ISJ 6: S21-S28, 2009 ISSN 1824-307X

REVIEW

Evolutionary mitogenomics of Chordata: the strange case of ascidians and vertebrates

C Gissi, F Griggio, F Iannelli

Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Milano, Italy

Accepted March 13, 2009

Abstract

The availability of almost one thousand complete mitochondrial genome (mtDNA) sequences of chordates provides an almost unique opportunity to analyse the evolution of this genome in the phylum Chordata, and to identify possible divergent evolutionary trends followed by the three chordate subphyla: Vertebrata, Cephalochordata and Tunicata.

Here, we review some genome-level features of mtDNA, such as genetic code, gene content, genome architecture and gene strand asymmetry, mostly focusing on differences existing between tunicates and remaining chordates. Indeed, tunicate mtDNAs show a surprisingly high variability in several genome-level features, even though the current tunicate taxon sampling is absolutely insufficient and is focused mainly on the class Ascidiacea. On the contrary, a stabilization of the mtDNA structural and evolutionary features is observed in both cephalochordates and vertebrates, where genome-level features are almost invariant. Thus, different evolutionary dynamics, probably related to divergent functional constraints, have modelled the overall mtDNA structure and organization of the three chordate subphyla.

Key Words: mitochondrial genome; evolution; chordates; tunicates; ascidians

Introduction

The mitochondrial genome (mtDNA) is used both as a model system for studying genome evolution, and as a molecular marker to reconstruct animal phylogeny, both at high and low taxonomic levels (Saccone *et al.*, 1999; Saccone *et al.*, 2002). This dual interest in mtDNA constitutes an advantage in phylogenetic studies, since the understanding of the evolutionary peculiarities of a given character greatly facilitates the evaluation of its reliability as a phylogenetic marker. For phylogenetic purposes, the sequences of one or more mitochondrial genes, or of the fast-evolving control region of vertebrates, are commonly analysed with traditional molecular evolutionary methods. Besides the sequence itself, genome-level features have been also used for phylogenetic reconstructions: gene order, gene content, and changes in the genetic code are all examples of mt genome-level features, used to clarify controversial phylogenetic relationships, especially at high taxonomic levels (Boore, 2006).

Corresponding author:
Carmela Gissi
Dipartimento di Scienze Biomolecolari e Biotecnologie
Università di Milano
via Celoria 26, 20133 Milano, Italy
E-mail: carmela.gissi@unimi.it

To date, the entire mtDNA sequence has been determined for 1206 metazoan species (Gissi *et al.*, 2008) but the taxon sampling is highly biased towards vertebrates and arthropods (GenEmbl, September 2008). The taxon sampling of Chordata is also quite heterogeneous, as it is almost exhaustive for vertebrates and cephalochordates (lancelets) but absolutely inadequate for tunicates, where sampling is almost exclusively derived from representatives of the class Ascidiacea (Fig. 1). In spite of the scarce data, it is evident that ascidian mtDNAs possess several unusual features compared to other chordates, such as a fast nucleotide substitution rate (Yokobori *et al.*, 1999, 2005), gene orders that are extremely variable both within the class and compared to other metazoans (Gissi *et al.*, 2004; Yokobori *et al.*, 2005; Iannelli *et al.*, 2007a), and a variable number of tRNA genes (Gissi and Pesole, 2003; Gissi *et al.*, 2004; Iannelli *et al.*, 2007a). Thus, several clues point to the existence of an almost opposite mtDNA evolutionary trend between tunicate and remaining chordates, resulting in a strong variability of mt genome features in ascidians/tunicates versus a relative stability in vertebrates and cephalochordates (Gissi *et al.*, 2008). The “strange case” of ascidian mtDNA is even more intriguing when we consider the controversies on the phylogenetic position of

tunicates within chordates and the debated phylogenetic relationships among the three tunicate classes. Traditional morphological data indicate that tunicates should constitute the basal chordate subphylum - clustering cephalochordates and vertebrates in the Euchordata group (Rowe, 2004). However, recent molecular phylogenetic analyses based on a large number of nuclear genes strongly support a sister relationship between tunicates and vertebrates (Olfactores clade) and a basal position of cephalochordates within chordates (Bourlat *et al.*, 2006; Delsuc *et al.*, 2006). In reality, the position of tunicates in molecular phylogenetic analyses is quite unstable and varies according to which genes or molecules (nuclear or mitochondrial DNA) are used (Zrzavy *et al.*, 1998; Cameron *et al.*, 2000; Winchell *et al.*, 2002; Oda *et al.*, 2004; Bourlat *et al.*, 2006; Delsuc *et al.*, 2006). With regard to phylogenetic relationships within Tunicata, the analyses of morphological and/or molecular characters have given rise to distinct phylogenetic hypotheses most of which suggest that Larvacea is the sister group to the rest of tunicates, and that Thaliacea is the sister group to phlebobranch ascidians, thus rendering Ascidiacea a paraphyletic group (see references in Stach and Turbeville, 2002; Zeng and Swalla, 2005).

In this review we will provide an overview of some structural and evolutionary peculiarities of the chordate mtDNA, investigated through the analysis of a carefully revised dataset of 865 complete mtDNAs (obtained from Gissi *et al.*, 2008). In particular, we will compare four genome-level features: genetic code, gene content, genome architecture and gene strand asymmetry, among the three subphyla of Vertebrata, Cephalochordata and Tunicata, highlighting both differences and similarities. Our synthesis should stimulate further comprehensive analyses of these and other mitochondrial genomic features within a phylogenetic framework.

Genetic code

The three chordate subphyla use each their own modified mitochondrial genetic code (Table 1), differing only in the meaning of AGR codons. The cephalochordate genetic code is quite controversial, as the AGR codons were initially assigned to glycine (Spruyt *et al.,* 1998) and then to serine (Boore, 1999) based on the analysis of a single lancelet mtDNA. Further studies, carried out on a large number of mtDNAs, and comparing the conservation within deuterostomes of amino acid sites with AGR codons in lancelets, have confirmed the hypothesis that AGR encodes serine (Nohara *et al.*, 2005a). It should be noted that the codon usage of lancelets is characterized by the absence or extremely low number of AGG codons, which occur only 4 times in a total of 14 available complete mtDNAs (see notes of Table 1), thus the meaning of AGG codon remains uncertain. These data leave the evolutionary history of the chordate mt genetic code an open question, particularly in the light of the recently reported basal phylogenetic position of lancelets compared to vertebrates and tunicates (Bourlat *et al.*, 2006; Delsuc *et al.*, 2006). In this respect, it is interesting that the genetic codes of both vertebrates and tunicates are unique among metazoans, while the genetic code of lancelets is shared by many protostome and deuterostome phyla (i.e., Annelida, Arthropoda, Mollusca, Nematoda and Xenoturbellida). Thus, the data on genetic code seem to support the basal position of cephalochordates compared to remaining chordates.

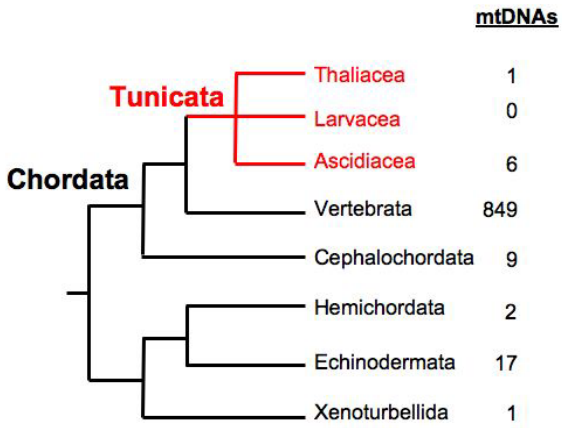

Fig. 1 Number of available complete mitochondrial genomes of deuterostomes, corresponding to deuterostome mtDNA dataset analysed in this review. The three taxonomic classes of tunicates are shown as a polytomy, due to phylogenetic controversies. Deuterostome phylogeny is in accordance with Bourlat *et al.* (2006).

Gene content

Within vertebrates, the most frequent mitochondrial gene content consists of 37 genes encoding for 13 protein subunits of the oxidative phosphorylation complexes, two ribosomal RNAs, and 22 tRNAs necessary for the translational machinery located within mitochondria.

Among the 865 analysed chordates, there are no cases of rRNA gene loss/acquisition, while only four vertebrate species lose/acquire one or two protein-coding genes (Table 2). This is in accordance with observations carried out on metazoan mtDNAs, where the number of protein-coding and rRNA genes changes rarely (Gissi *et al.*, 2008). On the contrary, the tRNA gene content is quite variable between different metazoan phyla and, in general, it depends on the genetic code employed and the taxonomic group (Gissi *et al.*, 2008). According to the relaxed wobble rules in codon-anticodon recognition and to the genetic code observed (Table 1), the mt translation machinery of Chordata should require a minimum set of 22 tRNAs in vertebrates and lancelets, and 23 tRNAs in tunicates. In fact two tRNAs are needed to decode each of the six-fold degenerate amino acids (leucine and serine in all chordates, plus glycine in tunicates), while only one tRNA is needed to decode each remaining amino acid. This tRNA number, predicted on basis of the genetic code, holds for

Table 1 Mitochondrial genetic code of deuterostomes compared to the universal code. "CodTab" indicates the number of the Table containing the entire genetic code, available at the web site "http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi". The genetic code of hemichordates is reported as CodTab 9, although exceptions in the meaning of the AAA codon indicate that this is a new mt genetic code

	UGA	**AUA**	**AGA**	**AGG**	**AAA**	**CodTab**
Universal	Stop	Ile	Arg	Arg	Lys	1
Mitochondrial						
Chordata						
Vertebrata	Trp	Met	Stop	Stop	=	2
Tunicata	Trp	Met	Gly	Gly	=	13
Cephalocordata [a]	Trp	Met	Ser	Ser/ absent	=	5
Hemichordata [b]	Trp	=	Ser	Ser	=	new
Balanoglossus	Trp	=	Ser	absent	absent	9
Saccoglossus	Trp	=	Ser	Ser	=	new
Echinodermata	Trp	=	Ser	Ser	Asn	9
Xenoturbellida	Trp	Met	Ser	Ser	=	5

a: AGG codons have been found only in *Epigonichthys maldivensis* (2 occurrences) and in *Branchiostoma belcheri* (1 occurrence in both the AB078191 and AB083383 mtDNA sequences), while they are absent in the mtDNA of remaining *Branchiostoma* species and of *Asymmetron* genus (7 and 4 mtDNA sequences, respectively).
b: the AC number of the two complete mtDNA sequences of hemichordates are: AF051097 (*Balanoglossus carnosus*) and AY336131 (*Saccoglossus kowalevskii*).

vertebrates and lancelets but not for tunicates, whose most frequent (i.e. "standard") tRNA gene number is 24 and includes an additional tRNA-Met. In fact, all tunicates encode for the common tRNA-Met with CAT anticodon and for a tRNA-Met with the unusual TAT anticodon, probably acting as initiator and elongator tRNA-Met, respectively (Yokobori *et al.*, 1999, 2003, 2005; Gissi *et al.*, 2004; Iannelli *et al.*, 2007a, 2007b). This additional tRNA-Met(TAT) gene is shared only with some mollusc bivalves, thus this feature has arisen independently in two distant animal lineages (Gissi *et al.*, 2008).

Variations from the standard tRNA gene content have been found less frequently in vertebrates and lancelets than in tunicates. As shown in Table 2, only 1.6 % of the 849 vertebrate mtDNAs analysed show loss/acquisition of tRNA genes. Moreover, the loss of tRNA genes is an uncommon event compared to tRNA acquisition, as it has been found only in three of 849 vertebrates (Table 2), suggesting that tRNA loss can not be easily compensated by an analogous function encoded by the nuclear genome. This may be due to differences in tRNA structure or difficulties in mitochondrial import of tRNAs. In addition, differences in tRNA content mainly concern amphibian species, which show acquisition of additional tRNA genes as results of duplication of mt regions and/or extensive gene order rearrangements (Liu *et al.*, 2005; Mueller and Boore, 2005; Zhang *et al.*, 2005; Kurabayashi *et al.*, 2006; San Mauro *et al.*, 2006). Among lancelets, only *Branchiostoma floridae* exhibits an additional tRNA gene (Spruyt *et al.*, 1998), although the modified cloverleaf structure and the presence of a non-canonical TCT anticodon suggest that this is a tRNA-like secondary structure rather than a functional tRNA. Finally, in tunicates the situation is extremely variable (see Table 2): about half of the seven available tunicate mtDNAs show differences from the expected tRNA gene number, moreover loss and acquisition of tRNA genes seem to be equally tolerated. Although the tunicate taxon sampling is still insufficient to draw definitive conclusions, the current data suggest that the tunicate tRNA gene number could vary in a species-specific manner.

Genome architecture

To compare the overall structure of mt genomes, we have introduced a new genomic feature, named genome "architecture" (AR). The genome AR takes into account both gene content and gene order, and it is defined as the order of the entire set of functional mt-encoded genes, including duplicated genes. Therefore, two mtDNAs with different architecture may show differences in gene content and/or gene order. The three chordate subphyla show different level of variability in genome AR, indeed genome AR is quite stable in vertebrates, moderately conserved in cephalochordates and highly variable in tunicates.

In vertebrates, 80 % of the analysed species show the same genome AR as the human mtDNA (hereafter named Std, "standard"), while remaining species exhibit one of the 42 additional genome ARs (Table 3). In particular, the Std architecture is shared by all eutherian and prototherian mammals, as well as by almost all fishes (except for few neopterygian species), and it is lost in all species belonging to Metatheria, Sauropsida (Aves plus Crocodylidae), and Hyperoartia (Table 3). In addition, non-standard ARs are highly similar to the standard AR, as in most cases they can be converted into the standard AR by the translocation of few genes, most of which are located near to the mtDNA replication origins (Boore, 1999).

Table 2 Chordate taxa with a non-standard mitochondrial gene content (different from the expected number of 37 genes in vertebrates and lancelets; 39 genes in tunicates). The AC number of the corresponding mtDNA sequences is also reported. The minus symbol indicates genes encoded by the minor strand. GSA: gene strand asymmetry, calculated as described in Table 4

Taxon (N° available mtDNAs)	AC number	Proteins		tRNA		GSA
		Additional	Lost	Additional	Lost	
Vertebrata (849)						
Amphibia (82)						
Polypedates megacephalus	NC_006408		*atp8, nad5*			0.49
Rhyacotriton variegatus	NC_006331			*trnT*		0.53
Plethodon elongatus	NC_006335			*trnT*		0.53
Mantella madagascariensis	NC_007888			*trnM*		0.53
Gegeneophis ramaswamii	NC_006301				*trnF*	0.50
Fejervarya limnocharis	NC_005055			*trnM*		0.53
Aneides hardii	NC_006338			*trnF, -trnE, trnL(UUR), -trnP*		0.46
Aves (73)						
Diomedea melanophris [a]	NC_007172	*-nad6*		*trnT, -trnP, -trnE*		0.41
Testudines (19)						
Malacochersus tornieri [b]	NC_007700			*trnF*		0.58
Lepidosauria (44)						
Cordylus warreni [c]	NC_005962			*trnT, -trnP*		0.49
Sphenodon punctatus	NC_004815		*nad5*	*trnK*	*trnH, trnT*	0.49
Neopterygii (374)						
Chionodraco rastrospinosus	DQ526431		*nad6*		*trnE*	0.60
Bathygadus antrodes	NC_008222			*trnL(UUR)*		0.53
Cephalochordata (9)						
Branchiostoma floridae	Y16474			*trn(TCT)* [d]		0.51
Tunicata (7)						
Halocynthia roretzi	AB024528			*trnF*		1
Phallusia fumigata	AM292602			*trnI, trnX*	*trnD*	1
Phallusia mammillata	AM292320				*trnD*	1

a: duplication of the region: +*trnT-trnP-nad6-trnE*+CR, where CR corresponds to the control region (Gibb *et al.*, 2007)
b: duplication of the region: +*trnF*+CR (Parham *et al.*, 2006), where CR corresponds to the control region
c: tandem duplication of the region: +*trnT-trnP* (Kumazawa, 2004)
d: non-conventional tRNA with TCT anticodon found in the *B. floridae* entry Y16474 [first published as *Branchiostoma lanceolatum* (Spruyt *et al.*, 1998), this sequence actually belongs to *B. floridae* (Spruyt *et al.*, 1998; Nohara *et al.*, 2005b)]

In Cephalochordates, the genome AR is slightly less conserved, as there are three different genome ARs in 9 distinct species, and the most represented AR has been found in 55 % of the sampled species.

Finally, tunicates show an extreme AR variability, as each species has its own specific genome AR, with no AR shared by two or more mtDNAs. This peculiarity appears even more surprising when considering that the seven sampled tunicates derive from a narrow taxonomic range and also include several congeneric ascidian species. Additional mtDNA sequences of ascidians produced in our laboratory (unpublished data) confirm this extreme AR variability. We have analysed the complete mtDNA of ten additional ascidians, sampled both as closely (congeneric) and distantly related species, and no genome AR is present in more than one species. Moreover, gene order variability at intra-genus level has been observed in all ascidian orders (Aplousobranchiata, Stolidobranchiata and Phlebobranchiata).

A comparative analysis among all chordates shows that no AR is shared by different subphyla when all genes are taken into account. On the contrary, the exclusion of tRNA genes makes the genome architectures (AR-tRNA) much more similar and sometimes identical between different subphyla. For example, the AR-tRNA of lancelets is identical or very similar to the standard vertebrate AR-tRNA depending on the genus (identical in *Branchiostoma* plus *Epigonichthys*; very similar in *Asymmetron*). Similarly, in vertebrates the exclusion of tRNA genes drastically reduces the number of different non-standard architectures (from 42 AR to 12 AR-tRNA; Table 3) and increases the number of taxonomic groups showing a Std architecture. In particular, excluding the tRNA genes, the Std architecture is acquired by Metatheria, Crocodylidae and Hyperoartia species, as well as by many lepidosaurians and amphibians (compare AR and AR-tRNA in Table 3). These observations indicate that the extant genome ARs of vertebrates and cephalochordates arose by modest rearrangements of the same ancestral AR, with modifications mainly concerning tRNA genes.

Table 3 Distribution of the standard (Std) and non-standard (Other) mtDNA architecture (AR) between vertebrate taxonomic groups. AR: genome architecture calculated for all mt genes. AR(-tRNA): genome architecture calculated excluding the tRNA genes. Std: standard AR, identical to the human mtDNA architecture. N_{other} : number of species showing a non-standard mtDNA architecture. Other_fishes: fish group including Polypteridae, Chondrostei, Chondrichthyes, Coelacanthiformes and Dipnoi

Taxon	mtDNA N°	N° AR			N° AR(-tRNAs)	
		Std	Other [a]	N_{other}	Std	Other
Vertebrata	849	1	42	173	1	12
Eutheria	197	1			1	
Prototheria	3	1			1	
Metatheria	22		1	22	1	
Testudines	19	1	2	2	1	1
Lepidosauria	44	1	10	26	1	2 [b]
Aves	73		2	73		2 [b]
Crocodylidae	8		1	8	1	
Amphibia	82	1	16	26	1	4 [b]
Neopterygii	374	1	13	14	1	6 [b]
Other_fishes	23	1			1	
Hyperoartia	2		1	2	1	
Hyperotreti	2	1			1	

a: four non-standard AR are shared by vertebrates belonging to different groups: AR-1 is shared by two neopterygians and two amphibians; AR-2 is shared by all metatherians and one amphibian; AR-3 is shared by one neopterygian and all crocodiles; AR-4 is shared by 72 birds and one lepidosaurian.
b: the same non-standard genome AR(-tRNA) is shared by the following species: 72 Aves, 2 Neopterygii, 2 Lepidosauria and 2 Amphibia

On the contrary, the genome architecture of tunicates does not show any similarity to that of remaining chordates. Even the exclusion of tRNA genes, neither significantly reduces the number of different tunicate genome architectures (6 different AR-tRNA against 7 different AR when considering all genes) nor makes the tunicate mtDNAs more similar in genome architecture to that of remaining chordates. These data remain valid also considering our enlarged and unpublished ascidian dataset.

Overall, these observations suggest a fast rate of mtDNA rearrangements in the tunicate subphylum, with changes equally involving tRNA and other mitochondrial genes, while the strong similarity in AR-tRNA between vertebrates, cephalochordates, and remaining deuterostomes (excluding tunicates) (data not shown) supports the conservation of mtDNA architecture during deuterostome diversification.

Gene strand distribution

A peculiarity of metazoan mtDNA is the asymmetric distribution of genes between the two strands, indeed a major and a minor strand are commonly recognized based on the number of encoded genes. This asymmetric gene distribution can be quantified using the GSA formula (Gene Strand Asymmetry) (Gissi *et al.*, 2008), calculated as the absolute value of the difference in gene number between the two strands, divided by the total gene number. GSA values range from 0 to 1, with values close to zero indicating an almost equal number of genes encoded by the two strands, and values higher than 0.5 corresponding to the presence of at least 75 % of the genes on the major strand.

The overall data on gene strand distribution in metazoans show a prevalence of high GSA values, while a symmetric gene distribution is very rare and it has been observed only in 17 phylogenetically distant species over 1206 complete metazoan mtDNAs (Gissi *et al.*, 2008).

Among chordates, the GSA value is high and almost invariant within each subphylum. In vertebrates, the GSA is equal to 0.51 in almost all taxa, and the minor strand encodes only for one protein-coding gene, *nad6,* and 8 tRNA genes (Table 4). The few exceptions (15 over 849 species) to this structure are related to inversion of a single tRNA gene (only two cases, Table 4) or to changes in gene content (Table 2). In the last cases, all additional duplicated genes are present on the same strand as the ancestral gene that gave rise to the duplication (Table 2). In cephalochordates, the GSA ranges from 0.41 to 0.51, this modest variability being due to the different number of tRNAs encoded by the minor strand (from 8 to 10 tRNA genes, plus only one protein-coding gene, *nad6* or *nad5* depending on the species) (Table 4). Thus, vertebrates and lancelets exhibit a very similar gene strand asymmetry, as consequence of the similar genome architecture (see previous section), but they strongly differ in the tendency to switch the sense strand of a gene. Indeed, lancelets show frequent gene inversions, both of large mt regions or single genes, while vertebrates show only two cases of gene inversion over a total of 849 analysed species (Table 4). In particular, a strand switch of a 2.5 kb region including four genes (*trnL(CUN*), *nad5*, *trnG*, and *nad6*) has been found in the *Asymmetron* genus compared to other lancelets and to the standard genome AR of vertebrates (Nohara *et al.*, 2005b; Kon *et al.*, 2007).

Table 4 Gene strand asymmetry (GSA) of chordate mtDNAs, including data on number and name of genes encoded by the minor strand. For vertebrates, only species with the common mt gene content (37 genes) are reported in this table, and the minor strand genes are indicated as differences from the "Standard" vertebrate situation. Numbers in brackets refer to the number of species showing a given GSA

Taxon (N° species)		GSA [a]	Minor strand genes		
			N°	tRNA	CDS
Vertebrata	Standard (833)	0.51	9	*trnQ, trnA, trnN, trnC, trnY, trnS(UCN), trnE, trnP*	*nad6*
Neopterygii	*Carapus bermudensis*	0.46	10	Standard plus *trnM*	*nad6*
Lepidosauria	*Calotes versicolor*	0.57	8	Standard except *trnP*	*nad6*
Cephalochordata	*Asymmetron inferum*	0.46	10	*trnN, trnA, trnC, trnY, trnS(UCN), trnG, trnL(CUN), trnE, trnP*	*nad5*
Cephalochordata	*Asymmetron lucayanum* (3)	0.41	11	*trnA, trnC, trnY, trnQ, trnN, trnS(UCN), trnG, trnL(CUN), trnE, trnP*	*nad5*
Cephalochordata	*Branchiostoma* (4)	0.51	9	*trnQ, trnN, trnA, trnC, trnY, trnS(UCN), trnE, trnP*	*nad6*
Cephalochordata	*Epigonichthys maldivensis*	0.51	9	*trnQ, trnN, trnA, trnC, trnY, trnS(UCN), trnE, trnP*	*nad6*
Tunicata (7)		1	0		

a: the GSA is calculated as the absolute value of the difference in gene number between the two strands, divided by the total gene number.

In addition, the inversion plus transposition of the *trnQ* gene has been observed between *Asymmetron inferum* and *Asymmetron lucayanus* (Kon *et al.*, 2007). In vertebrates, the two observed cases of gene inversion concern a single tRNA gene (Table 4) adjacent or translocated close to the control region (*trnP* and *trnM*, respectively), thus gene inversions have taken place near to the "hot spot" region of vertebrate genome rearrangements (Boore, 1999). Moreover, these rare events have occurred in two phylogenetically-distant species, a lizard and a fish: the inversion of *trnP,* found in the lizard *Calotes versicolor* and shared by all South Asian draconine agamids, appears incompatible with a tRNA remoulding process, and it has been explained by a homologous DNA recombination helped by accidentally formed inverted repeats (Amer and Kumazawa, 2007); no hypothesis has yet been published to explain the translocation/inversion of *trnM* in the bone fish *Carapus bermudensis* (see GenEmbl accession number AP004404; (Miya *et al.*, 2003)).

In general, gene inversion can not be explained by the tandem duplication/random gene loss model, the mechanism commonly invoked for mtDNA rearrangements in metazoans, but can be easily explained by homologous or illegitimate recombination. Thus, it has been proposed that the sporadic events of gene inversion observed in vertebrates are the result of rare DNA recombination processes, while DNA recombination is significantly high in the mtDNA of cephalochordates, thus promoting frequent gene inversions in this subphylum (Amer and Kumazawa, 2007; Kon *et al.*, 2007). Although the existence of recombination has been traditionally excluded in the mitochondria of metazoans, there is growing evidence for the occurrence of DNA recombination in several taxonomic groups (Lunt and Hyman, 1997; Hoarau *et al.*, 2002; Shao *et al.*, 2005; Tsaousis *et al.*, 2005; Guo *et al.*, 2006) and for the presence of homologous enzymatic recombination activities in vertebrates (Thyagarajan *et al.*, 1996; Kraytsberg *et al.*, 2004).

Differently from other chordates, tunicates have a GSA equal to one, i.e. they show the most extreme gene strand asymmetry. Thus, all genes of tunicates are encoded on the same strand, rendering the "minor" strand devoid of genes. This feature is shared by all tunicate species, including our unpublished dataset of ascidian mtDNAs, making the gene strand asymmetry the only "invariant" mitochondrial feature currently found in tunicates. Therefore, we can suppose that strong functional constraints have forced the tunicate mtDNA to maintain all genes on the same strand, in spite of the numerous changes in genome AR and rearrangements of gene order: the overall transcription mechanism or processes regulating the level of transcription could be the most obvious functional constraints for the maintenance of all genes on the same strand.

Conclusions

The data here summarized underline the existence of a significant variability between the three subphyla of Chordata in several genome-level features.

The genetic code exhibits a different specificity in each of the three subphyla, with cephalochordates retaining a genetic code more similar to that of remaining deuterostomes (Table 1).

Similarities in genome architecture, i.e. in structural features such as gene content, gene order and gene strand distribution, cluster the vertebrates and cephalochordate subphyla to the exclusion of tunicates. In fact, tunicates are characterized by the

most extreme gene strand asymmetry, and a strong variability in genome AR, which is different between species, even when we analyse our unpublished dataset of ten additional ascidian species. On the contrary, cephalochordates are characterized by moderate rearrangements of genome AR, frequent gene inversions, and a gene order and gene strand asymmetry that can be easily traced back to that of vertebrates. Finally, the genome AR of vertebrates is almost stable but not so invariant as previously reported: while gene inversions are very rare in all vertebrates, changes in gene content and gene order are quite frequent in Amphibia and Lepidosauria (see Table 2 and Table 3), and sporadic cases of genome rearrangements have been also observed in a few bony fishes, birds and turtles. Thus, the initial dogma of a frozen genome AR in vertebrates has been broken by the identification of several deviations from the "standard" AR in some vertebrate groups (Table 3). Many of these non-standard genome ARs involve gene duplications or changes in the order of genes located close to the control region or surrounding the L-strand replication origin, suggesting that errors in the mtDNA replication have given rise to these rearrangements. Moreover, the tRNA genes are the most common gene category implicated in genome AR changes.

Although these differences in chordate mtDNA evolutionary dynamics are already manifest, the extent of the surprising features of tunicate mtDNA needs to be further confirmed through the analysis of a large number of ascidian mtDNAs (our research field) and the sampling of the remaining tunicate classes of Thaliacea and Larvacea, for which mtDNA sequences are almost absent.

Abbreviations

cox1, *cox2*, *cox3*: cytochrome oxidase subunit I, II, and III protein genes; *cob*: cytochrome b gene; *atp6*, *atp8*: ATP synthase subunit 6 and 8 genes; *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*: NADH dehydrogenase subunit 1–6, 4L genes; *rrnS*, *rrnL*: ribosomal rRNA of the small and large subunit, respectively. Transfer RNA genes are named as *trn* followed by the one-letter code of the specified amino acid and, in brackets, the recognized codon or the used anticodon.

Acknowledgements

This work was supported by the "Italian Ministry of University and Research" and the University of Milan (COFIN-2006 and FIRST projects to CG). We thank G Pesole and D Horner for valuable suggestions on the manuscript.

References

Amer SA, Kumazawa Y. The mitochondrial genome of the lizard *Calotes versicolo*r and a novel gene inversion in South Asian draconine agamids. Mol. Biol. Evol. 24: 1330-1339, 2007.

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

Boore JL. The use of genome-level characters for phylogenetic reconstruction. Trends Ecol. Evol. 21: 439-446, 2006.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, *et al*. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444: 85-88, 2006.

Cameron CB, Garey JR, Swalla BJ. Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. Proc. Natl. Acad. Sci. USA 97: 4469-4474, 2000.

Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439: 965-968, 2006.

Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol. 24: 269-280, 2007.

Gissi C, Iannelli F, Pesole G. Complete mtDNA of *Ciona intestinalis* reveals extensive gene rearrangement and the presence of an *atp8* and an extra *trnM* gene in ascidians. J. Mol. Evol. 58: 376-389, 2004.

Gissi C, Iannelli F, Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101: 301-320, 2008.

Gissi C, Pesole G. Transcript mapping and genome annotation of ascidian mtDNA using EST data. Genome Res. 13: 2203-2212, 2003.

Guo X, Liu S, Liu Y. Evidence for recombination of mitochondrial DNA in triploid crucian carp. Genetics 172: 1745-1749, 2006.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. Mol. Biol. Evol. 19: 2261-2264, 2002.

Iannelli F, Griggio F, Pesole G, Gissi C. The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. BMC Evol. Biol. 7: 155, 2007a.

Iannelli F, Pesole G, Sordino P, Gissi C. Mitogenomics reveals two cryptic species in *Ciona intestinalis*. Trends Genet. 23: 417-422, 2007b.

Kon T, Nohara M, Yamanoue Y, Fujiwara Y, Nishida M, Nishikawa T. Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences. BMC Evol. Biol. 7: 127, 2007.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, *et al*. Recombination of human mitochondrial DNA. Science 304: 981, 2004.

Kumazawa Y. Mitochondrial DNA sequences of five squamates: phylogenetic affiliation of snakes. DNA Res. 11: 137-144, 2004.

Kurabayashi A, Usuki C, Mikami N, Fujii T, Yonekawa H, Sumida M, *et al*. Complete nucleotide sequence of the mitochondrial genome of a Malagasy poison frog *Mantella madagascariensis*: evolutionary implications on mitochondrial genomes of higher anuran

groups. Mol. Phylogenet. Evol. 39: 223-236, 2006.

Liu ZQ, Wang YQ, Su B. The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. Gene 346: 145-151, 2005.

Lunt DH, Hyman BC. Animal mitochondrial DNA recombination. Nature 387: 247, 1997.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, *et al*. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 26: 121-138, 2003.

Mueller RL, Boore JL. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol. Biol. Evol. 22: 2104-2112, 2005.

Nohara M, Nishida M, Miya M, Nishikawa T. Evolution of the mitochondrial genome in cephalochordata as inferred from complete nucleotide sequences from two *Epigonichthys* species. J. Mol. Evol. 60: 526-537, 2005a.

Nohara M, Nishida M, Nishikawa T. New complete mitochondrial DNA sequence of the lancelet *Branchiostoma lanceolatum* (Cephalochordata) and the identity of this species' sequences. Zoolog Sci. 22: 671-4, 2005b.

Oda H, Akiyama-Oda Y, Zhang S. Two classic cadherin-related molecules with no cadherin extracellular repeats in the cephalochordate amphioxus: distinct adhesive specificities and possible involvement in the development of multicell-layered structures. J. Cell. Sci. 117: 2757-2767, 2004.

Parham JF, Macey JR, Papenfuss TJ, Feldman CR, Turkozan O, Polymeni R, *et al*. The phylogeny of Mediterranean tortoises and their close relatives based on complete mitochondrial genome sequences from museum specimens. Mol. Phylogenet. Evol. 38: 50-64, 2006.

Rowe T. Assembling the tree of life. Oxford University Press, Oxford, UK, 2004.

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

Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238:195-209, 1999.

San Mauro D, Gower DJ, Zardoya R, Wilkinson M. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol. Biol. Evol. 23: 227-234, 2006.

Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. J. Mol. Evol. 60: 764-773, 2005.

Spruyt N, Delarbre C, Gachelin G, Laudet V. Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome: relations to vertebrates. Nucleic Acids Res. 26: 3279-3285, 1998.

Stach T, Turbeville JM. Phylogeny of Tunicata inferred from molecular and morphological characters. Mol. Phylogenet. Evol. 25: 408-428, 2002.

Thyagarajan B, Padua RA, Campbell C. Mammalian mitochondria possess homologous DNA recombination activity. J. Biol. Chem. 271: 27536-27543, 1996.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. Widespread recombination in published animal mtDNA sequences. Mol. Biol. Evol. 22: 925-933, 2005.

Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol. Biol. Evol. 19: 762-776, 2002.

Yokobori S, Oshima T, Wada H. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. Mol. Phylogenet. Evol. 34: 273-283, 2005.

Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, *et al*. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics 153: 1851-1862, 1999.

Yokobori S, Watanabe Y, Oshima T. Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. J. Mol. Evol. 57: 574-587, 2003.

Zeng L, Swalla BJ. Molecular phylogeny of protochordates: chordate evolution. Can. J. Zool. 83: 24-33, 2005.

Zhang P, Zhou H, Liang D, Liu YF, Chen YQ, Qu LH. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. Gene 346: 133-143, 2005.

Zrzavy J, Mihulka S, Kepka P, Bezdek A, Tietz D. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. Cladistics 14: 249-285, 1998.