison program (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and MUSCLE multiple sequence alignments were used to corroborate whether the annotations were correct or not. We compared each gene that was reported as being inverted relative to the putative ancestral arrangement with the orthologous genes of closely related species in each orientation with 80% sequence identity as the threshold for determining the correct orientation. The MITOS web server (http://mitos.bioinf.uni-leipzig.de/index.py) [18] was used to corroborate the annotation, size and orientation of tRNAs and coding genes The tRNAscan-SE 2.0 program [19] was used to detect tRNA-encoding genes and confirm their orientation in the mitochondrial genomes.

The gene rearrangements proportion of each Vertebrate class and/or order, by means of the following equation (1):

$$PRo: \frac{NRo}{NE * 37} * 100$$

PRo: Proportion of rearrangements for class and/or order; NRo: Number of rearrangements for orders; NE: Number of species of each order; 37: Total mitochondrial genome genes.

The gene rearrangements proportion of each Vertebrate class for gene, by means of the following equation (2):

$$PRg: \frac{NRg}{NE} * 100$$

RRg: Proportion of rearrangements for gene; NRg: Number of rearrangements for gene; NE: Number of species of each order.

These results were compared with those obtained by the qMGR program [4]. The results generated by the analysis of gene rearrangements rate and those of qMGR of each

Vertebrate order were graphed using the Heatmapper program (http://www.heatmapper.ca/) [20]. A phylogeny was constructed from reference molecular phylogeny of the Vertebrate [21]; contrasting them with those presented in this study.

*Ancestral state estimation and phylogenetic analysis*

Several methodologies were used to estimate the ancestral status in the vertebrate mitochondrial genome. Common interval analysis was conducted using CREx [18] for pairwise comparisons and TreeREx [22] for inference of ancestral genome states. Paired CREx comparisons of the representative mitochondrial genome of some taxonomic classes or orders against the vertebrate ancestral mitochondrial genome were performed to determine the number of minimal genome rearrangement events separating each taxonomic order from the ancestral state. The CREx and TreeREx programs use the same set of rearrangement events: transpositions, inversions, reverse transpositions and TDRL, as well as the same algorithm called common interval, however, they cannot analyze gene duplications or deletions.

Mitochondrial genomes within each vertebrate taxonomic order were aligned using the MAUVE aligner software v.2.3. progressive alignment algorithm [23]. Alignments were made using the following parameters: skip-refinement and seed-weight = 15, total alignment, determining local collinear blocks (LCB) and pairs of LCBs. The option of using seed families in the anchorage and linear genomes was ignored. the Mauve phylogenetic trees were used for the of TreeRex analysis. Phylogenetic trees were visualized and edited using FigTree ver 1.4 [24].

We conducted a multiple genome rearrangement using GRIMM software (http://grimm.ucsd.edu) which uses Hannenhalli´s and Pevzner's (HP) algorithms [25] for computing uni-chromosomal and multi-chromosomal genomic distances [26]; to infer convergent and parallelism evolution on the vertebrate mitochondrial genomes.

## Results

*Gene Annotation errors in the Vertebrate mitochondrial genome*

There are a total of 104904 genes annotated in the 2831 vertebrate mitochondrial genomes found in the NCBI database, including 178 gene duplications detected by the presence of an identical copy of such gene. A total of 1951 cases are annotated to be in arrangements differing from the ancestral vertebrate mitochondrial gene order (Supplementary Table S1). Of all these reorganizations, 389 were identified as gene annotation errors, an error percentage of 20%. Of these, tRNAs have the highest level of annotations errors (94.6%) with the highest values in trnE (15.3%) and trnP (14.6%), 4.3% are errors in rRNA genes, represented mostly by errors in rrnS (3%), and 1.1% of the annotation errors in coding genes are only in the case of nad6 gene. Actinopteri, Reptilia, Aves, and Mammalia with 126, 89, 77, and 68 genes with annotation errors, respectively (Table 1).

*Rearrangement level in the Vertebrate mitochondrial genome*

In this study a total of 1562 reorganizations in the vertebrate mitochondrial genome were confirmed. Our results show that 33.2% (942 of 2,831) of the mitogenomes (present in 6 of the 12 class) have at least one rearrangement (Table 1). The present work revealed that the

153 species with the most rearrangements were in the classes Actinopteri (628), Reptilia (475)
154 and Amphibia (262), compared to other classes (Table 1). However, when calculating the
155 rearrangements rate by class, these vary from one to another; showing that the classes with
156 low rearrangement rate (>0 to 5%) are Reptilia (4.85%), Amphibia (3.27%), Petromyzonti
157 (2.70%), Mammalia (1.66%), Actinopteri (1.24%); and Aves (0.65%). The only taxonomic
158 groups with vertebrate ancestral order were Myxini, Elasmobranchii, Holocephali,
159 Cladistii, Coelacanthi and Dipneusti (Table 1 and Fig. 1).

160 Within some taxonomic classes, there are orders wherein all sampled species retain the
161 ancestral order and others with gene rearrangements. For example, in the Actinopteri class,
162 34 of the 59 taxonomic orders have all sampled species retaining the ancestral vertebrate
163 mitochondrial gene arrangement; the other orders have species with one or more
164 rearrangements. The order Saccopharyngiformes (belonging to Actinopteri class), has many
165 species (72.8%) with rearranged mtDNAs in comparison with other orders of the same
166 class as Anguilliformes (5.1%); observing gene rearrangements in only 4 analyzed
167 genomes. Within the Reptilia class, the Serpentes and Crocodilia orders have higher
168 proportion of reorganization (9.8% and 8.1%, respectively) compared to Amphisbaena
169 (4.0%), Lacertilia (2.7%) or Testudines (0.19%). Anura order (belonging to the Amphibia
170 class), presents a major proportion of reorganization (4.8%) in comparison with Caudata
171 (0.78%) or Gymnophiona (1.30%) orders (Fig. 2, Supplementary Table S2).

172 On the other hand, our results show that the most frequent rearrangements in vertebrates are
173 inversions/translocations with 87.3% (1363/1562), followed by duplications with 11.4%
174 (181/1562) and deletions with 1.3% (21/1562). However, in some classes, certain events
175 predominate as in the case of Aves, where 94.4% (68/72) of the rearrangements are

duplications and 5.6 % (4/77) are deletions (There are no confirmed inversions/translocations in this class.) In the case of Mammalia, only inversions/translocations were confirmed (Table 1). In addition, of all rearrangements, 85.3% (1332/1562) are of the tRNA genes, mainly associated with inversions/translocations (92%) (Supplementary Table S1).

Our results show differences in quantifying mitochondrial genome rearrangement proportion in each gene by taxonomic order between manual and qMGR analysis. For example within the class Actinopteri, our analyses confirm that most or all of the genes present a high proportion of gene rearrangements. In Batrachoidiformes rearrangements were found in 100% of genes whilst in Saccopharyngiformes 72.8% of genes presented rearrangements. In Saccopharyngiformes, manual analysis finds a proportion of gene rearrangements of 72.8%, while qMGR shows 37.8% (Fig. 2). In this sense, the most rearrangements in vertebrate mitochondrial genome were observed mainly in tRNA genes, concentrating in certain clusters such as *trnL, nad1, trnI, -trnQ, trnM; trnW, -trnA, -trnN, -trnC* and *trnH, trnS, trnL, nad5, -nad6, -trnE, cob, trnT, -trnP* (Figure 2). Of the 1562 genes with confirmed rearrangements, 103, 101, 145, 70, and 58 are present in the *trnL, nad1, trnI, -trnQ, trnM* region, respectively; while 55, 56, 56 and 60 are observed in *trnW,-trnA,-trnN, -trnC*; 38, 44, 108, 30, 70, 80, 50, 81 and 85 are in *trnH, trnS, trnL, nad5, -nad6, -trnE, cob, trnT, -trnP*. In contrast, it can be observed that within the vertebrate mitogenome there are relatively conserved blocks, as is the case of *cox2, trnK, atp8, atp6, cox3, trnG, nad3, trnR, nad4L, nad4* (9, 13, 11, 10, 9, 11, 10, 12, 10, and 12 rearrangements, respectively) (Fig. 2). However, it can be observed that each taxonomic order in vertebrates, presents rearrangement proportions in different regions. For example,

while in most taxonomic orders, the rearrangements are concentrated in adjacent genes of the control region, high rearrangements proportions in *trnI, -trnQ, trnM* and adjacent genes are shown in Pleuronectiformes, Myctophiformes (Actinopteri class) and Serpentes; Crocodilia in *nad4, trnH, trnS, trnL.* Also Marsupialia and mentioned taxonomic orders of Actinopteri class, share rearrangements proportions in *trnW, -trnA, -trnN, -trnC* (Supplementary Table S2, Fig. 2).

On the other hand, of the 178 duplications present in vertebrate mitochondrial genome, are concentrated mainly in tRNA genes with the 83.7% (149/178), observed in the *nad6-trnE-cob-trnT-trnP* region (19, 23, 22, 29, and 35 copies of each gene respectively) and of the *trnM* gene (35 copies of the single gene; a single copy per mitogenome). The taxonomic class with the highest number of duplicated genes is Aves with 38.2% (68/178) followed by amphibians with 30.9% (55/178) of all gene duplications detected in vertebrates. Most of the duplications observed in amphibians occurred in *trnM* gene (23 copies), all observed in certain species of the order Anura. Another characteristic observed in Aves was the presence of duplicated pseudogenes (ineffective copies). In this taxonomic group, 42 of these copies were observed, most of them (21) correspond to pseudogenes of *cob*; mainly observed in the order Pelecaniformes. In this same taxonomic order, it was also observed copies in the form of pseudogenes of *nad6, trnE, trnT* and *trnP* genes (6, 2, 7, and 5, respectively). Although it is common to find a single additional copy (duplicate) of one or two genes, in certain genomes more than two presumably effective copies of the same gene are observed; in the case of *Cnemaspis limi* (Reptilia, Squamata) that contains four copies of the *trnA* gene. Similarly, tandem duplications of complete mitochondrial regions are observed, as in the case of *Aeluroscalabotes felinus* (Reptilia, Squamata) with 53 genes in

total (18974 bp) and in *Breviceps adspersus* (Amphibia, Anura) with 51 genes (28757 bp) (Supplementary Table S1). Of the 21 deletions that we founded, they also are concentrated in the tRNA genes (19/21), mainly in *trnT-trnP* region with 3 and 6 deletions, respectively. The most frequent taxonomic group of deletions is Actinopteri with 9 deletions, mainly in the *trnP* gene with 5 deletions (Supplementary Table S1).

*Gene Arrangement Convergence in the vertebrate mitochondrial genome*

Our results indicate that within the 1020 reorganized mitogenomes (with respect to the order of the ancestral genes of vertebrate), 138 different genetic architectures were identified; 11 of these are grouped in convergences in the genetic order observed in 764 species (Table 2). The remaining 127 are unique genetic arrangements in vertebrates, as in the case of *Phrynocephalus przewalskii* (Reptilia, Squamata), with duplications of the *trnF* and -*trnP* genes, and inversions/translocations in *trnQ* and in the second copy of -*trnP* gene (Supplementary Table S1).

Our results indicate that there are taxonomic groups with many different architectures. For example, all the sampled species within Aves share a reorganization from the vertebrate ancestral order, but relatively conserved within this taxonomic class. Only 29 species of the 620 analyzed have any type of rearrangement with respect to the Aves ancestral order (see CREX analysis). Among Saccopharyngiformes (Actinopteri), there are three different architectures in the four analyzed species, while in other taxonomic groups they are more conserved; for example, within the class Mammalia, all marsupials (29 mt genomes) share a single rearrangement from the vertebrate ancestral order (Table 1).

In the convergences (11) four of them (8-11) are between orders belonging to the same taxonomic class of fish, in addition, 10 and 11 are convergences in deletion and duplication events of a gene, respectively. All cases of convergence between different taxonomic orders are represented in Fig. 3.

*Analysis of gene order with CREX and TreeREx*

CREX analysis was done with the purpose of determining a putative ancestral gene order for the taxonomic orders that presented more rearrangements, here are the scenarios of events that had to occur to produce a change between the ancestral order to the order of the current genomes (See Fig. S1). In many cases, a modest rearrangement from the ancestral vertebrate gene order occurred at the base of a large group of organisms and is shared broadly within that group, including: (1) a reverse transposition of *cob, trnT, -trnP* in Actenopteri - Anguiliformes; (2) a reverse transposition of *-trnQ* and *-trnC, -trnY*, a reverse transposition of *-trnS, trnD, cox2*, *trnK*, and a reversal of the region of *cox2 to trnH* in Actinopteri - Pleuronectiformes; (3) a transposition of *trnL* in the Amphibia - Anura; (4) a reverse transposition of *nad1*, *trnI* genes, a reversal of the region *trnL* to *nad1*, and a reversal of the region *cox2* to *trnH* in Reptilia - Serpentes; (5) a transposition of *trnS* in Reptilia - Crocodilia; (6) a transposition of the region *cob*, *trnT, -trnP* in Aves; and (7) a Tandem-Duplication-Random-Loss (TDRL) event for the region *trnW, -trnA, -trnN, -trnC, -trnY* in Mammalia - Marsupialia.

On the other hand, TreeREx is a useful algorithm for assigning rearrangements to the edges of a given phylogenetic tree, with which we reconstruct the ancestral genetic orders at the

interior nodes of the most rearranged orders and identify a significant number of transpositions, reverse transpositions, inversions and Tandem-Duplication-Random-Loss (TDRL) events (Supplementary Fig. S2). In Actinopteri - Saccopharyngiformes, TreeREx detected that most nodes are inconsistent and that the most common event is TDRL (Fig. S2A). In Actinopteri - Myctophiformes and Pleuronectiformes, TreeREx detected that most nodes are consistent; the most common events being transpositions (Fig. S2B) and TDRL, respectively (Fig. S2C). In Amphibia - Anura, TreeREx detected that nodes are mostly consistent, with transpositions being the most common event causing reorganizations in this order (Fig. S2D). In Reptilia - Lacertilia, TreeREx detected all the nodes are consistent and the most common event is the inversions (Fig. S2E).

## Discussion

*Gene annotation errors in vertebrate mitochondrial genome*

Due to the advance of sequencing techniques in the last two decades, hundreds of mitochondrial genomes of different taxonomic groups of vertebrates have been deposited in the NCBI database and then subsequently curated into RefSeq files. However, it has been reported that the animal mitochondrial genomes in this database contain many errors of gene annotation, the majority of which are readily detectable [12,27,28]. One source of these errors may be from software that provides automated gene annotation [28], but some misannotations stem from simply failing to note the proper orientation of a gene or an erroneous error of naming. Errors are also sometimes perpetuated by presuming that previously made annotations are correct and then following this for a newly sequenced

mtDNA, the so-calle “percolation of errors” [12]. Prada and Boore[12] reported that 36.3% of mammalian mitochondrial genomes obtained from the NCBI database analyzed had annotation errors.

Our results, similar to those observed solely in mammals, show that tRNA genes are more susceptible to errors in gene notation (94.6%) than other mitochondrial genes, with a higher error percentage in certain taxonomic groups (Actinopteri, Reptilia, and Aves). Popadin et al. [29] suggest performing verification of tRNA gene annotations manually to ensure a higher level of accuracy in annotation, although it could also be done through a combination of semi-automated bioinformatics techniques [12], where the curator would play an important role in detecting these errors. Due to the large number of mitochondrial genomes being reported annually in the database, the probability that these errors in gene notation will continue to spread is high, so a curation of the data by the scientific community is recommended [30]. NCBI is also encouraged to implement elementary error checking mechanisms when promoting a submitted sequence into their RefSeq database as described previously [12,27].

*Genome Rearrangements and vertebrate mitochondrial genome Evolution*

Traditionally, vertebrate mitochondria have been considered to have a "conserved gene order” from primitive vertebrates (fish) to primates [1,31,32]. However, in some lineages, there is a variation in the rate of its reorganization. For example, within vertebrates, gene rearrangements have been found for some species of lizards, amphibians, fish, crocodilians, snakes, tuatara, and lamprey [2,10,16]. Nevertheless, in these studies, it is not identified which

taxonomic groups are the most reorganized within them, or what is the differential reorganization rate among them.

Our results show a differential reorganization proportion within and between taxonomic classes, confirming that Reptilia, Amphibia, Petromyzonti (Lamprey), Mammalia, Actinopteri and Aves, present a proportion between 4.85 and 0.65, in order of highest to lowest. For example, within the Reptilia class, Crocodylia and snakes present a specific reorganization in comparison with the ancestral order of vertebrates, although it is preserved within the same group. All Crocodylia have the order *trnS-trnH-trnL* (derived from an *trnS* transposition from the ancestral order of vertebrates), results consistent with those reported previously [2,33,34]; observing a unique genetic architecture for all this group (Alligatoridae and Crocodylidae). Similarly, most snakes have the order *trnI-trnL-trnQ-trnM*, *trnL* (derived from a reverse transposition and reversal of the ancestral order of vertebrates), which differs from the Viperidae with the order *trnI-trnL-trnP-trnQ-trnM* (derived from a *trnP* translocation), similar results with those reported previously [2,34,35]. However, our results show that the two species of snakes belonging to the Typhlopidae family (*Amerotyphlops reticulatus* and *Indotyphlops braminus*), present the ancestral order of vertebrates and not that of other snakes, consistent with an earlier report [36]. In contrast, within the suborder Lacertilia (Reptilia), we observe that some taxonomic families such as Agamidae, Chamaelonidae and some species of the family Gekkonidae, present a higher rate of reorganization than others families such as Iguanidae or Lacertidae that retain the ancestral vertebrate order [34,37].

Amphibians have been reported to be a more conservative group compared to reptiles [2]. However, our results show that within amphibians, the Anura order (124 species analyzed) is more reorganized, with a total of 103 species showing at least one reorganization event and only 21 with the ancestral vertebrate order. Of these reorganized species, most (86) share different gene orders, and the rest have unique architectures (18). Xia et al.[6] presents the *cob, trnL, trnT, -trnP* gene order as an extensive reorganization in anurans (Neobatrachia), our results confirm this, and that the Anura mitochondrial genome presents greater variability in terms of genetic order than in comparison with other amphibians. Previous studies have postulated that the genetic order *nad6-trnT-trnE-cob-trnP* is common among lampreys [2,38] , known to be the earliest diverged vertebrates with a time of divergence inferred to be 550 mya [39]. Although our results confirm the presence of this reorganization in two species of lampreys (*Lethenteron camtschaticum* and *Petromyzon marinus*) in the family Petromyzontidae, the three remaining species of this family with a reported mtDNA gene arrangement and the one species in the Geotriidae family retain the vertebrate ancestral gene order, so it is not correct to assert that it is an ancestral character of all lampreys.

According to Satoh et al., [10], most of the fishes had the typical gene order widely shared among vertebrate mt genomes, noting that only 14% (35/250) of these species observed have at least one rearrangement, a percentage over three times as high (4.1%; 52/1255) as that presented by Gong et al., [40]. Although our results show a relatively low proportion, similar to those previously published, some taxonomic orders such as Anguilliformes, Saccopharyngiformes, Myctophiformes, Gadiformes, Batrachoidiformes, Pleuronectiformes, Perciformes, and Perciformes (and some listed as sedis mutabilis) are

355 highly reorganized in comparison with other orders in fish. Many of such gene order are
356 unique to a specific taxon, but some are shared polyphyletically between distantly related
357 species.

358 Our results confirm that the majority of species within the class Aves share the *cob, trnT, -*
359 *trnP, -nad6, -trnE* rearrangement from the ancestral vertebrate arrangement as described in
360 early work [41]. In addition to this, multiple independent rearrangements have occurred in
361 some species of birds, including genetic duplications of *trnE, trnT, trnP*, CR (Control
362 Region) and *cob* pseudogenes as has been previously reported [9,42]. This work clarifies that
363 these duplications are concentrated in certain taxonomic groups such as the family
364 Ardeidae (Pelecaniformes), Procellariiformes and Suliformes, marine aquatic birds with
365 great diving capacity, suggesting this gene order is the ancestral pattern within these birds
366 and persisted in most lineages perhaps through concerted evolution [42]. According to Gibb et
367 al., [43], it is very possible that in species of birds of different orders, they have hidden
368 duplications in the genome that also include the control region and cannot be observed
369 because the genome assembly programs for short sequencing reads artifactually collapse
370 these regions. Nevertheless, further physiological and molecular analyses are necessary to
371 assess the potential selective advantages of the mitogenome duplications [44].

372

373 *Convergence in a Hotspot of Gene Rearrangement*

374 Genomes with small size, such as mitochondrial genomes, have fewer mutational targets
375 compared to genomes with large sizes (such as the nuclear genome), so convergent
376 evolution through homologous site mutations is expected to occur more commonly in

377 smaller genomes [45]. Therefore, homoplastic gene orders have been identified in vertebrate
378 mitochondrial genomes[13,46].

379 Our results, as found in previous analysis, have shown that the mitochondrial genome of
380 vertebrates, in certain taxonomic groups, have a considerable proportion of reorganization.
381 Most of these rearrangements involve tRNA genes, *nad5*, and/or the D-loop region, as has
382 been observed for smaller studies[2,3,13-15,37,47]. There are constraints and 'hotspots' for
383 rearrangement, making evolutionary convergence more likely than it would be if there were
384 equal probabilities of change at each gene position[11]. In this work, the identification of
385 three "hot spots" (*trnL, nad1, trnI, -trnQ, trnM*; *trnW, -trnA, -trnN, -trnC* and *-nad6, -trnE,*
386 *cob, trnT, -trnP*) in the mitochondrial genome of vertebrates, in certain lineages, could
387 increase the possibility of evolutionary convergences between different groups of species.
388 The large number of species involved in the convergences (764) observed in this work,
389 favors the view that convergent evolution is a general phenomenon of the vertebrate
390 mtDNA, at least in these hotspot regions, as had been earlier predicted [8,11]. Our results
391 support that convergences occur in two cases, (A) in which nearest neighbor tRNA genes
392 exchange their position as is the case in convergences 4, 6, 7 and 8 and (B) in genes
393 flanking either of the two origins of replication as occurs in convergences 1, 2, 3, 5 9, 10
394 and 11 reported in this study.

395

396 **Conclusion**

397 The analysis of vertebrate mitochondrial genomes available in the database that we
398 performed in this work identified a high error percentage in gene annotations, in addition, it

shows a number of significant rearrangement events (especially in tRNA genes) in these organisms, contrary to what has been believed for many years. In addition, we show that the types and frequency of rearrangements in genomes behave differently between vertebrate classes and between taxonomic orders of classes. We report new convergence events in the gene order among vertebrate mitogenome in this paper.


## Acknowledgements

We thank the University of Tolima for supporting this research. The first author thanks to Oficina de Investigaciones y Desarrollo Científico de la Universidad del Tolima by the early researcher scholarship.


## Author Contributions

MPML: Methodology, Software, Validation, Formal analysis, Investigation and Data Curation for Amphibian and Reptile, Writing - Original Draft, Writing - Review & Editing

MAMC: Methodology, Software, Validation, Formal analysis, Investigation and Data Curation for mammals.

MOC: Software, Validation, Formal analysis, Investigation and Data Curation for fishs.

NSM: Software, Validation, Formal analysis, Investigation and Data Curation for birds.

JLB: Methodology, Validation, Writing - Original Draft, Writing - Review & Editing

CFPQ: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing Visualization, Supervision, Project administration  Funding acquisition.

## Ethics approval and consent to participate

There was no animal experimentation undertaken in this study.

## Competing interests

The authors declare that they have no competing interests.

## References


1 Boore, J. L. Animal mitochondrial genomes. *Nucleic acids research* **27**, 1767-1780, doi:10.1093/nar/27.8.1767 (1999).

2 Pereira, S. L. Mitochondrial genome organization and vertebrate phylogenetics. *Genetics and Molecular Biology* **23**, 745-752 (2000).

3 Formenti, G. *et al.* Complete vertebrate mitogenomes reveal widespread gene duplications and repeats. *bioRxiv*, 2020.2006.2030.177956, doi:10.1101/2020.06.30.177956 (2020).

4 Zhang, J. *et al.* qMGR: A new approach for quantifying mitochondrial genome rearrangement. *Mitochondrion* **52**, 20-23, doi:10.1016/j.mito.2020.02.004 (2020).

5 Gong, L. *et al.* Novel gene rearrangement pattern in Cynoglossus melampetalus mitochondrial genome: New gene order in genus Cynoglossus (Pleuronectiformes:

Cynoglossidae). *International journal of biological macromolecules* **149**, 1232-1240, doi:10.1016/j.ijbiomac.2020.02.017 (2020).

6 Xia, Y., Zheng, Y., Murphy, R. W. & Zeng, X. Intraspecific rearrangement of mitochondrial genome suggests the prevalence of the tandem duplication-random loss (TDLR) mechanism in Quasipaa boulengeri. *BMC genomics* **17**, 965, doi:10.1186/s12864-016-3309-7 (2016).

7 Al Arab, M., Bernt, M., Höner Zu Siederdissen, C., Tout, K. & Stadler, P. F. Partially local three-way alignments and the sequence signatures of mitochondrial genome rearrangements. *Algorithms Mol Biol* **12**, 22, doi:10.1186/s13015-017-0113-0 (2017).

8 Boore, J. L. in *Comparative genomics* 133-147 (Springer, 2000).

9 Mindell, D. P., Sorenson, M. D. & Dimcheff, D. E. Multiple independent origins of mitochondrial gene order in birds. *Proceedings of the National Academy of Sciences* **95**, 10693-10697 (1998).

10 Satoh, T. P., Miya, M., Mabuchi, K. & Nishida, M. Structure and variation of the mitochondrial genome of fishes. *BMC genomics* **17**, 719, doi:10.1186/s12864-016-3054-y (2016).

11 Boore, J. L. & Brown, W. M. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current opinion in genetics & development* **8**, 668-674, doi:10.1016/s0959-437x(98)80035-x (1998).

12 Prada, C. F. & Boore, J. L. Gene annotation errors are common in the mammalian mitochondrial genomes database. *BMC genomics* **20**, 73, doi:10.1186/s12864-019-5447-1 (2019).

13 San Mauro, D., Gower, D. J., Zardoya, R. & Wilkinson, M. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. *Molecular biology and evolution* **23**, 227-234, doi:10.1093/molbev/msj025 (2006).

14 Zhong, J., Li, G., Liu, Z. Q., Li, Q. W. & Wang, Y. Q. [Gene rearrangement of mitochondrial genome in the vertebrate]. *Yi chuan xue bao = Acta genetica Sinica* **32**, 322-330 (2005).

15 Gissi, C., Iannelli, F. & Pesole, G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **101**, 301-320, doi:10.1038/hdy.2008.62 (2008).

16 Sankoff, D. *et al.* Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6575-6579, doi:10.1073/pnas.89.14.6575 (1992).

17 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).

18 Bernt, M. *et al.* MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution* **69**, 313-319, doi:10.1016/j.ympev.2012.08.023 (2013).

19 Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* **1962**, 1-14, doi:10.1007/978-1-4939-9173-0_1 (2019).

20 Babicki, S. *et al.* Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* **44**, W147-153, doi:10.1093/nar/gkw419 (2016).

21 Amemiya, C. T. *et al.* Analysis of the African coelacanth genome sheds light on tetrapod evolution. *Nature* **496**, 311, doi:10.1038/nature12027 (2013).

22 Bernt, M., Merkle, D. & Middendorf, M. in *RECOMB International Workshop on Comparative Genomics.* 143-157 (Springer).

23 Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394-1403, doi:10.1101/gr.2289704 (2004).

24 Rambaut, A. (2014).

25 Hannenhalli, S. & Pevzner, P. A. in *Proceedings of IEEE 36th annual foundations of computer science.* 581-592 (IEEE).

26 Tesler, G. GRIMM: genome rearrangements web server. *Bioinformatics* **18**, 492-493, doi:10.1093/bioinformatics/18.3.492 (2002).

27 Boore, J. L. Requirements and standards for organelle genome databases. *Omics* **10**, 119-126, doi:10.1089/omi.2006.10.119 (2006).

28 Donath, A. *et al.* Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic acids research* **47**, 10543-10552, doi:10.1093/nar/gkz833 (2019).

29 Popadin, K. Y., Mamirova, L. A. & Kondrashov, F. A. A manually curated database of tetrapod mitochondrially encoded tRNA sequences and secondary structures. *BMC Bioinformatics* **8**, 441, doi:10.1186/1471-2105-8-441 (2007).

30 Brenner, S. E. Errors in genome annotation. *Trends Genet* **15**, 132-133, doi:10.1016/s0168-9525(99)01706-0 (1999).

31 Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465, doi:10.1038/290457a0 (1981).

32 Noack, K., Zardoya, R. & Meyer, A. The complete mitochondrial DNA sequence of the bichir (Polypterus ornatipinnis), a basal ray-finned fish: ancient establishment of the consensus vertebrate gene order. *Genetics* **144**, 1165-1180 (1996).

33 Janke, A. & Arnason, U. The complete mitochondrial genome of Alligator mississippiensis and the separation between recent archosauria (birds and crocodiles). *Molecular biology and evolution* **14**, 1266-1272, doi:10.1093/oxfordjournals.molbev.a025736 (1997).

34 Kumazawa, Y. & Nishida, M. Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Molecular biology and evolution* **12**, 759-772, doi:10.1093/oxfordjournals.molbev.a040254 (1995).

35 Kumazawa, Y., Ota, H., Nishida, M. & Ozawa, T. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular biology and evolution* **13**, 1242-1254, doi:10.1093/oxfordjournals.molbev.a025690 (1996).

36 Yan, J., Li, H. & Zhou, K. Evolution of the mitochondrial genome in snakes: gene rearrangements and phylogenetic relationships. *BMC genomics* **9**, 569, doi:10.1186/1471-2164-9-569 (2008).

37 Macey, J. R., Larson, A., Ananjeva, N. B., Fang, Z. & Papenfuss, T. J. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular biology and evolution* **14**, 91-104, doi:10.1093/oxfordjournals.molbev.a025706 (1997).

38 Delarbre, C. *et al.* The complete nucleotide sequence of the mitochondrial DNA of the agnathan Lampetra fluviatilis: bearings on the phylogeny of cyclostomes.

*Molecular biology and evolution* **17**, 519-529, doi:10.1093/oxfordjournals.molbev.a026332 (2000).

39 Lee, W. J. & Kocher, T. D. Complete sequence of a sea lamprey (Petromyzon marinus) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics* **139**, 873-887 (1995).

40 Gong, L., Shi, W., Si, L. Z. & Kong, X. Y. [Rearrangement of mitochondrial genome in fishes]. *Dong wu xue yan jiu = Zoological research* **34**, 666-673 (2013).

41 Desjardins, P. & Morais, R. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *Journal of molecular biology* **212**, 599-634, doi:10.1016/0022-2836(90)90225-b (1990).

42 Zhou, X., Lin, Q., Fang, W. & Chen, X. The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. *BMC genomics* **15**, 573, doi:10.1186/1471-2164-15-573 (2014).

43 Gibb, G. C. *et al.* New Zealand Passerines Help Clarify the Diversification of Major Songbird Lineages during the Oligocene. *Genome biology and evolution* **7**, 2983-2995, doi:10.1093/gbe/evv196 (2015).

44 Urantówka, A. D., Kroczak, A. & Mackiewicz, P. New view on the organization and evolution of Palaeognathae mitogenomes poses the question on the ancestral gene rearrangement in Aves. *BMC genomics* **21**, 874, doi:10.1186/s12864-020-07284-5 (2020).

45 Stern, D. L. The genetic causes of convergent evolution. *Nat Rev Genet* **14**, 751-764, doi:10.1038/nrg3483 (2013).

46 Macey, J. R., Papenfuss, T. J., Kuehl, J. V., Fourcade, H. M. & Boore, J. L. Phylogenetic relationships among amphisbaenian reptiles based on complete

mitochondrial genomic sequences. *Molecular phylogenetics and evolution* **33**, 22-31, doi:10.1016/j.ympev.2004.05.003 (2004).

47 Pääbo, S., Thomas, W. K., Whitfield, K. M., Kumazawa, Y. & Wilson, A. C. Rearrangements of mitochondrial transfer RNA genes in marsupials. *J Mol Evol* **33**, 426-430, doi:10.1007/bf02103134 (1991).

## List of tables and figures

Table 1. Number and distribution of analyzed mitochondrial genomes with summary of types of deletions, duplications, inversions, translocations that have been reported in GenBank correctly or as gene annotation errors detected in this work. * The order has no species with the ancestral gene order. A: Ancestral architecture, in some taxonomic groups in addition to the different architectures (with rearrangements) organisms with the ancestral organization can also be found. Dupli. - Duplications, Dele. -Deletions, Inver. - Inversions, Trans. -Translocations.

Table 2. Gene Arrangement: Convergence in the mitochondrial genome of Vertebrata.

Fig. 1. Gene rearrangement proportion values for each of the vertebrate classes sampled.

Fig. 2. Heat map of gene rearrangement analysis among vertebrate classes. Phylogenetic relationships are as interpreted in [21]. Dark green colors show a low proportion of change and red colors show a high number of rearrangement events for each of the individual genes within each taxonomic order that exhibited rearrangements in mitochondrial sequences. Orange diamonds show the number of convergences detected in that taxonomic

579 order; aquamarine blue circles indicate that a CREx representation was performed for that
580 taxonomic order.

581 Fig. 3. Evolutionary convergence of mitochondrial gene order rearrangements. This shows
582 only the subset that are rearranged; all other genes share the ancestral arrangement. Arrows
583 show transcriptional orientation. tRNA-encoding genes are labeled with the one-letter code
584 for the corresponding amino acid. In parentheses are the number of species involved in the
585 convergence.

586

587 **Supplementary Material**

588 Table S1. Analysis of gene arrangements of vertebrate mitochondrial genome. NCBI with
589 Reference IDs shown for each species. The numeral in parentheses indicates the number of
590 species belonging to each taxonomic group. Each gene is assigned a number 1 to 37 at the
591 top of the Table which is then used to describe the annotated gene arrangement for each
592 species. Each gene is transcribed left-to-right as shown except for those with a minus (-)
593 symbol to indicate opposite orientation. The ancestral gene order has been inferred to be as
594 postulated by [1]. Highlighting indicates all deviations from that ancestral arrangement as
595 confirmed by the bioinformatics tools used in this study, with being yellow to gene
596 inversion-translocation, light blue to gene duplication, purple to pseudogenes and red to
597 gene deletions.

598

599 Table S2. Mitochondrial genome rearrangement rate in each gene by taxonomic order in
600 Hexapoda. (M) manual and qMGR rearrangement rate analysis are indicated by each
601 mitochondrial gene.

602 Fig. S1. Parsimonious evolutionary scenario leading to the extant gene orders in vertebrate
603 by CREx analysis. Arrows show transcriptional orientation. tRNA-encoding genes are
604 labeled with the one-letter code for the corresponding amino acid. (A) Actenopteri-
605 Anguiligormes partial ancestral order, (B) Actenopteri-Pleuronectiformes ancestral order,
606 (C) Amphibia-Anura partial ancestral, (D) Reptilia-Lacertilia-Chamaelonidae and
607 Agamidae ancestor order, (E) Reptilia-Serpentes ancestor order, (F) Reptilia-Crocodilia
608 ancestor order, (G) Aves ancestral order, (H) Mammalia-Marsupialia ancestor order. Each
609 evolutionary event (reversal, transposition and reversed translocations) is signaled to
610 transform the ancestral genome into the derived mitochondrial genome. The superscript
611 indicates the taxonomic orders that have partial ancestor (ancestor with reorganizations and
612 with ancestral order).

613

614 Fig. S2. TreeREx analysis for vertebrate class. (A) Actinopteri - Saccopharyngiformes, (B)
615 Actinopteri - Myctophiformes, (C) Actinopteri, (D) Amphibia – Anura. The rearrangements
616 on the branches are given as Transposition, Inversion, Inverse transposition and Tandem-
617 Duplication-Random-Loss events (TDRLs).

618

619

620

621 Table 1. Number and distribution of analyzed mitochondrial genomes with summary of types of deletions, duplications, inversions, translocations

622 that have been reported in GenBank correctly or as gene annotation errors detected in this work.

| Class | No. orders | No. species | No. of different genomes within the order | Genes differing in rearrangement | Numer confirmed | Dup | Del | Inv-tra | Numer refuted | % of reorganization |
|---|---|---|---|---|---|---|---|---|---|---|
| Myxini | 1 | 2 | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Petromyzonti | 1 | 6 | 1+A | 6 | 6 | 0 | 0 | 6 | 0 | 2.70 |
| Elasmobranchii | 12 | 74 | A | 2 | 0 | 0 | 0 | 0 | 2 | 0 |
| Holocephali | 1 | 5 | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cladistii | 1 | 2 | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Actinopteri | 59 | 1259 | 60+A | 751 | 628 | 25 | 9 | 594 | 123 | 1.61 |
| Coelacanthi | 1 | 1 | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Dipneusti | 1 | 3 | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Amphibia | 3 | 241 | 34+A | 292 | 262 | 55 | 4 | 203 | 30 | 3.27 |
| Reptilia | 6 | 314 | 26+A | 564 | 475 | 30 | 4 | 441 | 89 | 4.85 |
| Aves | 28 | 620 | 15* | 149 | 72 | 68 | 4 | 0 | 77 | 0.65 |
| Mammalia | 29 | 304 | 2+A | 187 | 119 | 0 | 0 | 119 | 68 | 1.66 |
| | | **2831** | **-** | **1951** | **1562** | **178** | **21** | **1363** | **389** | |

623

624 **Table 2.** Gene Arrangement: Convergence in the mitochondrial genome of Vertebrata.

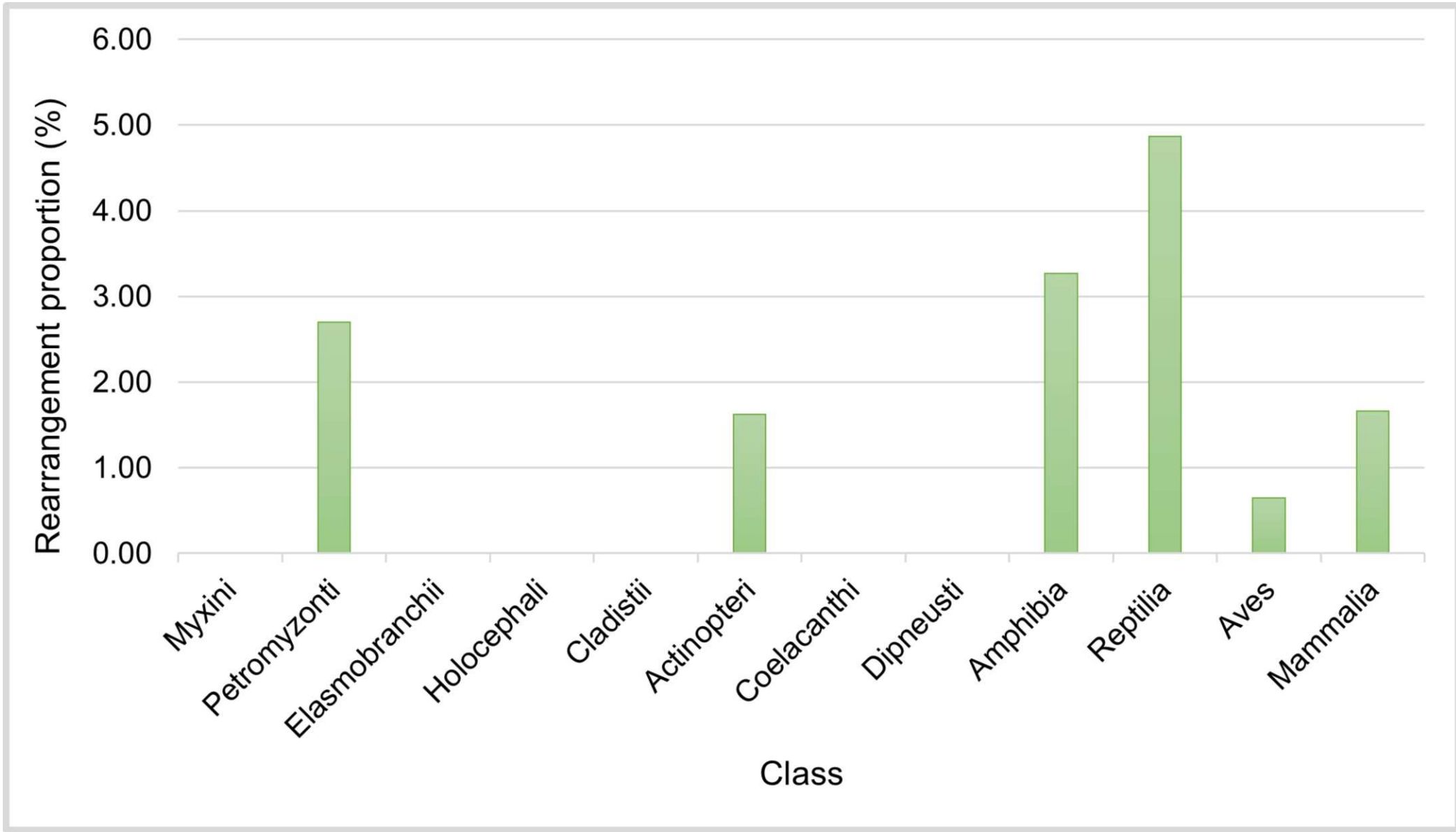
| Convergence | Taxonomic level | Order | N. genome |
|---|---|---|---|
| Convergence 1 (-*trnA, -trnC, trnW, -trnN*) | *Siphonops annulatus* | Gymnophiona (Amphibia) | 1 |
| | | Marsupialia (Mammalia) | 29 |
| Convergence 2 (*cob, trnT, -nad6, -trnE)* | *Coloconger cadenati* | Anguiliformes (Actinopteri) | 14 |
| | *Ariosoma shiroanago* | | |
| | *Paraconger notialis* | | |
| | *Conger japonicus* | | |
| | *Congriscus sp.* | | |
| | *Heteroconger hassi* | | |
| | *Derichthys serpentinus* | | |
| | *Nessorhamphus ingolfianus* | | |
| | *Cynoponticus ferox* | | |
| | *Muraenesox bagio* | | |
| | *Facciolella oxyrhyncha* | | |
| | *Hoplunnis punctata* | | |
| | *Nettastoma parviceps* | | |
| | *Ophisurus macrorhynchos* | | |
| | *Chaenocephalus aceratus* | Perciformes (Actinopteri) | 3 |
| | *Chionodraco hamatus* | | |
| | *Notothenia coriiceps* | | |
| | *Aneides flavipunctatus* | Caudata (Amphibia) | 2 |
| | *Stereochilus marginatus* | | |
| Convergence 3 (*trnT, -trnP, -nad6, -trnE)* | *Pagothenia borchgrevinki* | Perciformes (Actinopteri) | 1 |
| | *Rhineura floridana* | Amphisbaenia (Reptilia) | 1 |
| | | Aves | 591 |
| Convergence 4 (-*trnQ, trnI, trnM)* | *Kurtus gulliveri* | Kurtiformes (Actinopteri) | 1 |
| | *Ponticola kessleri* | Gobiiformes (Actinopteri) | 1 |
| | *Brookesia decaryi* | Lacertilia (Reptilia) | 29 |
| | *Chamaeleo africanus* | | |
| | *Chamaeleo arabicus* | | |
| | *Chamaeleo calcaricarens* | | |
| | *Chamaeleo calyptratus* | | |
| | *Chamaeleo chamaeleon* | | |
| | *Chamaeleo dilepis* | | |
| | *Chamaeleo monachus* | | |
| | *Chamaeleo zeylanicus* | | |
| | *Furcifer oustaleti* | | |

| | | | |
|---|---|---|---|
| | *Kinyongia fischeri* | | |
| | *Trioceros melleri* | | |
| | *Acanthosaura armata* | | |
| | *Acanthosaura lepidogaster* | | |
| | *Chlamydosaurus kingii* | | |
| | *Hydrosaurus amboinensis* | | |
| | *Leiolepis boehmei* | | |
| | *Leiolepis guttata* | | |
| | *Phrynocephalus albolineatus* | | |
| | *Phrynocephalus axillaris* | | |
| | *Phrynocephalus grumgrzimailoi* | | |
| | *Phrynocephalus guinanensis* | | |
| | *Phrynocephalus helioscopus* | | |
| | *Phrynocephalus mystaceus* | | |
| | *Phrynocephalus putjatai* | | |
| | *Pogona vitticeps* | | |
| | *Pseudotrapelus sinaitus* | | |
| | *Uromastyx benti* | | |
| | *Xenagama taylori* | | |
| Convergence 5 (-*nad6, cob, trnT, -trnP, -trnE*) | *Cetonurus globiceps* | Gadiformes (Actinopteri) | 3 |
| | *Coelorinchus kishinouyei* | | |
| | *Ventrifossa garmani* | | |
| | *Uroplatus fimbriatus* | Lacertilia (Reptilia) | 1 |
| Convergence 6 (*nad1, trnI, trnL)* | *Squalogadus modificatus* | Gadiformes (Actinopteri) | 1 |
| | *Trachyrincus murrayi* | | 1 |
| | | Alethinophidia-Serpentes (Reptilia) | 57 |
| Convergence 7 (*trnS, trnH*) | *Aulorhynchus flavidus* | Perciformes (Actinopteri) | 1 |
| | | Crocodylia (Reptilia) | 18 |
| Convergence 8 (*trnD, -trnS)* | *Normichthys operosus* | Alepocephaliformes (Actinopteri) | 1 |
| | *Ambassis gymnocephalus* | Perciformes ‘sedis mutabilis’ (Actinopteri) | 1 |
| Convergence 9 (*cob, -trnP, trnT)* | *Dallia pectoralis* | Esociformes (Actinopteri) | 1 |
| | *Rudarius ercodes* | Tetraodontiformes (Actinopteri) | 1 |
| Convergence 10 (-*trnP* deletion) | *Trichiurus japonicus* | Scombriformes (Actinopteri) | 1 |
| | *Hapalogenys analis* | Acanthuriformes (Actinopteri) | 1 |

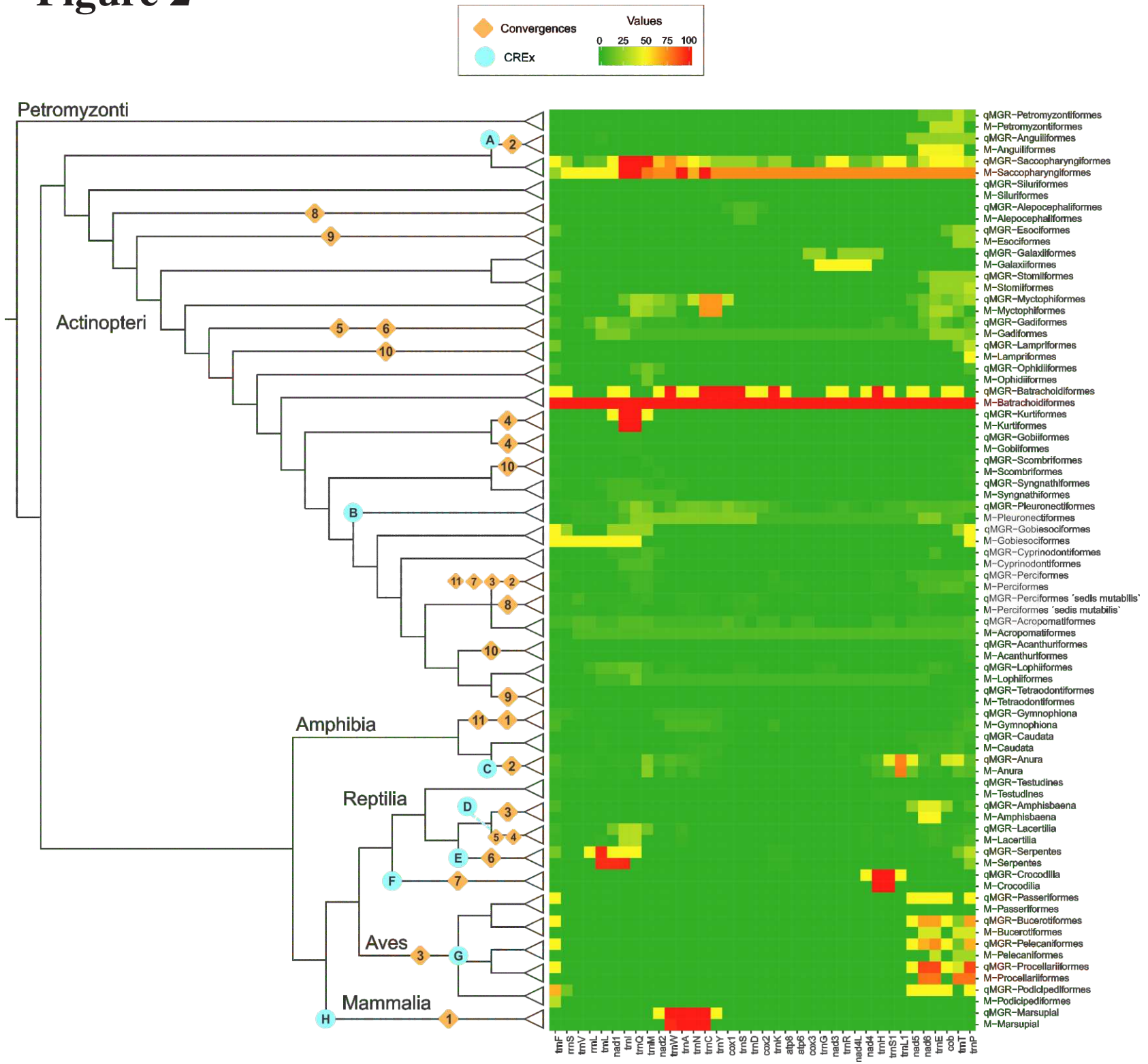
| | | | |
|---|---|---|---|
| | *Lampris guttatus* | Lampriformes (Actinopteri) | 1 |
| Convergence 11 (-*trnP* duplication) | *Clinocottus analis* | Perciformes (Actinopteri) | 1 |
| | *Boulengerula taitana* | Gymnophiona (Amphibia) | 1 |
| | | **Total** | 764 |

625

**Figure 1**

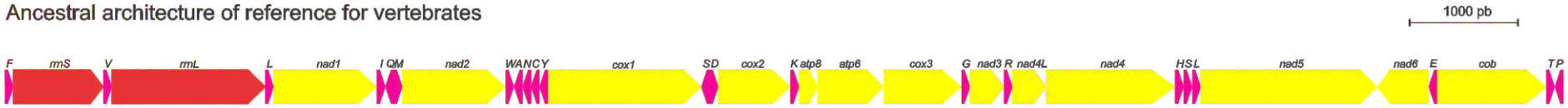

6.00
5.00
4.00
3.00
2.00
1.00
0.00
Rearrangement proportion (%)
Myxini
Petromyzonti
Elasmobranchii
Holocephali
Cladistii
Actinopteri
Coelacanthi
Dipneusti
Amphibia
Reptilia
Aves
Mammalia
Class

# Figure 2


Convergences
CREx
Values
0 25 50 75 100
Petromyzonti
Actinopteri
Amphibia
Reptilia
Aves
Mammalia
A
B
C
D
E
F
G
H
qMGR−Petromyzontiformes
M−Petromyzontiformes
qMGR−Anguilliformes
M−Anguilliformes
qMGR−Saccopharyngiformes
M−Saccopharyngiformes
qMGR−Siluriformes
M−Siluriformes
qMGR−Alepocephaliformes
M−Alepocephaliformes
qMGR−Esociformes
M−Esociformes
qMGR−Galaxiiformes
M−Galaxiiformes
qMGR−Stomiiformes
M−Stomiiformes
qMGR−Myctophiformes
M−Myctophiformes
qMGR−Gadiformes
M−Gadiformes
qMGR−Lampriformes
M−Lampriformes
qMGR−Ophidiiformes
M−Ophidiiformes
qMGR−Batrachoidiformes
M−Batrachoidiformes
qMGR−Kurtiformes
M−Kurtiformes
qMGR−Gobiiformes
M−Gobiiformes
qMGR−Scombriformes
M−Scombriformes
qMGR−Syngnathiformes
M−Syngnathiformes
qMGR−Pleuronectiformes
M−Pleuronectiformes
qMGR−Gobiesociformes
M−Gobiesociformes
qMGR−Cyprinodontiformes
M−Cyprinodontiformes
qMGR−Perciformes
M−Perciformes
qMGR−Perciformes 'sedis mutabilis'
M−Perciformes 'sedis mutabilis'
qMGR−Acropomatiformes
M−Acropomatiformes
qMGR−Acanthuriformes
M−Acanthuriformes
qMGR−Lophiiformes
M−Lophiiformes
qMGR−Tetraodontiformes
M−Tetraodontiformes
qMGR−Gymnophiona
M−Gymnophiona
qMGR−Caudata
M−Caudata
qMGR−Anura
M−Anura
qMGR−Testudines
M−Testudines
qMGR−Amphisbaena
M−Amphisbaena
qMGR−Lacertilia
M−Lacertilia
qMGR−Serpentes
M−Serpentes
qMGR−Crocodilia
M−Crocodilia
qMGR−Passeriformes
M−Passeriformes
qMGR−Bucerotiformes
M−Bucerotiformes
qMGR−Pelecaniformes
M−Pelecaniformes
qMGR−Procellariiformes
M−Procellariiformes
qMGR−Podicipediformes
M−Podicipediformes
qMGR−Marsupial
M−Marsupial
trnF rrnS trnV rrnL trnL nad1 trnI trnQ trnM nad2 trnW trnA trnN trnC trnY cox1 trnS trnD cox2 trnK atp8 atp6 cox3 trnG nad3 trnR nad4L nad4 trnH trnS1 trnL1 nad5 nad6 trnE cob trnT trnP

# Figure 3

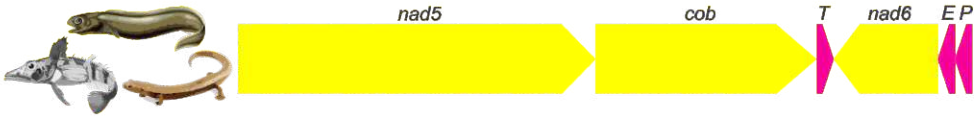
Ancestral architecture of reference for vertebrates




Convergences present in vertebrates. The number of species involved is indicated in parentheses

Convergence 1. Amphibia: Gymnophiona (1) - Mammalia: Marsupialia (29)


Convergence 2. Actinopteri: Anguilliformes (14), Perciformes (3) - Amphibia: Caudata (2)




Convergence 3. Actinopteri: Perciformes (1) - Amphibia: Amphisbaenia (1) - Aves (591)

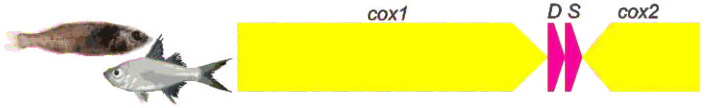

Convergence 4. Actinopteri: Kurtiformes (1), Gobiiformes (1) - Reptilia: Lacertilia (29)


Convergence 5. Actinopteri: Gadiformes (3) - Reptilia: Lacertilia (1)


Convergence 6. Actinopteri: Gadiformes (2) - Reptilia: Serpentes (57)


Convergence 7. Actinopteri: Perciformes (1) - Reptilia: Crocodilia (18)

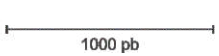

Convergence 8. Actinopteri: Alepocephaliformes (1), Perciformes *sedis mutabilis* (1)




Convergence 9. Actinopteri: Esociformes (1), Tetradontiformes (1)


Convergence 10. Actinopteri: Scombriformes (1), Acanthuriformes (1), Lampriformes (1)


Convergence 11. Actinopteri: Perciformes (1) - Amphibia: Gymnophiona (1)

# Figures

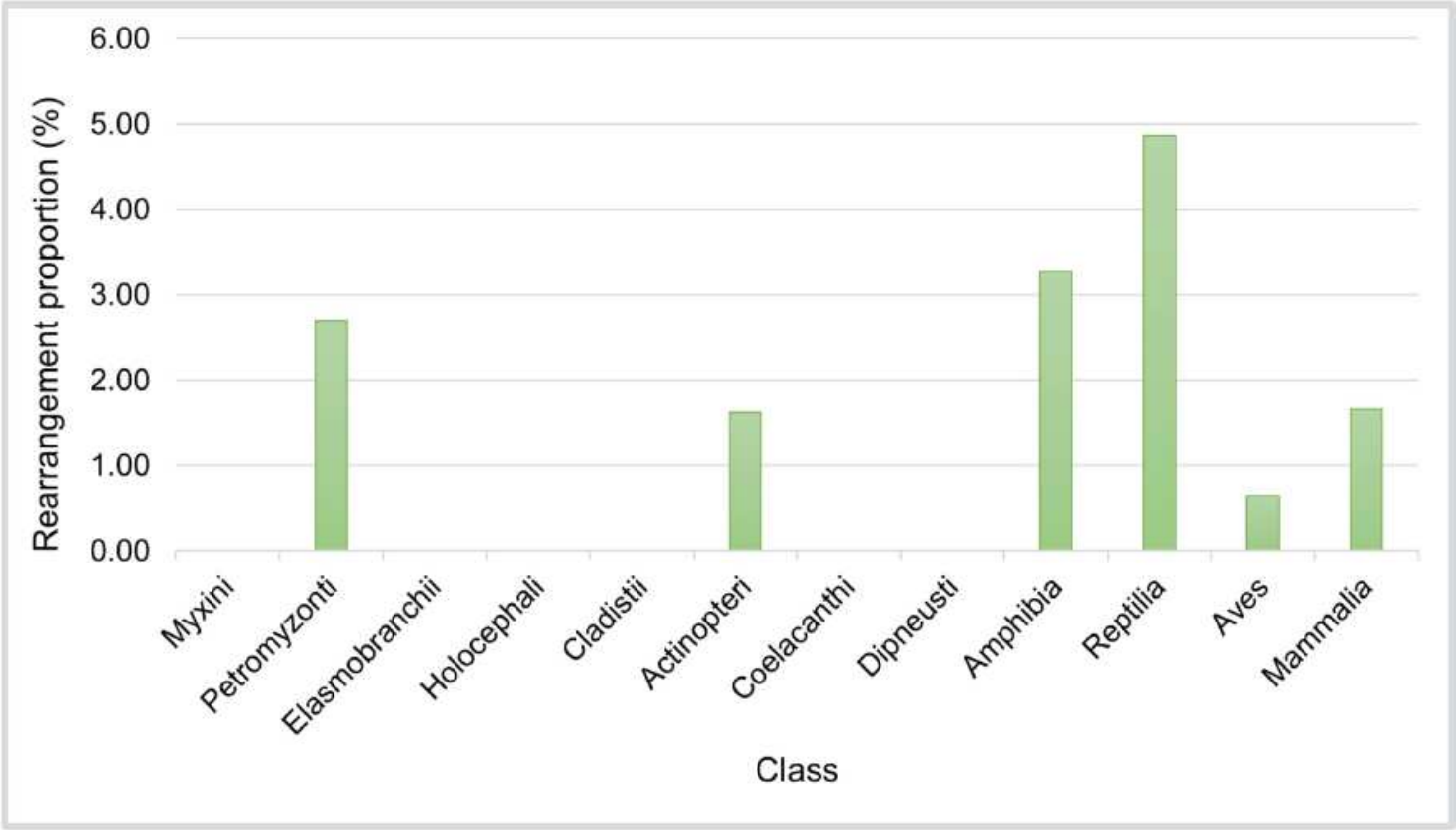
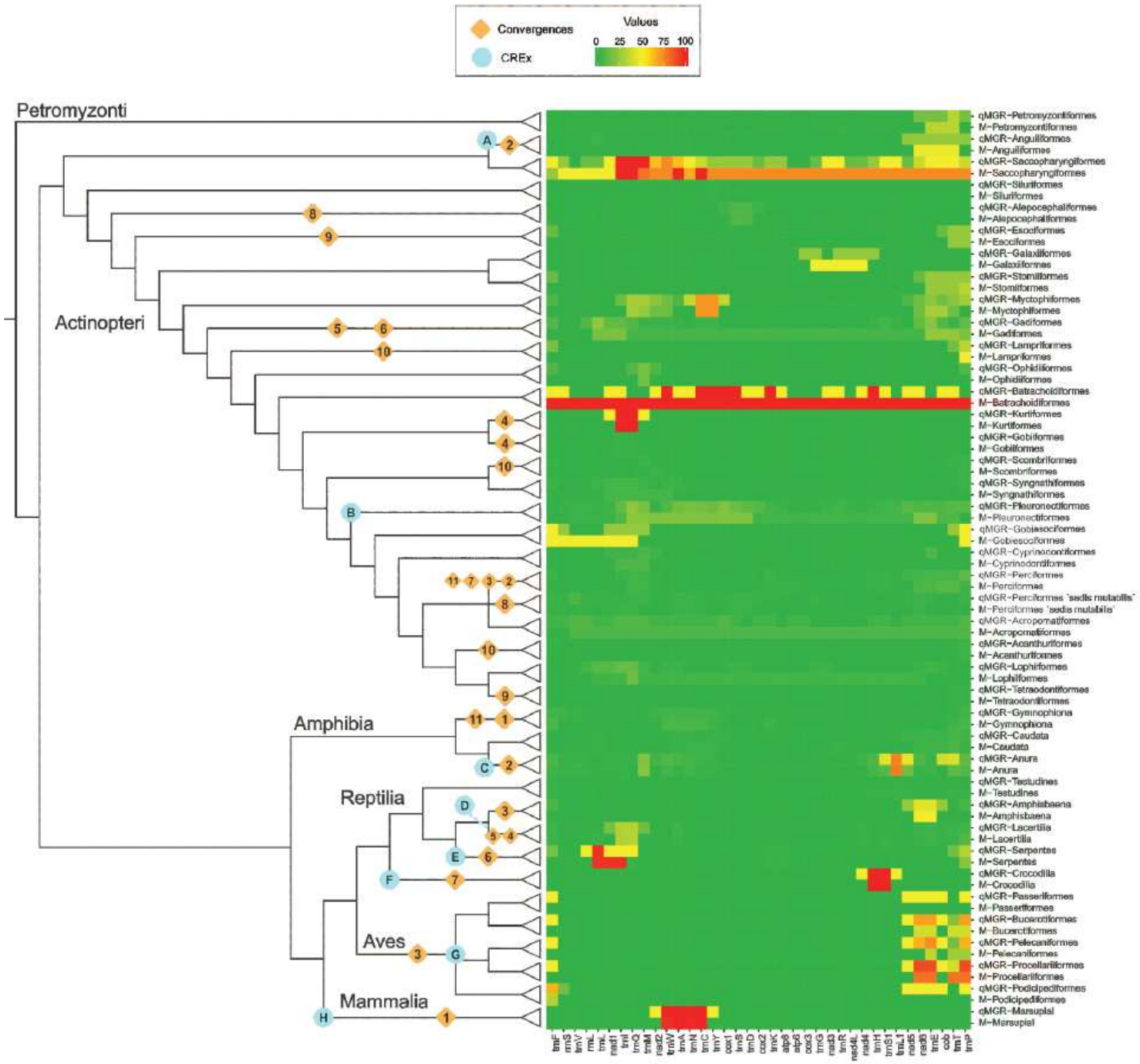



**Figure 1**

Gene rearrangement proportion values for each of the vertebrate classes sampled.




## Figure 2

Heat map of gene rearrangement analysis among vertebrate classes. Phylogenetic relationships are as interpreted in 21. Dark green colors show a low proportion of change and red colors show a high number of rearrangement events for each of the individual genes within each taxonomic order that exhibited rearrangements in mitochondrial sequences. Orange diamonds show the number of convergences detected in that taxonomic order; aquamarine blue circles indicate that a CREx representation was performed for that taxonomic order.

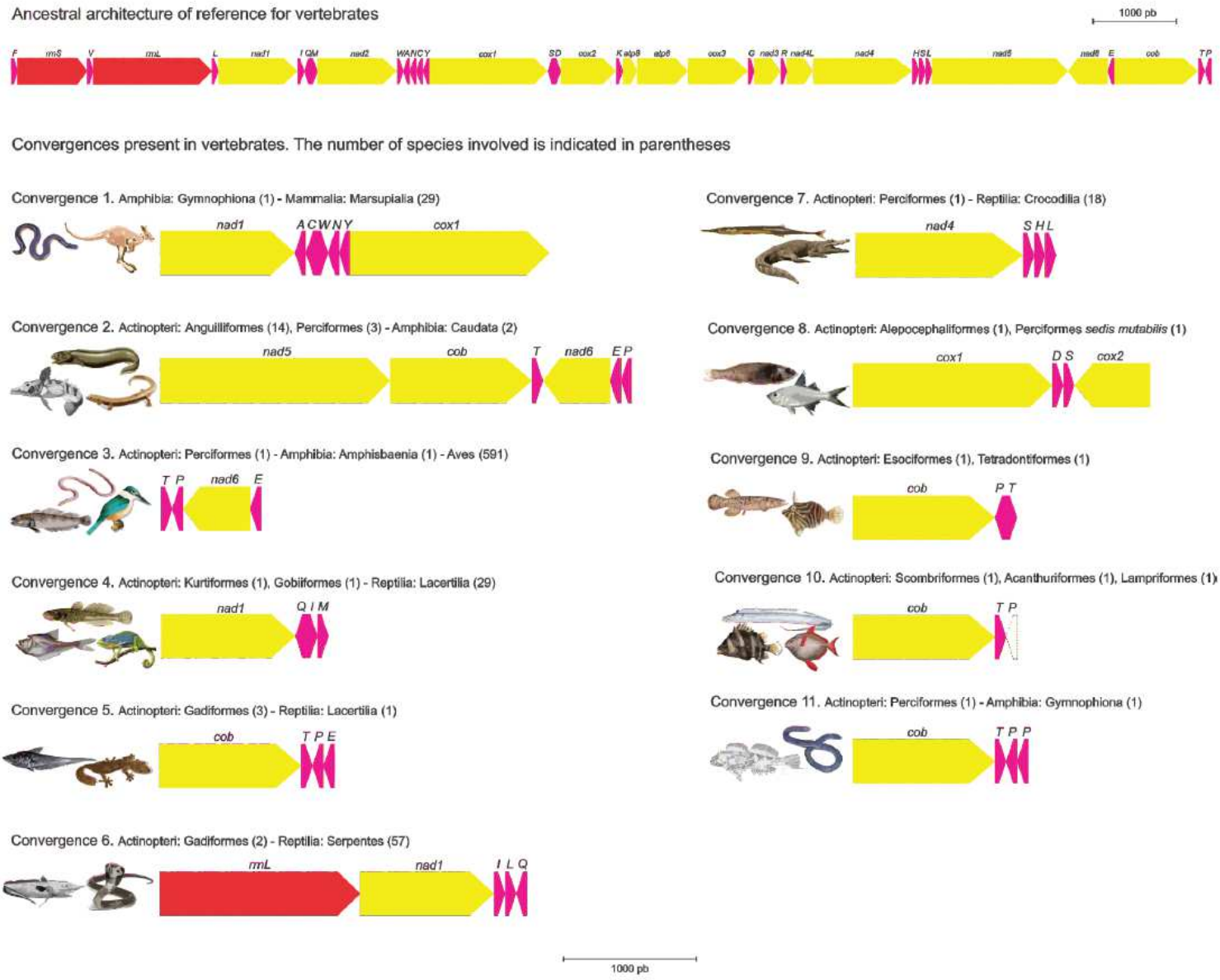



**Figure 3**

Evolutionary convergence of mitochondrial gene order rearrangements. This shows only the subset that are rearranged; all other genes share the ancestral arrangement. Arrows show transcriptional orientation. tRNA-encoding genes are labeled with the one-letter code for the corresponding amino acid. In parentheses are the number of species involved in the convergence.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS1.pdf
- SupplementaryFigureS2.pdf
- SupplementaryTableS1.xlsx

- SupplementaryTableS2.xlsx