

Research article

Open Access

## The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level

Fabio Iannelli<sup>1</sup>, Francesca Griggio<sup>1</sup>, Graziano Pesole<sup>1,2</sup> and Carmela Gissi\*<sup>1</sup>

Address: <sup>1</sup>Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Via Celoria 26, 20133 Milano, Italy and <sup>2</sup>Dipartimento di Biochimica e Biologia Molecolare "E. Quagliariello", Università di Bari, Via Orabona 4, 70126 Bari, Italy

Email: Fabio Iannelli - [fabio.iannelli@unimi.it](mailto:fabio.iannelli@unimi.it); Francesca Griggio - [francesca.griggio@unimi.it](mailto:francesca.griggio@unimi.it); Graziano Pesole - [graziano.pesole@biologia.uniba.it](mailto:graziano.pesole@biologia.uniba.it); Carmela Gissi\* - [carmela.gissi@unimi.it](mailto:carmela.gissi@unimi.it)

\* Corresponding author

Published: 31 August 2007

Received: 29 May 2007

*BMC Evolutionary Biology* 2007, **7**:155 doi:10.1186/1471-2148-7-155

Accepted: 31 August 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/155>

© 2007 Iannelli et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Within Chordata, the subphyla Vertebrata and Cephalochordata (lancelets) are characterized by a remarkable stability of the mitochondrial (mt) genome, with constancy of gene content and almost invariant gene order, whereas the limited mitochondrial data on the subphylum Tunicata suggest frequent and extensive gene rearrangements, observed also within ascidians of the same genus.

**Results:** To confirm this evolutionary trend and to better understand the evolutionary dynamics of the mitochondrial genome in Tunicata Ascidiacea, we have sequenced and characterized the complete mt genome of two congeneric ascidian species, *Phallusia mammillata* and *Phallusia fumigata* (Phlebobranchiata, Ascidiidae). The two mtDNAs are surprisingly rearranged, both with respect to one another and relative to those of other tunicates and chordates, with gene rearrangements affecting both protein-coding and tRNA genes. The new data highlight the extraordinary variability of ascidian mt genome in base composition, tRNA secondary structure, tRNA gene content, and non-coding regions (number, size, sequence and location). Indeed, both *Phallusia* genomes lack the *trnD* gene, show loss/acquisition of DHU-arm in two tRNAs, and have a G+C content two-fold higher than other ascidians. Moreover, the mt genome of *P. fumigata* presents two identical copies of *trnI*, an extra tRNA gene with uncertain amino acid specificity, and four almost identical sequence regions. In addition, a truncated cytochrome b, lacking a C-terminal tail that commonly protrudes into the mt matrix, has been identified as a new mt feature probably shared by all tunicates.

**Conclusion:** The frequent occurrence of major gene order rearrangements in ascidians both at high taxonomic level and within the same genus makes this taxon an excellent model to study the mechanisms of gene rearrangement, and renders the mt genome an invaluable phylogenetic marker to investigate molecular biodiversity and speciation events in this largely unexplored group of basal chordates.

## Background

The mitochondrial genome (mtDNA) is considered a model system for studying genome evolution, and a good molecular marker to reconstruct the phylogeny of Metazoa, both at high and low taxonomic level. In general, evolutionary genomics and phylogenetic studies are strictly related, because the understanding of the evolutionary peculiarities of a given character greatly facilitates the evaluation of its suitability as phylogenetic marker. For phylogenetic purposes, the mitochondrial sequences of single or multiple genes or regions are commonly analysed by traditional molecular evolutionary methods. However, besides the sequence itself, genome-level features have been proposed as powerful phylogenetic characters [1]. Gene arrangement, gene content, tRNA secondary structures, and changes in genetic code are all mitochondrial genome-level characters, used to clarify some controversial phylogenetic relationships, especially at high taxonomic levels. However, the phylogenetic reliability of these characters needs to be investigated more in detail, considering the variability of the mitochondrial genome in metazoans.

The mtDNA of Metazoa is a small circular genome, of around 14–16 kb in length, characterized by a compact structure with no introns and short intergenic regions. Only one large non-coding region is typically present in the genome and is known as the control region (CR), because it is assumed that elements controlling replication and transcription of the genome are located in this region, although functional studies have been carried out only in vertebrates and a few other species [2-4]. In most invertebrate mtDNAs where only short non-coding regions are present, the CR has sometimes been tentatively identified through the presence of repeated sequences, secondary structures, or a strong compositional bias, because these features are often associated with the control region [5,6]. The mt gene content is extremely conserved and typically consists of 37 genes encoding two rRNAs, 22 tRNAs, and 13 proteins, all of which are components of the respiratory chain complexes. Variations in gene content are rarely due to changes in the protein-coding gene complement, and mostly concern changes in the content of tRNA genes. Indeed, only the *atp8* gene, encoding for subunit 8 of ATP synthase, is lost in Platyhelminthes [7], and in most mollusc *Bivalvia* [8] and *Nematoda* [9]. On the contrary, drastic changes in tRNA gene number are related to the usage of a modified genetic code [10,11] or are restricted to specific taxonomic groups. Thus, the mtDNA of Cnidaria encodes only for one or two tRNAs (tRNA-Met and/or -Trp) [12,13], whereas the mtDNA of Porifera usually present extra and/or duplicated tRNAs in variable number, but can also have an incomplete tRNA set [14,15]. In addition, the loss or duplication of one or few

tRNA genes occurs sporadically in many taxa, sometimes associated with changes in gene order. Mitochondrial gene arrangement varies both across phyla and within a given phylum, and exhibits a high degree of variability in invertebrates [16], while the vertebrate gene order is almost frozen and highly similar to that of lancelets (Cephalochordata) [17]. Extensive gene rearrangements have been observed in Mollusca [5], Nematoda [18], and some arthropod lineages such as hemipteroid insects [19,20] and several parasitic groups [20-23]. Although gene order is generally conserved at low taxonomic levels, such as within a genus or a family, exceptions have been reported in the most diverse taxonomic groups, even those commonly characterized by a stable gene order. For example, within Vertebrata, salamanders of the family Plethodontidae show extensive gene rearrangements, moreover the presence of pseudogenes and additional gene copies testifies to ancient duplications of large mt regions [24]. Similarly, lancelets of the genus *Epigonichthys* shows an inversion and several gene translocations, explained by recombination and tandem duplications, respectively [17]. Genome rearrangements at intra-genus level have been observed in Mollusca and Arthropoda, such as in marine gastropods of the genus *Dendropoma* [25], bivalves of the genus *Crassostrea* [26], and even between gender-specific mtDNAs of bivalves with doubly uniparental inheritance (DUI). In particular, the extent of rearrangements in DUI mitotypes ranges from several gene inversions, duplications and transpositions in *Inversidens japonensis*, to an almost-identical or identical gene order in *Tapes philippinarum* and *Mytilus galloprovincialis* [27,28]. Among arthropods, chigger mites of the genus *Leptotrombidium* (Acari) show translocation of a tRNA gene and duplications of non-coding and rRNA sequences, in addition to an overall gene order drastically different from the arthropod ancestral gene order [29]. Finally, parasitic flatworms of the genus *Schistosoma* exhibit extensive rearrangements due to translocations of protein-coding and tRNA genes [30,31].

The investigation of gene rearrangements in closely related species, such as congeneric species, can be extremely useful to discover the mechanisms underlying gene order variability, as it allows the observation of recent rearrangements and even the identification of intermediary stages of the process itself. To date, the mtDNA of more than 1000 metazoan species has been sequenced but the taxon sampling is highly biased towards vertebrates and arthropods (771 and 133 species, respectively), with few or no complete genome sequences available for many taxonomic groups (GenEmbl, July 2007). The availability of congeneric mt genomes is also extremely limited. Within Tunicata, traditionally considered the basal subphylum of Chordata (but see also [32]), the mtDNA has been completely sequenced only in one representative

of the class Thaliacea (*Doliolum nationalis*) and four representatives of the class Ascidiacea, including three species of the genus *Ciona*. Surprisingly, substantial gene rearrangements have been observed within the genus *Ciona* [33], and have significantly contributed to demonstrate the existence of two cryptic species in *Ciona intestinalis* [34]. Moreover, ascidian mt genomes show several unusual features compared to other chordates, such as a modified genetic code [10], a fast nucleotide substitution rate [35,36], a gene order extremely rearranged compared to other chordates and metazoans [33,36], the presence of two distinct tRNA-Met genes [37], and the absence of a discernible control region [33]. These features suggest that further investigations on tunicate mtDNA should be of interest. Indeed, the highly variable gene order observed even at intra-genus level in the currently limited tunicate sample suggests that this taxon may be a good model for the investigation of mechanisms of genome rearrangement. Moreover, the unexpected evolutionary dynamics of the mtDNA of tunicates suggests caution in the usage of this molecule to reconstruct the phylogeny of Chordata, which is in turn a prerequisite to understand the origin of Vertebrata [32]. As first step to verify the reliability of tunicate mtDNA as phylogenetic marker in chordates, it is fundamental to increase the number of available sequences.

To investigate the variability of ascidian mtDNA at low taxonomic levels, we have sequenced the complete mtDNA of the two congeneric ascidians *Phallusia fumigata* and *Phallusia mammillata*, and carried out detailed comparisons of the structural and evolutionary features of ascidian mtDNAs at the intra-genus level. The *Phallusia* genus was chosen for this study as it belongs to the same order as *Ciona* (order Phlebobranchiata) but to a different family (Ascidiidae versus Cionidae), thus the divergence between these organisms is adequate to confirm whether the intra-genus evolutionary dynamics observed in *Ciona* are a common ascidian trait.

## Results and discussion

### Gene content

The mitochondrial genome of both *Phallusia* species contains the canonical set of two rRNA genes (*rrnS* and *rrnL*) and 13 protein-coding genes, including the *atp8* gene previously considered absent in tunicates [35]. The complement of tRNA genes is unusual, being different between the two congeneric species and with respect to other ascidians. In accordance with the tunicate mt genetic code [10,35], two tRNA genes for Leu, Ser, and Gly codons have been identified. In addition, tunicate mtDNAs specify two tRNA-Met, one with the common 5'-CAU-3' and the other with the unusual 5'-UAU-3' anticodon, thus the canonical tRNA set for tunicates consists of 24 tRNAs [33,37]. Both tRNA-Met genes have been found in *Phallusia* mtDNAs, confirming the hypothesis that this is a general feature of ascidian mtDNA. However, the tRNA-Asp gene (*trnD*) is lost in both genomes, and in *P. fumigata* two extra tRNA genes have been identified: an additional gene for tRNA-Ile (*trnI-2*), and a *trnX* gene that can be folded into two alternative cloverleaf structures showing different anticodons (see paragraph on tRNA genes). The two copies of *trnI* are almost identical, a situation similar to that found in the mtDNA of the distantly related species *Halocynthia roretzi*, where there are two almost identical copies of *trnF* [38]. In conclusion, 23 tRNA genes are present in the mtDNA of *P. mammillata*, and 25 in *P. fumigata*.

### Genome size and compactness

The mtDNAs of *P. mammillata* and *P. fumigata* are 14579 and 15535 bp long respectively, which is consistent with the typical size of chordate mt genomes. Surprisingly, the mtDNA of *P. mammillata* is 6.5% (956 bp) shorter than that of *P. fumigata*, a somewhat unexpected observation give the almost identical size of the mtDNA of the three *Ciona* species (maximum difference: 1.4% – 199 bp – between *C. intestinalis* sp. A and sp. B).

The more compact structure of the *P. mammillata* mtDNA compared to *P. fumigata* is highlighted not only by the smaller genome size but also by the higher number/per-

**Table 1: Statistics on gene overlaps of ascidian mitochondrial genomes**

Species <sup>a</sup>	Gene overlap			Longest overlap	
	%	Total bp	N° (L ≥ 1)	bp	Gene pair
<i>Phallusia mammillata</i>	0.30	44	6	14	<i>cox2-cob</i>
<i>Phallusia fumigata</i>	0.15	23	3	14	<i>cox2-cob; nad1-nad2</i>
<i>Ciona intestinalis</i> sp. A	0.26	37	3	26	<i>cox2-cob</i>
<i>Ciona intestinalis</i> sp. B	0.25	38	3	29	<i>cox2-cob</i>
<i>Ciona savignyi</i>	0.22	33	3	26	<i>cox2-cob</i>
<i>Halocynthia roretzi</i>	0.23	34	4	13	<i>nad4L-trnM(CAT)</i>

<sup>a</sup> Sequence AC numbers are reported in Additional file 2.

centage of gene overlaps, and the lower percentage of non-coding (NC) sequences. As reported in Table 1, overlapping gene pairs account for 44 bp in *P. mammillata*, and 23 bp in *P. fumigata* (0.30 against 0.15% of the total genome). Thus, there are six and three overlapping gene pairs in *P. mammillata* and *P. fumigata*, respectively, with only two overlapping gene pairs common to both species (*cox2-cob* and *trnL(CUN)-trnT*). In parallel, the percentage of NC sequences is quite low in the more compact mtDNA of *P. mammillata* (1.94%) and higher in *P. fumigata* (5.89%) (Table 2). Distinct from the *Phallusia* genus, the three species of *Ciona* reveal almost identical values of gene overlap (Table 1), and a notably lower %NC only in *C. intestinalis sp. B* (Table 2) [34]. Overall, the involvement of both NC total size and gene overlaps in the determination of mtDNA compactness is demonstrated by the highly significant negative correlation between these parameters in a comparative analysis between all the ascidian mtDNAs ( $r = 0.9$ ).

In conclusion, different evolutionary trends toward compactness have shaped the mtDNAs of the *Phallusia* genus, giving rise to substantial differences in genome size between *P. mammillata* and *P. fumigata*. By way of contrast, the compactness is almost the same in the three mtDNAs of the genus *Ciona*, where mtDNA size variability is mainly due to differences in NC size [34].

### Non-coding regions

No single non-coding region comparable in size to the control region of vertebrates is found in the mtDNA of the two *Phallusia* species.

NC sequences constitute 21–25 short intergenic regions ( $L < 200$  bp), and the NC size distribution is quite different between the two species (Table 2). Indeed, five NC longer than 100 bp are present in *P. fumigata* but none in *P. mammillata*. Moreover, variability in NC size distribution is common to all other analysed ascidians (Table 2). The largest NC region is more than two times longer in *P.*

*fumigata* compared to *P. mammillata*, as observed in *C. savignyi* compared to the two *C. intestinalis* species (Table 2).

The position of NC regions longer than 40 bp is not conserved in ascidians. Indeed, only one NC region is located exactly in the same position in the two *Phallusia* species (NC region between *trnC-nad4L*). Similarly, only one NC region is located in a similar but not identical position in the three available *Ciona* species (NC region downstream *nad1*).

No significant sequence similarity has been found between NC sequences of the two *Phallusia* species, neither between any of these NC and those of other ascidians.

Sequence similarities between NC of the same mtDNA have been found only in *P. fumigata* (see Figure 1, where NC are named according to their size). In particular, the block NC(123 bp)-*trnI2*-NC(125 bp) shows a 95% identity over 208 bp to the block NC(109 bp)-*trnI1-rrnS* (first 38 bp), including an identical sequence 165 bp long (Figure 2). The NC(134 bp) located between *cox1* and *trnG(GGN)* contains a sequence 29 bp-long duplicated in the region downstream *nad1* (Figure 3). Moreover, this 29 bp-long sequence constitutes a repeat located upstream and downstream of the gene pair *trnG(GGN)-nad1*, similarly to the situation found in *C. intestinalis sp. A*, where a completely different sequence, 25 bp-long, flanks the gene pair *trnK-nad1* [33]. Thus, in both *C. intestinalis sp. A* and *P. fumigata*, a completely different duplicated sequence is located upstream and downstream a gene pair including a tRNA gene (*trnK* or *trnG-GGN*) and *nad1*. These sequences can be explained as remnants of gene rearrangement events involving *nad1* and occurred independently in the two species, or as sequences that promoted the rearrangement event. Alternatively, they could also be unknown functional sequences involved in the regulation of *nad1* expression.

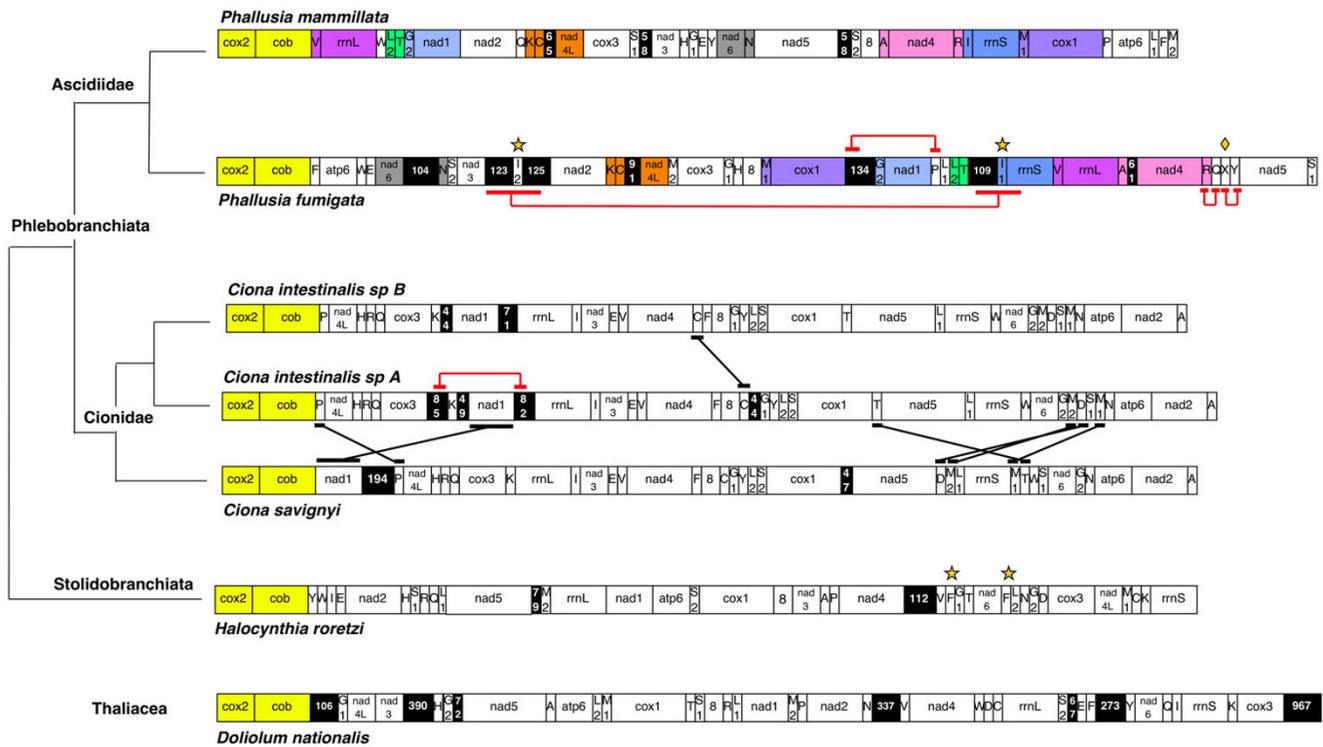
**Table 2: Statistics on non-coding (NC) sequences of ascidian mitochondrial genomes**

Species <sup>a</sup>	% NC (bp)	N <sup>b</sup>	NC length distribution			Longest NC	
			20–39 bp	40–99 bp	> 99 bp	bp	Location <sup>c</sup>
<i>Phallusia mammillata</i>	1.94 (283)	21	0	3	0	65	<i>trnC-nad4L</i>
<i>Phallusia fumigata</i>	5.89 (915)	25	2	2	5	134	<i>cox1-trnG(GGN)</i>
<i>Ciona intestinalis sp. A</i>	2.9 (429)	22	2	4	0	85	<i>nad1-rrnL</i>
<i>Ciona intestinalis sp. B</i>	1.8 (263)	23	1	2	0	71	<i>cox3-trnK</i>
<i>Ciona savignyi</i>	2.9 (428)	25	1	1	1	194	<i>nad1-trnP</i>
<i>Halocynthia roretzi</i>	3.09 (456)	29	5	1	1	112	<i>nad4-trnV</i>

<sup>a</sup> Sequence AC numbers are reported in Additional file 2.

<sup>b</sup> Number of non-coding sequences, including spacers 1 bp long.

<sup>c</sup> The location of NC sequences is defined by flanking genes.



**Figure 1**  
**Gene organization of tunicate mitochondrial genomes, and taxonomic classification of the analysed species.**  
 Gene blocks conserved between *P. mammillata* and *P. fumigata* are reported in different colours, with yellow indicating the gene pair conserved in all analysed tunicates. Sequences almost identical in the same mtDNA are underlined and linked by a red line. Genes transposed among the three *Ciona* species are underlined and linked by a black line. Non-coding (NC) regions equal or longer than 40 bp are indicated by black background, with numbers corresponding to their size (in bp). Abbreviations for protein-coding and rRNA genes are as in the main text, except for *atp8* gene (abbreviation: 8). Transfer RNA genes are indicated according to the transported amino acid, except for: G1: Gly(AGR); G2: Gly(GGN); L1: Leu(UUR); L2: Leu(CUN); M1: Met(AUG); M2: Met(AUA); S1: Ser(AGY); S2: Ser(UCN). Diamond: *trnX* gene of *P. fumigata*. Star: intra-genome duplicated genes, that is *trnI* genes in *P. fumigata* (named I1 and I2), and *trnF* in *Halocynthia roretzi*. Feature table of *Halocynthia roretzi* mtDNA is as reported in [38].

Additional elements characterize the long NC regions of *P. fumigata*: two palindromes 8 bp long are found in the NC upstream of the duplicated *trnI* genes (see Figure 2); a tRNA-like secondary structure, with sequence similarity to the beginning of tRNA-Pro, has been found in the NC-134 bp (Figure 3). It is striking that palindromes or tRNA-like structures have often been identified in the control region of metazoans, which suggests they may be involved in the mt replication or transcription processes.

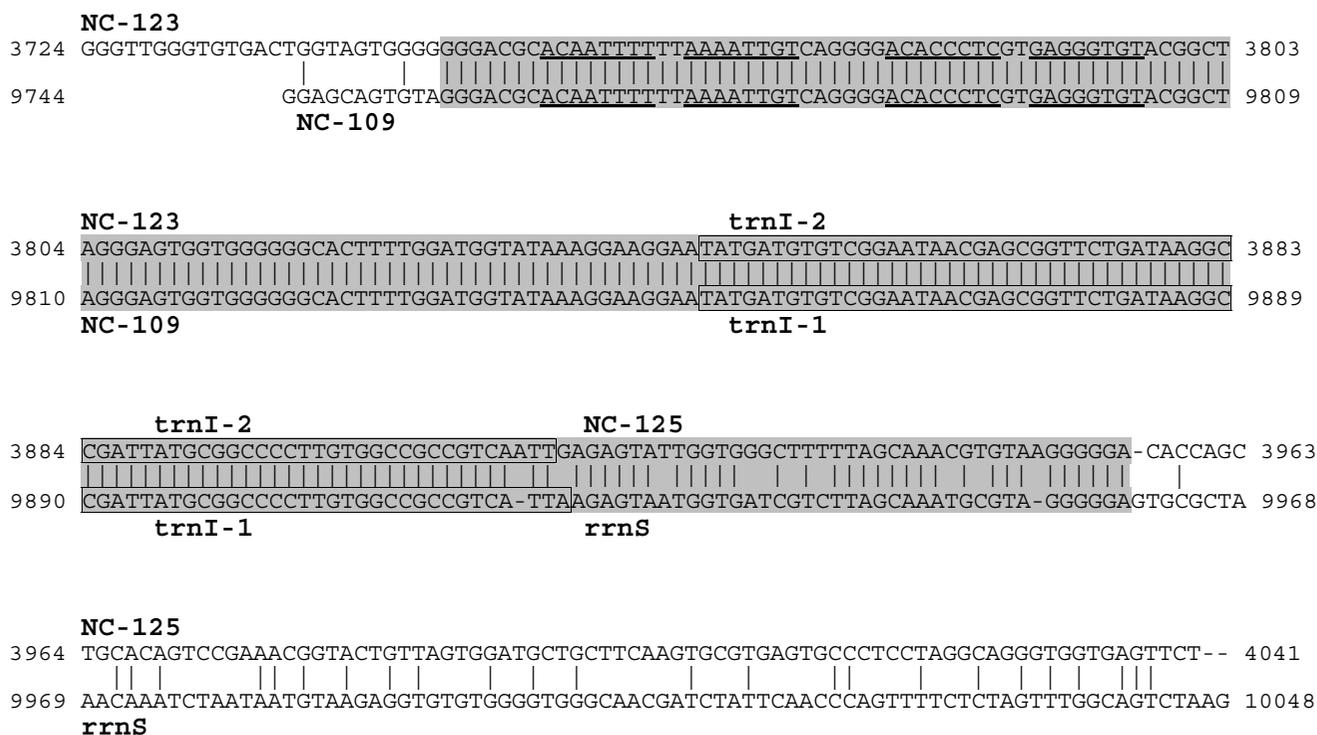
In conclusion, NC regions appear highly variable in number, size, sequence and location in the *Phallusia* as in the *Ciona* genus suggesting that a high variability in NC is quite common between closely related ascidian species. The extensive reorganization of NC regions could partially be related to the extensive changes in gene order, indeed a process of "duplication-random gene loss" can generate

NC fossil sequences, as remnants of duplicated genes erased by evolution.

**Gene order**

In both *Phallusia* species all genes are transcribed from the same strand, as in other tunicates so far studied.

Gene order is extremely rearranged between *P. mammillata* and *P. fumigata*: only nine gene blocks, including about half of the total mitochondrial genes, are conserved in both species, and the largest of these blocks consists of only three genes (Figure 1). The breakpoint analysis confirms this observation (Table 3), as the relative breakpoint distance is very high and quite close to the value expected for random permutations (0.99). Moreover, gene rearrangements are not due to a high mobility of tRNA genes, indeed omitting the tRNAs the relative breakpoint distance between *Phallusia* species increases (Table 3) and



**Figure 2**  
**Similarity between the two *trnI* sequences of *P. fumigata*, including flanking non-coding (NC) regions.** A gray background highlights the 208 bp sequence with 95% identity between the two regions, with *trnI-1* and *trnI-2* sequences boxed. Gene abbreviations are as in the main text. Non-coding regions are named according to their length (in bp). Underlined sequences indicate inverted repeats.

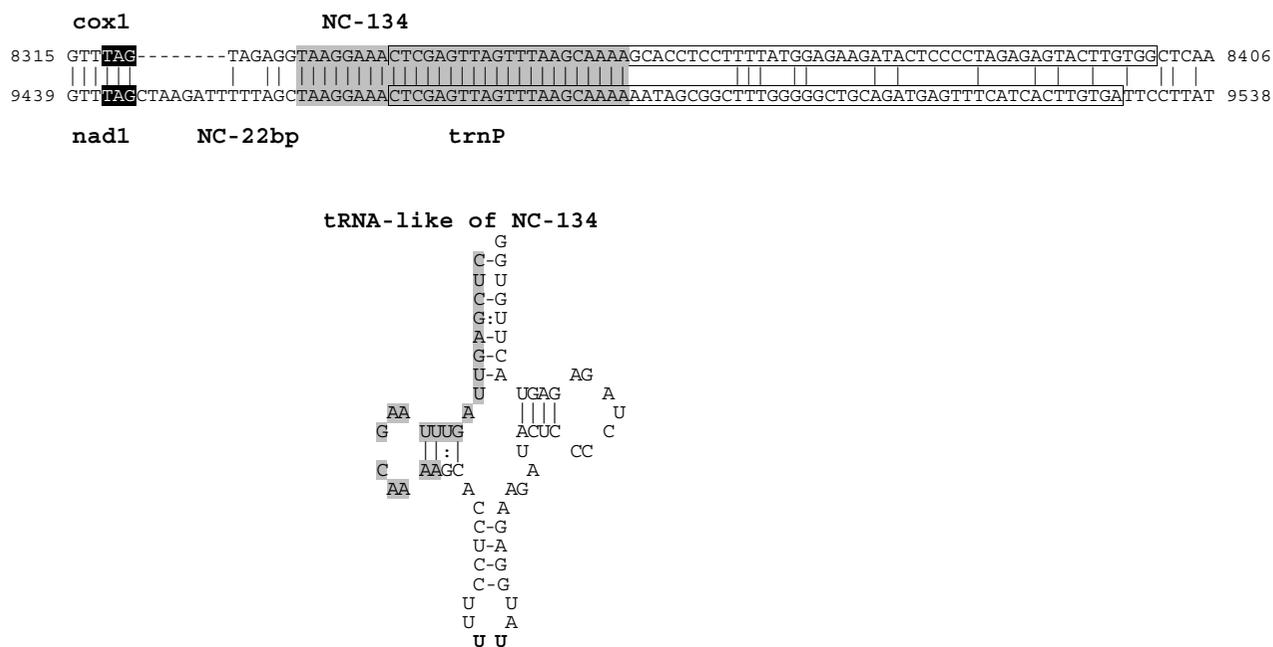
only two gene blocks appear conserved (*nad2-nad4L-cox3*, and the previously identified *cox2-cob* gene pair, accounting for one-third of non-tRNA genes).

Gene rearrangements in the *Phallusia* appear to be more extensive than those observed in the *Ciona* and in other congeneric protochordate pairs (Table 3). As reported in Figure 1, only the tRNA-Cys gene (*trnC*) is translocated in *C. intestinalis sp. A* compared to *C. intestinalis sp. B* [34], whereas between *C. intestinalis sp. A* and *C. savignyi* there are six translocated genes (*nad1* and five tRNA genes) [33]. Thus, the ancestral gene order state could be reconstructed quite easily for the *Ciona* genus but not for *Phallusia*, where several duplication/transposition steps will be required to relate the two gene orders. On the contrary, gene order is almost invariant in lancelets (Table 3), with translocations and inversions found only in the genus *Epigonichthys*, specifically only in *E. lucayanus* compared to all other lancelets studied [17].

Gene order comparisons among all tunicate mtDNAs, both with and without tRNA genes, do not identify large gene blocks shared by most species. Thus, including all

genes, only *cox2-cob* is conserved in all tunicates, probably because of the ORF overlap (see below), and *trnR-trnQ* is conserved in almost all ascidians except for *P. mammillata* (Figure 1). Excluding tRNA genes, no additional or largest gene blocks appear conserved in all tunicates or ascidians, and only *nad4L-cox3* is shared by the order Phlebobranchiata (i.e. *Ciona* and *Phallusia* genera).

The absence of a consensus gene order for ascidian mtDNA complicates comparisons with the almost frozen gene order of vertebrates and cephalochordates [16,17]. The few instances of gene blocks shared by tunicates and other-chordates appear due to fortuitous similarities, rather than to a common evolutionary origin, as indicated by their presence in only one or few tunicates. For example, the *trnV-rrnL* gene pair is shared by vertebrates, cephalochordates, and *Phallusia* species (Phlebobranchiata), but it is absent in all remaining tunicates including other phlebobranchies of the genus *Ciona* (see Figure 1). Similarly, the gene block *rrnS-trnV-rrnL* is shared only by vertebrates and *P. fumigata*. Although these gene blocks are quite primitive, as indicated by their presence even in other deuterostomes, their occurrence only in the *Phallu-*



**Figure 3**  
**Similarity between the non-coding region 134 bp-long (NC-134) and the trnP gene of *P. fumigata*, together with the tRNA-like structure assumed by NC-134.** A gray background indicates the 29-bp identical sequences, and boxed sequences identify tRNA and tRNA-like structures. In the tRNA-like structure, canonical and G-U base pairing are differently indicated. Stop codons are in black background.

*sia* genus among all tunicates suggests that this may be the result of multiple gene rearrangements, rather than the remnant of a primitive gene order.

Although complete mtDNA sequences are still underrepresented for ascidians, it is evident that gene rearrangements are frequent in this group, and quite common at

low taxonomic levels at least in the order Phlebobranchiata. It should be verified if intra-genus rearrangements are frequent even in ascidians belonging to the remaining orders of Stolidobranchiata and Aplousobranchiata. The evolutionary trend of ascidians toward gene rearrangements is completely different from other chordates, and even from other metazoans. In most metazoans, mito-

**Table 3: Relative breakpoint distance and sequence divergence for several intra-genus pairs of protochordates**

Genus	Species-1	Species-2	Breakpoint <sup>a</sup>		Protein-genes <sup>b</sup>		Protein % id <sup>c</sup>	rRNA <sup>d</sup>		
			All	-tRNA	dN	dS		rrnS	rrnL	
Asciacea	<i>Phallusia</i>	<i>mammillata</i>	<i>fumigata</i>	0.71	0.80	0.182	56.60	72.36	0.452	0.412
	<i>Ciona</i>	<i>intestinalis</i> sp. A	<i>intestinalis</i> sp. B	0.08	0	0.055	3.25	88.79	0.113	0.147
	<i>Ciona</i>	<i>intestinalis</i> sp. A	<i>savignyi</i>	0.33	0.20	0.156	129.21	72.66	0.403	0.295
	<i>Ciona</i>	<i>intestinalis</i> sp. B	<i>savignyi</i>	0.41	0.20	0.169	83.0	71.42	0.396	0.310
Cephalochordata	<i>Branchiostoma</i>	<i>floridae</i>	<i>lanceolatum</i>	0	0	0.087	3.12	85.84	0.269	0.287
	<i>Branchiostoma</i>	<i>belcheri</i>	<i>lanceolatum</i>	0	0	0.097	3.94	84.26	0.257	0.236
	<i>Branchiostoma</i>	<i>floridae</i>	<i>belcheri</i>	0	0	0.110	3.75	82.54	0.289	0.263
	<i>Epigonichthys</i>	<i>lucyanus</i>	<i>maldivensis</i>	0.22	0.33	0.171	10.29	74.87	0.405	0.402

<sup>a</sup> Relative breakpoint distance calculated dividing the pairwise breakpoint distance by the number of genes shared by corresponding genomes, for a dataset including all mt genes (All), and a dataset without tRNA genes (-tRNA).

<sup>b</sup> dN (nonsynonymous substitutions) and dS (synonymous substitutions) calculated on the 13 mt protein-coding genes according to the codon substitution model of Goldman and Yang [73].

<sup>c</sup> Percentage of amino acid identity calculated on the 13 mt-encoded proteins

<sup>d</sup> Substitutions per sites calculated according to the GTR model [75].

chondrial gene order has been demonstrated to evolve in a non-clocklike, saltatory manner, and changes are commonly described as rare, unique events unlikely to be subjected to reversion or convergent evolution [1,16,39]. Therefore, this genome-level character has been considered highly reliable as a phylogenetic marker (review in [1]). In ascidians, the frequent gene rearrangements could impair the reliability of this feature as a phylogenetic marker. This high frequency of rearrangement could also imply differences in the ascidian gene rearrangement mechanisms compared to other metazoans. Gene rearrangements are commonly explained by a process of "tandem duplication-random gene loss" [40], assuming that duplicated portions of the mtDNA can be generated by several processes, and that supernumerary copies of each gene will be selected at random to be subsequently eliminated [41]. A few cases of mt gene rearrangements have been explained by different mechanisms, such as a tandem duplication process followed by non-random gene loss based on gene transcriptional polarities, suggesting that functional constraints could overcome the random process of gene loss [42]. Finally, intra- and inter-genome illegitimate recombination has also been observed [43], or hypothesized [44] in organisms such as nematodes and mites. As simple speculation, the high frequency of rearrangements in ascidians could be due to a completely different mechanism, or to the exacerbation of one of the previously hypothesized mechanisms (i.e. due to an increased activity of enzymes already involved in the process or due to the acquisition of new enzymatic activities). The "duplication-random gene loss" model does not fit well the gene rearrangements of the *Phallusia* genus, because a large number of gene movements are required to explain the observed gene order, however it is still compatible with the situation found in the *Ciona*. Thus, the high frequency of gene rearrangements observed in ascidians could be partially related to the existence of long divergence times between morphologically closely related species. Moreover, the absence of data on the replication process and the location of control region in ascidian mtDNA prevents further speculations on the rearrangement process. Indeed, in most metazoan mtDNAs the replication origins act as hot spots for gene rearrangements [16], suggesting that components of the replication machinery are involved in the step of segmental duplication.

#### Sequence divergence

Table 3 reports the sequence divergence between the two *Phallusia* species and other congeneric pairs of protochordates, calculated separately for rRNA genes, synonymous (dS) and nonsynonymous sites (dN) of protein-coding genes, and amino acid sequences. As expected, dN values and percentage of amino acid identity are negatively correlated ( $r = 0.98$ ), and a strong positive correlation is

observed between dN and rRNA sequence divergence ( $r = 0.97$  and  $0.88$  for *rrnS* and *rrnL*, respectively), indicating that different mt genes follow the same evolutionary trends. Synonymous substitutions appear saturated in all intra-genus comparisons, as indicated by dS values higher than one (Table 3). Both nonsynonymous sites and rRNA genes indicate a sequence divergence between *P. mammillata* – *P. fumigata* highly similar to that observed between lancelets *E. lucayanus* – *E. maldivensis*, and in comparisons between *C. intestinalis* – *C. savignyi*. Assuming a constant evolutionary rate, a similar divergence time can be inferred for the above-mentioned species pairs, whereas a more recent divergence can be hypothesized for the species pair *C. intestinalis* sp. A – *C. intestinalis* sp. B, which shows the lowest dN and rRNA divergence values (Table 3).

In the protochordate sample here analysed, there is no statistically significant correlation between the relative breakpoint distance and the evolutionary rate of rRNA or protein-coding genes, although in arthropods a general correlation between the rate of genome rearrangement and the rate of nucleotide substitution has been found [5,45]. However in our small sample, species pairs with the highest gene rearrangements also exhibit the highest sequence divergence in both Cephalochordata and Ascidiacea (Table 3).

Considering the single mitochondrial-encoded proteins, COX1 is the most conserved protein (93.8% identity in *Phallusia*), whereas ATP8, NAD6 and NAD2 are among the less conserved proteins in all analysed ascidians (NAD6 is the least conserved protein in *Phallusia*, with 45.1% identity).

#### Base composition

As reported in Table 4, the G+C content of *P. mammillata* and *P. fumigata* mtDNAs is significantly higher than that of other ascidians (average values  $46.9 \pm 0.2\%$  in *Phallusia*, and  $24.8 \pm 4.7$  in other ascidians). The trend toward a GC-rich genome is even more pronounced at the third codon position (P3) of protein-coding genes, where the G+C content is  $51.5 \pm 1.1\%$  in *Phallusia* and  $15.8 \pm 5.9\%$  in remaining ascidians. This high G+C content is due to the strong increase of G in *Phallusia* compared to remaining species (Table 4). Indeed, in other ascidians the most common base is T, and the T/G ratio is always higher than one, ranging from 2.4 to 8.2 in P3. On the contrary, in both *Phallusia* species the G content is almost as high as the T content, thus the T/G ratio is close to one and becomes lower than one at P3 in *P. fumigata*. At intra-genus level, base composition is constant in *Phallusia* and quite variable in *Ciona*, as shown by the standard deviation of the average GC content calculated on P3 in intra-

**Table 4: Compositional features of tunicate mitochondrial genomes**

	G+C%		T/G ratio		%T		%G		%A		%C		GC-skew		AT-skew	
	All <sup>a</sup>	P3 <sup>b</sup>	All	P3	All	P3	All	P3	All	P3	All	P3	All	P3	All	P3
<b>Asciacea</b>																
<i>Phallusia mammillata</i>	47.1	50.7	1.18	1.12	33.3	32.1	28.2	28.6	19.5	17.2	18.9	22.1	0.20	0.13	-0.26	-0.30
<i>Phallusia fumigata</i>	46.8	52.2	1.07	0.86	33.3	30.7	31.1	35.5	19.9	17.1	15.7	16.7	0.33	0.36	-0.25	-0.29
<i>Ciona intestinalis</i> sp. A	21.4	10.5	3.72	8.18	44.4	49.3	11.9	6.0	34.2	40.2	9.5	4.4	0.12	0.15	-0.13	-0.10
<i>Ciona intestinalis</i> sp. B	23.5	15.5	3.47	5.95	43.5	46.9	12.5	7.9	33.1	37.6	10.9	7.7	0.07	0.01	-0.14	-0.11
<i>Ciona savignyi</i>	22.7	13.0	3.22	4.99	45.3	51.9	14.1	10.4	32.0	35.2	8.7	2.6	0.24	0.61	-0.17	-0.19
<i>Halocynthia roretzi</i>	31.7	24.1	1.90	2.38	44.0	50.0	23.2	21.0	24.3	26.0	8.5	3.1	0.46	0.74	-0.29	-0.32
<b>Thaliacea</b>																
<i>Doliolum nationalis</i>	39.0	39.4	1.64	1.74	32.8	30.9	20.0	17.8	28.3	29.7	19.0	21.7	0.03	-0.10	-0.07	-0.02

<sup>a</sup> All: whole mitochondrial genome.

<sup>b</sup> P3: third codon position of the 13 protein-coding genes.

genus comparisons ( $51.5 \pm 1.1\%$  in the *Phallusia* genus and  $13 \pm 2.5\%$  in the *Ciona* genus).

An asymmetric distribution of complementary bases between the two strands is observed in all ascidians: on the coding strand, the GC-skew is positive and the AT-skew is negative, with absolute values of both skews quite low in most species (Table 4). Surprisingly, the compositional asymmetry can be different even between species with a similar G+C content. Indeed, in the *Phallusia* genus the GC-skew at P3 is quite different between the two species (0.13 and 0.36), although the G+C content is almost the same (see Table 4).

The high variability of base composition in ascidians has a strong influence on synonymous codon usage and even on amino acid composition of mitochondrial proteins. Thus, leucine is the most abundant amino acid in ascidians (average  $15.1\% \pm 0.7\%$ ) as in other metazoan mt proteins but the usage of the two codon families (CTN and TTR) is very different between ascidian species (see standard deviation: CTN =  $4.8\% \pm 3.4\%$ ; TTR =  $10.2\% \pm 2.9\%$ ). Similarly, AGR and GGN glycine codon families are used differently between species (see standard deviation: GGN =  $6.2\% \pm 4.8\%$ ; AGR =  $4.7\% \pm 2.1\%$ ), as consequence of the different G content. Strong differences in amino acid abundance are also observed for amino acids encoded by quartet codons, such as Ala-GCN ( $3.9\% \pm 1.4\%$ ) and Val-GTN ( $9.7\% \pm 3.1\%$ ), and some amino acid encoded by AT-rich codons (Asn-AAY, Lys-AAR, and Ile-ATY). In all these cases, the standard deviation of the amino acid content exceeds 30% of the average value (data not shown).

#### Protein-coding genes

As already reported, *cox2* and *cob* genes overlap in both *Phallusia* species and in all other tunicate mtDNAs so far sequenced (length overlap from 11 to 29 bp). This gene overlap could be responsible of the linkage between *cox2*

and *cob* (Figure 1), preventing gene rearrangements involving these two genes. EST analyses have demonstrated that *cox2-cob* genes are transcribed to a single bicistronic mature mRNA in *C. intestinalis* sp. A and in *H. roretzi* [38]. Thus, a bicistronic mature mRNA for *cox2-cob* can be hypothesized even in *Phallusia*. The conservation of this overlap could be due to the need to produce a bicistronic mRNA or to a functional constraint at protein level, such as the need to preserve a specific amino acid pattern in the upstream ORF (i.e. in *cox2*). A careful analysis of the overlap region shows that the sequence included in both ORFs encodes for an amino acid region not conserved at the C-terminal region of COX2, but well conserved at the N-terminal of the COB protein (the downstream gene), where an arginine is present in all analysed deuterostomes. Thus, a functional constraint at protein level for this gene overlap can be excluded.

In *P. mammillata*, there are two additional cases of overlapping ORFs: *nad1-nad2*, with a 14 bp overlap, and *nad4L-cox3*. In the latter case, a complete TAA stop codon for *nad4L* is located 233 bp inside *cox3*, giving rise to a NAD4L protein showing a longer C-terminal region with no similarity to other tunicate NAD4L. Assuming the existence of an incomplete stop codon, a NAD4L with no frame overlap and length consistent to the homolog of *P. fumigata* can be inferred [see Additional file 1]. As in the case of *cox2-cob*, a frame overlap for *nad4L-cox3* implies the presence of a bicistronic mature mRNA, whereas an incomplete stop codon implies the synthesis of distinct polyadenylated transcripts and the presence of a polyadenylation signal downstream *nad4L*. It is interesting that even in *H. roretzi*, *nad4L* gene exhibits a complete stop codon well inside the downstream tRNA gene, that is 76 bp downstream of the predicted incomplete stop codon [38].

Most *Phallusia* genes are inferred to have a complete termination codon, except for *cob* of *P. fumigata* and *nad4L* of *P. mammillata* [see Additional file 1]. Such abbreviated stop codons are assumed to be completed to a TAA codon during the process of maturation and polyadenylation of precursor polycistronic transcripts, as demonstrated by experimental data in mammals [46] and by "in silico" analyses of mt ESTs in two ascidians [38]. In these cases, a complete stop codon is often located a few bases downstream of the incomplete one, as a mechanism to prevent a translational readthrough in case of incorrect transcript maturation [47]. Curiously, this situation does not hold in *Phallusia*, given that complete stop codons are well inside the downstream gene in both *cob* and *nad4L* genes where an incomplete stop codon has been found (position 20 of the downstream *trnF* gene for *cob* of *P. fumigata*, and position 233 of the downstream *cox3* gene for *nad4L* of *P. mammillata*).

As expected from the high G content (Table 4), the preferred stop codon is TAG in *P. fumigata* (10 genes) while in *P. mammillata* TAG and TAA are equally used as stop codons [see Additional file 1]. Similarly, the preferred start codon is GTG (8 in *P. fumigata* and 5 in *P. mammillata*), and most inferred start codons end with a G (12 in *P. fumigata*, and 9 in *P. mammillata*) [see Additional file 1].

### Cytochrome b

The cytochrome b protein (COB) of *Phallusia* and other tunicates is significantly shorter than in other deuterostomes, indeed the COB mean length is  $363 \pm 3$  amino acids in tunicates, against  $381 \pm 4$  in other deuterostomes (see Additional file 2 for species list). This length difference is due to the lack of a carboxy-terminal sequence of 10–24 amino acids, which is highly conserved from vertebrates to echinoderms and even in *Xenoturbella*. The lost region include the ENK consensus sequence, with a negatively charged amino acid (E) conserved in all analysed deuterostomes, and a positively charged amino acid in the last position of the consensus (K or R in 15 over 23 sequences) (Figure 4). COB is a membrane-spanning protein whose structure has been widely studied both at evolutionary and functional level [48–50]. This protein consists of three functional domains, each evolving at a different rate: an intermembrane, a transmembrane, and a matrix domain. The intermembrane domain is the one that evolves most slowly, because it contributes structurally to the Qo redox center. The transmembrane domain consists of eight hydrophobic helices, including those binding the two heme prosthetic groups [51], and is characterized by a preponderance of conservative amino acid changes [48]. The matrix domain includes three extramembrane loops protruding into the mitochondrial matrix, the N-terminal and C-terminal segments of the protein. This domain is poorly conserved, with amino acid

replacements fitting neutral evolutionary expectations and reflecting the relative lack of function of this region [52]. Moreover, a previous analysis of more than 800 COB from bacteria to mammals has shown that the last 100 C-terminal residues of this protein are highly variable and that the immediately upstream H helix is one of the two least conserved transmembrane helices [50]. The COB fragment lost in tunicates is the C-terminal segment of the matrix domain, thus the tunicate protein terminates with the hydrophobic transmembrane H helix (Figure 4). The poor conservation of this region in COB of other organisms is compatible with a full functionality of this truncated protein. Moreover, the C-terminal segment following the H-helix is lost also in COB of trypanosomes [50], and greatly shortened or lost in some arthropods, nematodes and molluscs (data not shown).

### rRNA genes

*rrnS* is of similar size in the two *Phallusia* species (738 and 724 bp in *P. mammillata* and *P. fumigata*, respectively), whereas *rrnL* is significantly longer in *P. fumigata* than in *P. mammillata* because of the presence of an additional 120 bp-long sequence at the 5' end (1275 and 1147 bp in *P. mammillata* and *P. fumigata*, respectively). However, the additional *rrnL* sequence can be also considered as a non-coding region located upstream *rrnL* (i.e. between *trnV* and *rrnL*). Indeed, a long NC region has been also annotated upstream *rrnL* in *C. intestinalis* sp. A and sp. B, where the 5' end of this gene was tentatively determined based on sequence conservation between orthologous genes of the three available *Ciona* species [33,34]. In general, the lack of transcripts and experimental data on mt rRNAs prevents the exact identification of the rRNA boundaries, and leaves open the possibility of the existence of non-coding regions flanking these genes.

The rRNA length of *Phallusia* species confirms a previous observation showing that tunicates have some of the shortest rRNAs among deuterostomes and even among metazoans [33], their size being comparable to the shortest rRNAs of nematodes and platyhelminthes [53]. Indeed, the average size of tunicate rRNAs is  $695 \pm 35$  for *rrnS*, and  $1134 \pm 68$  for *rrnL*, against an average rRNA size in other deuterostomes of  $864 \pm 33$  for *rrnS*, and  $1461 \pm 93$  for *rrnL* (species listed in Additional file 2).

### tRNA genes

The mtDNA of both *Phallusia* species lacks the tRNA-Asp gene (*trnD*). We sought this gene without success on both genome strands, even as an unusual cloverleaf structure with only a standard anticodon arm, or as a common cloverleaf structure with a non canonical Asp anticodon (5'-VYC-3' instead of 5'-GUC-3'). The absence of the *trnD* is inconsistent with the *Phallusia* codon usage, indeed the percentage of GAY codons for Asp is similar between *Phal-*

P_fumigata	342	YTHLAQVSG.FLYFALIALFVCL.....	363
P_mammillata	342	FTHLAQVSG.FLYFSLIILFVWL.....	363
C_intestinalis_A	340	FIFMAQISS.FIYFFFLFLMMII.....	361
C_intestinalis_B	341	FIFMAQISS.FIYFFFLFLLMVI.....	362
C_savignyi	340	FIYMAQISS.FIYFFFLILMLI.....	361
Halocynthia	343	YVYMAQVCGCPLFYFVIYVYCLSFFV.....	369
Doliolum	338	YVYLSQICS.VLYFGLIGGMLAL.....	359
Homo	347	FTIIGQVAS.VLYFTTILILMPTISLIENKMLKWA.....	380
Chrysemys	348	FITIGQVAS.ILYFSTLILIPAGVIENKMLT.....	379
Squalus	348	FILVGQIAS.VTYFSLFLIIPLTGWENKMLNLN.....	381
Latimeria	347	FMTIGQIAS.ITYFSLFLILIPMTGWLENKAMNWN.....	380
Eptatretus	348	FILMSQISS.FIYFFIFVFMFPLISRNEKILEFHAMYS.....	385
Lampetra	351	FILMTQIAS.TVYFMIFILIFPILGRLENKILLSKNTGKFNWNLT.....	396
B_floridiae	347	YIFLGOAAS.VIYFVNILLIPIVGYVENKLLFS.....	379
B_lanceolatum	348	YIFLGOVAS.FFYFFNILLIPMVGYAENKLLFH.....	380
E_lucayanus	348	YITLGOVSS.FLYFFNILLIPVGLLENKLLFK.....	380
E_maldivens	348	YIFLGOVAS.FFYFFNILLIPVGLVENELLFK.....	380
Balanoglossus	348	YIILGQLS.AIYFLCFVLIPVGLLENRLNLNPL.....	382
Saccoglossus	348	FIEIGQIAS.SIFFALFVIIIPSIHALENKLLYNSW.....	382
Acanthaster_p	348	FILLGQIAS.TAYFLFLFAIPLTSELENKLIF.....	379
Asterias	348	FILLGQISS.ILYFMLFILFIPITAQVENNLIF.....	379
Asterina	348	FILLGQISS.IAYFALFFLAFPLASELENKLIF.....	379
Pisaster	348	FILLGQISS.ILYFMLFILFIPITAQIENNLIF.....	379
Florometra	348	YIELSQVVS.FFYFSIFFIFFPLISYLENFFIFN.....	380
Gymnocrinus	348	YIELGOAVS.VFYFSLFIFWLPFTSYLENLLFIE.....	380
Paracentrotus	348	FVLIGQIAS.LLYFSLFIFGFPLVSSLENKMIFS.....	380
Cucumaria	348	FIIIGQVAS.ALYFSIFIIFFPFVATLENTIIKI.....	380
Ophiopholis	347	YIVLGOLAS.VFYFVGFIFGLNPLISYSEKLFADNIFRGVRYPR.....	389
Ophiura	348	FIVLGOLAS.VYFIFNFLIIMPLIGAFEKEMWNN.....	380
Xenoturbella	348	YITLGOVFS.MLYFLFFLTA.PMINMLEKKLSLKQ.....	380

\*\*\*

**Figure 4**

**Alignment of the C-terminal region of cytochrome b.** All available tunicate species and representatives of remaining deuterostomes are shown. Identical and conserved residues are shown by reverse contrast with black and dark gray background, respectively. Similar residues are indicated by light gray background. The "ENK" consensus sequence of non-ascidian proteins is highlighted by asterisks. Numbers refer to amino acid position. Species classification and AC number of the analysed sequences are listed in Additional file 2.

*lusia* and other ascidians ( $1.83 \pm 0.01\%$  in *Phallusia* against  $1.83 \pm 0.13\%$  in other ascidians). Moreover, the GAT codon is preferred over GAC, as expected from compositional data (see Table 4). The mitochondrial import of a nuclear-encoded tRNA-Asp could functionally compensate the lack of *trnD* in the mtDNA. Alternatively, it is possible that a highly modified mitochondrial *trnD*, with a transcript matured through RNA-editing, has not been identified in the mtDNA because of its unusual secondary structure. Both mechanisms, tRNA import in mitochondria and tRNA-editing by anticodon modification, have been observed in marsupial mtDNA [54-56].

In both *Phallusia* species, three tRNAs (-Asn, -Ser(AGY), and -Cys) show an unusual cloverleaf structure [see Additional file 3 and Additional file 4].

In tRNA-Asn, the spacer between the DHU and anticodon (AC) stems is 2 nt long, instead of the canonical 1 nt [57]. This configuration also holds for tRNA-Asn of the stolidobranch *H. roretzi* and the thaliacean *D. nationalis* but not for other phlebobranchs (i.e *Ciona*).

The tRNA-Ser(AGY) retains a DHU-arm with a stem 3 bp long and a large loop (11 nt long in *P. mammillata* and 15 nt *P. fumigata*), while in most metazoans the DHU-arm is replaced by a loop of variable size [57]. This peculiar cloverleaf structure has not been found in other tRNA-

Ser(AGY) of analysed tunicates and deuterostomes [see Additional file 2], and it is probably a derived character of the *Phallusia* genus or of a narrow ascidian group.

The tRNA-Cys lacks the DHU arm, which is replaced by a loop 5 nt long, and shows several mismatches in stem base pairing compared to remaining tRNAs (4 total mismatches in *P. fumigata* and 7 in *P. mammillata*) [see Additional file 3 and Additional file 4]. Among analysed deuterostomes [see Additional file 2], this feature is shared by ascidians of the order Phlebobranchiata (*Ciona* and *Phallusia*) and by cephalochordates. Moreover, literature data report that the DHU arm of tRNA-Cys is sporadically lost also in vertebrates such as acrodont lizards [58] and the tuatara *Sphenodon punctatus* (Rhynchocephalia) [59]. Thus, current data suggest that the ancestral secondary structure of tRNA-Cys in deuterostome was the standard one, and that the DHU arm was lost independently in multiple lineages. Moreover, the non-canonical structure exhibited by tRNA-Cys even in some protostomes and flatworms [30,60-62] could indicate a tendency of this tRNA to better tolerate alterations of the cloverleaf structure.

It is remarkable that in ascidians unusual tRNA cloverleaf structures with lack/acquisition of the DHU-stem have been found in tRNA genes differing for the last nucleotide of the anticodon sequence (5'-GCU-3' in tRNA-Ser(AGY) and 5'-GCA-3' in tRNA-Cys). Exchanges in amino acid specificity of these tRNAs by a single nucleotide mutation in the anticodon or relaxed constraints related to the anticodon specificity could explain this observation.

As previously reported, the mtDNA of *P. fumigata* presents two *trnI*, and a *trnX* gene encoding for a tRNA with uncertain anticodon specificity (Figure 1).

The two copies of *trnI* are about 6 kb apart and differ only by two nucleotides at the 3' end of the amino acid (AA) stem: *trnI-1* is located in a position similar to the orthologous gene of *P. mammillata* (upstream *rrnS*, see Figure 1), thus it is likely to be the ancestral *trnI* gene; *trnI-2* is in the middle of a large duplicated sequence, embedded between two non-coding regions (see Figure 2), thus probably arose by a recent gene duplication event. The region including *trnI-2* reported in Figure 2 could be the result of a "duplication-random gene loss" mechanism, with the gene loss step not having occurred yet.

The *trnX* gene is part of the tRNA cluster *trnR-trnQ-trnX-trnY*, and contains a sequence 30 bp-long showing 93% identity to a portion of the downstream *trnY* gene (Figure 5). With slightly different boundaries, the *trnX* gene can be folded into two canonical cloverleaf secondary structures with a distinct amino acid specificity (Figure 5): a

tRNA isoacceptor for Tyr (5'-GUA-3' anticodon), as expected by the sequence similarity to *trnY*; and a tRNA isoacceptor for Thr, with the 5'-UGU-3' anticodon overlapped to the anticodon of the alternative tRNA-Tyr structure. In both cloverleaf structures the anticodon is preceded by U and followed by the purine A, and there are only few mismatches in the stems (see Figure 5), thus both structures are compatible with functional tRNAs. Moreover, the low sequence similarity of *trnX* of *P. fumigata* with both *trnY* and *trnT* of *P. mammillata* suggests that *trnX* is not the true ortholog of these genes but a newly acquired gene (data not shown).

In *P. fumigata* a similarity like that observed between *trnX* and *trnY* is found also in the upstream gene pair *trnR-trnQ* (Figure 1). Indeed, *trnR* contains a sequence 34 bp-long showing 85% identity to the downstream *trnQ* gene, including a 17 bp-long identical sequence (Figure 5). In other ascidian mtDNAs, the genes *trnR* and *trnQ* do not share any identical long sequence (i.e. > 7 bp), however the *trnR-trnQ* gene arrangement is conserved in all available ascidians, except for *P. mammillata*. These data could suggest that an alloacceptor gene recruitment mechanism [63,64] gave origin to these tRNA genes in *P. fumigata*, consistent with the existence of a single nucleotide difference between the anticodons of *trnR* (UCG) and *trnQ* (UUG). However, evolutionary reconstructions based on tRNA sequence similarity are unable to support this hypothesis, because of the low resolution of most nodes of the tree (data not shown).

## Conclusion

The newly determined mtDNAs of *Phallusia* species and the comparative data reported here highlight the strong differences in mtDNA evolutionary dynamics between ascidians and remaining chordates.

Apart from the genetic code [10], our comparative analyses identify and/or confirm some mt features as distinctive of ascidians, and possibly of all tunicates, that is: same transcriptional polarities of all genes; presence of two *trnM* genes with different anticodon specificity; short rRNAs, comparable in size to the shortest mt rRNAs of nematodes and platyhelminthes; and a "truncated" cytochrome b protein lacking the C-terminal tail, which usually protrudes into the mitochondrial matrix.

The most striking ascidian features is the high frequency of gene order rearrangements, which is observed even in congeneric ascidians of two distinct genera (*Phallusia* and *Ciona*) and is not restricted to the tRNA genes, which often show an unstable mt genomic position. This extreme plasticity may represent a valuable source of data to study the mechanisms of gene rearrangement. Indeed, the analysis of gene order variability at low taxonomic distances



the complete mtDNA was amplified by long-PCR in several overlapping fragments, in order to avoid the amplification of short mitochondrial pseudogenes sometimes present in the nuclear genome (Numts). Long-PCR was carried out with Expand High Fidelity PCR System (Roche Applied Science), Expand Long Template PCR System (Roche Applied Science), or LA-Taq enzyme (TaKaRa), using several heterologous primers designed on the most conserved mt ascidian sequences: large subunit rRNA (*rrnL*), cytochrome b (*cob*) and cytochrome c oxidase subunit (*cox1*, *cox2*, and *cox3*) genes. Given the absence of information on the mtDNA gene order in *Phallusia* species, several combinations of these heterologous primers were initially used in PCR reactions, and only those reactions that gave a bright single band during electrophoretic analysis were further processed. Species-specific primers were designed on the sequences of the first successfully amplified and sequenced fragments, and used to amplify the remaining regions of the genome. In summary, the mtDNA of *P. mammillata* was amplified in three long overlapping fragments (size range: 2.5 – 9.4 kb), whereas the mtDNA of *P. fumigata* was amplified in five overlapped fragments (size range: 1.2 – 5 kb). Moreover, in both species an additional fragment was amplified to confirm the overlap surrounding the *rrnL* gene. Amplified fragments, primers sequences and PCR conditions are reported in Additional file 5. Fragments shorter than 3 kb were directly sequenced, after purification with the Montage PCR Filter Units (Millipore), using a primer walking strategy (MWG Sequencing Service). Longer fragments were re-amplified in short overlapping segments (information available on request), and then sequenced by primer walking.

The complete mtDNA sequences of *P. mammillata* and *P. fumigata* were deposited at EMBL database under accession numbers [AM292320](#) and [AM292602](#), respectively.

Protein-coding gene annotations were based on sequence similarity to orthologous genes of ascidians. ATG and non-standard initiation codons reported in Wolstenholme [6] were considered as reliable start codons. The start of a protein-coding gene was inferred as the initiation codon that does not cause overlap with the upstream gene and maximizes the similarity between proteins of the two *Phallusia* species. Whenever possible, incomplete stop codons were inferred in order to avoid overlap with the downstream genes.

Transfer RNA genes were identified by their potential cloverleaf secondary structures. The tRNAscan-SE v1.23 program (UNIX version) [65] was locally run with relaxed parameters specifically defined for ascidians (used options: -O -y -X 0 -L 10 -g ascidiancode); the PatSearch program [66] was used to search for very unusual tRNA

structures, such as those lacking an arm. All putative tRNAs were manually checked, considering the reliability of secondary structure and the alignment with known ascidian tRNAs. Thus, transfer RNA identity was specified by anticodon sequence, and sequence similarity.

Boundaries of rRNA genes of small and large ribosomal subunit (*rrnS* and *rrnL*, respectively) were inferred from the boundaries of flanking genes.

Compositional features and codon usage similarities were calculated with Codontree [67]. The GC- and AT-skews, which indicate compositional differences between the two strands, were calculated according to the formulae of Perna and Kocher [68]

Exact direct repeats longer than 10 bp were identified using the Reputer program package [69].

Comparative analyses were carried out on complete mtDNAs of deuterostomes [see Additional file 2]. In addition to *Phallusia*, 49 sequences were analysed, including 10 representatives of Vertebrata, and all available species of Agnatha (4 species), Cephalochordata (11), Tunicata (5), Hemichordata (2), Echinodermata (16), and Xenoturbellidae (1).

Alignments of the 13 mitochondrial proteins were performed with CLUSTAL W v1.82 [70] and manually revised, then the equivalent nucleotide (nt) alignments were "back-aligned" from the protein data. Transfer RNA genes were aligned based on their secondary structure.

Gene order rearrangements were analysed on two datasets, with or without tRNA genes, and the extra genes of *P. fumigata* (*trnI-2* and *trnX*) were excluded from all analyses. The GeneSyn program [71] allowed the detection of gene strings conserved in chordate mtDNAs. The BPAnalysis program [72] was used to calculate the breakpoint distances between pairs of genomes with identical gene content (i.e. the number of gene adjacencies present in one genome but absent in the other). This breakpoint measure was selected because it makes no assumptions about the rearrangement mechanisms. Given the differences in gene content between species and between datasets, a relative breakpoint distance was calculated dividing the breakpoint distance by the number of genes shared by each pair of analysed genomes.

The number of nonsynonymous (dN) and synonymous (dS) substitutions per site were calculated according to the codon substitution model of Goldman and Yang [73], using the Codeml program of the PAML v3.15 package [74]. The program was run on the gap-free concatenated alignment of the 13 protein-coding genes, with codon fre-

quencies calculated from the average nucleotide frequencies at the three codon positions (option CodonFreq = 2), and maximum likelihood (ML) estimations in pairwise comparisons (option RunMode = -2). Sequence divergence of rRNA genes was calculated on gap-free alignments with the Baseml program (PAML v3.15 package) [74] according to the GTR nucleotide substitution model [75].

### Authors' contributions

FI amplified, sequenced and annotated the complete mtDNA of *P. mammillata*; FG amplified, sequenced and annotated the complete mtDNA of *P. fumigata*. In addition, FI and FG participated in the bioinformatics analyses. GP participated in the study design, and critically revised the manuscript. CG conceived the study, supervised the research, carried out most bioinformatics analyses and drafted the manuscript. All authors have provided critical reviews of the manuscript content, read and approved the final manuscript.

### Additional material

#### Additional file 1

Length, start and stop codons of the mitochondrial protein-coding genes of the two *Phallusia* species. Length, start and stop codons of the mitochondrial protein-coding genes of the two *Phallusia* species.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-155-S1.doc>]

#### Additional file 2

Deuterostome mitochondrial genomes analysed in this study. Accession numbers of the deuterostome mtDNA sequences analysed in this study. Organism classification is also reported.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-155-S2.doc>]

#### Additional file 3

Putative secondary structures of tRNAs encoded by *Phallusia fumigata* mtDNA. Putative secondary structures of tRNAs encoded by *P. fumigata* mtDNA. Canonical and G-U base pairing are differently indicated. Yellow background indicates overlapped sequences belonging to adjacent tRNA genes. Nucleotides in square brackets indicate differences in tRNA-Ile2 compared to the reported tRNA-Ile1 sequence.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-155-S3.pdf>]

#### Additional file 4

Putative secondary structures of tRNAs encoded by *Phallusia mammillata* mtDNA. Putative secondary structures of tRNAs encoded by *P. mammillata* mtDNA. Canonical and G-U base pairing are differently indicated. Yellow background indicates overlapped sequences belonging to adjacent tRNA genes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-155-S4.pdf>]

### Additional file 5

Primer sequences and PCR conditions used to amplify the two *Phallusia* mtDNAs. Primer sequences and PCR conditions used to amplify the whole mtDNA of the two *Phallusia* species in several overlapped fragments.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-155-S5.doc>]

### Acknowledgements

This work was supported by the "Italian Ministry of University and Research" and the University of Milano (COFIN-2004 and FIRST projects to CG). We thanks Giuliana Zega, Roberta Pennati and Fiorenza De Bernardi of the Department of Biology (University of Milano) for the gift of *Phallusia* specimens, Francesco Mastrototaro for helpful discussion on ascidians, David Horner and an anonymous reviewer for comments on the manuscript.

### References

- Boore JL: **The use of genome-level characters for phylogenetic reconstruction.** *Trends Ecol Evol* 2006, **21**:439-446.
- Saito S, Tamura K, Aotsuka T: **Replication origin of mitochondrial DNA in insects.** *Genetics* 2005, **171**:1695-1705.
- Cantatore P, Roberti M, Loguercio Polosa P, Mustich A, Gadaleta MN: **Mapping and characterization of Paracentrotus lividus mitochondrial transcripts: multiple and overlapping transcription units.** *Curr Genet* 1990, **17**:235-245.
- Christianson TW, Clayton DA: **In vitro transcription of human mitochondrial DNA: accurate termination.** *Proc Natl Acad Sci USA* 1986, **83**:6277-6281.
- Serb JM, Lydeard C: **Complete mtDNA sequence of the North American freshwater mussel, Lampsilis ornata (Unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in Bivalvia (Mollusca).** *Mol Biol Evol* 2003, **20**:1854-1866.
- Wolstenholme DR: **Animal mitochondrial DNA: structure and evolution.** *International Rev Cytol* 1992, **141**:173-216.
- von Nickisch-Roseneck M, Brown WM, Boore JL: **Complete sequence of the mitochondrial genome of the tapeworm Hymenolepis diminuta: gene arrangements indicate that Platyhelminths are Eutrochozoans.** *Mol Biol Evol* 2001, **18**:721-730.
- Dreyer H, Steiner G: **The complete sequences and gene organization of the mitochondrial genomes of the heterodont bivalves Acanthocardia tuberculata and Hiatella arctica - and the first record for a putative Atpase subunit 8 gene in marine bivalves.** *Front Zool* 2006, **3**:13.
- Lavrov DV, Brown WM: **Trichinella spiralis mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement related to those of coelomate metazoans.** *Genetics* 2001, **157**:621-637.
- Durrheim GA, Corfield VA, Harley EH, Ricketts MH: **Nucleotide sequence of cytochrome oxidase (subunit III) from the mitochondrion of the tunicate Pyura stolonifera: evidence that AGR encodes glycine.** *Nucleic Acids Res* 1993, **21**:3587-3588.
- Telford MJ, Herniou EA, Russell RB, Littlewood DT: **Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms.** *Proc Natl Acad Sci U S A* 2000, **97**:11359-11364.
- Beaton MJ, Roger AJ, Cavalier-Smith T: **Sequence analysis of the mitochondrial genome of Sarcophyton glaucum: conserved gene order among octocorals.** *J Mol Evol* 1998, **47**:697-708.
- Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL: **Naked corals: skeleton loss in Scleractinia.** *Proc Natl Acad Sci U S A* 2006, **103**:9096-100. Epub 2006 Jun 5.
- Wang X, Lavrov DV: **Mitochondrial genome of the homoscleromorph Oscarella carmela (Porifera, Demospongiae)**

- reveals unexpected complexity in the common ancestor of sponges and other animals. *Mol Biol Evol* 2007, **24**:363-373.
15. Erpenbeck D, Voigt O, Adamski M, Adamska M, Hooper JN, Worheide G, Degnan BM: **Mitochondrial diversity of early-branching metazoa is revealed by the complete mt genome of a haplosclerid demosponge.** *Mol Biol Evol* 2007, **24**:19-22.
  16. Boore JL: **Animal mitochondrial genomes.** *Nucleic Acids Res* 1999, **27**:1767-1780.
  17. 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* 2005, **60**:526-537.
  18. Hu M, Gasser RB: **Mitochondrial genomes of parasitic nematodes-progress and perspectives.** *Trends Parasitol* 2006, **22**:78-84. Epub 2005 Dec 27..
  19. Shao R, Campbell NJ, Schmidt ER, Barker SC: **Increased rate of gene rearrangement in the mitochondrial genomes of three orders of hemipteroid insects.** *Mol Biol Evol* 2001, **18**:1828-1832.
  20. Thao ML, Baumann L, Baumann P: **Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha).** *BMC Evol Biol* 2004, **4**:25.
  21. Shao R, Campbell NJ, Barker SC: **Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, Heterodoxus macropus (Phthiraptera).** *Mol Biol Evol* 2001, **18**:858-865.
  22. Covacin C, Shao R, Cameron S, Barker SC: **Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta).** *Insect Mol Biol* 2006, **15**:63-68.
  23. Tjensvoll K, Hodneland K, Niilsen F, Nylund A: **Genetic characterization of the mitochondrial DNA from Lepeophtheirus salmonis (Crustacea; Copepoda). A new gene organization revealed.** *Gene* 2005, **353**:218-230.
  24. Mueller RL, Boore JL: **Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders.** *Mol Biol Evol* 2005, **22**:2104-2112.
  25. Rawlings TA, Collins TM, Bieler R: **A major mitochondrial gene rearrangement among closely related species.** *Mol Biol Evol* 2001, **18**:1604-1609.
  26. Milbury CA, Gaffney PM: **Complete mitochondrial DNA sequence of the eastern oyster Crassostrea virginica.** *Mar Biotechnol (NY)* 2005, **7**:697-712.
  27. Mizi A, Zouros E, Moschonas N, Rodakis GC: **The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel Mytilus galloprovincialis: implications for the doubly uniparental inheritance mode of mtDNA.** *Mol Biol Evol* 2005, **22**:952-967.
  28. Passamonti M, Boore JL, Scali V: **Molecular evolution and recombination in gender-associated mitochondrial DNAs of the Manila clam Tapes philippinarum.** *Genetics* 2003, **164**:603-611.
  29. Shao R, Barker SC, Mitani H, Takahashi M, Fukunaga M: **Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus Leptotrombidium (Acari: Acariformes).** *J Mol Evol* 2006, **63**:251-261.
  30. Littlewood DT, Lockyer AE, Webster BL, Johnston DA, Le TH: **The complete mitochondrial genomes of Schistosoma haematobium and Schistosoma spindale and the evolutionary history of mitochondrial genome changes among parasitic flatworms.** *Mol Phylogenet Evol* 2006, **39**:452-467.
  31. Le TH, Blair D, Agatsuma T, Humair PF, Campbell NJ, Iwagami M, Littlewood DT, Peacock B, Johnston DA, Bartley J, Rollinson D, Herniou EA, Zarlenga DS, McManus DP: **Phylogenies inferred from mitochondrial gene orders: a cautionary tale from the parasitic flatworms.** *Mol Biol Evol* 2000, **17**:1123-1125.
  32. Delsuc F, Brinkmann H, Chourrou D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439**:965-968.
  33. 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* 2004, **58**:376-389.
  34. Iannelli F, Pesole G, Sordino P, Gissi C: **Mitogenomics reveals two cryptic species in Ciona intestinalis.** *Trends Genet* 2007, **23**(9):419-422.
  35. Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K: **Complete DNA sequence of the mitochondrial genome of the ascidian Halocynthia roretzi (Chordata, Urochordata).** *Genetics* 1999, **153**:1851-1862.
  36. 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* 2005, **34**:273-283.
  37. 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* 2003, **57**:574-587.
  38. Gissi C, Pesole G: **Transcript Mapping and Genome Annotation of Ascidian mtDNA Using EST Data.** *Genome Res* 2003, **13**:2203-2212.
  39. Rokas A, Holland PW: **Rare genomic changes as a tool for phylogenetics.** *Trends Ecol Evol* 2000, **15**:454-459.
  40. Moritz C, Brown WM: **Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA.** *Science* 1986, **233**:1425-1427.
  41. Boore JL: **The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals.** In *Comparative Genomics Volume 1*. Edited by: Sankoff D and Nadeau JH. Dordrech/Boston/London, Kluwer Academic; 2000:133-147.
  42. Lavrov DV, Boore JL, Brown WM: **Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss.** *Mol Biol Evol* 2002, **19**:163-169.
  43. Lunt DH, Hyman BC: **Animal mitochondrial DNA recombination.** *Nature* 1997, **387**:247.
  44. 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* 2005, **60**:764-773.
  45. Xu W, Jameson D, Tang B, Higgs PG: **The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes.** *J Mol Evol* 2006, **63**(3):375-392.
  46. Ojala D, Montoya J, Attardi G: **tRNA punctuation model of RNA processing in human mitochondria.** *Nature* 1981, **290**:470-474.
  47. Boore JL: **Complete mitochondrial genome sequence of the polychaete annelid Platynereis dumerilii.** *Mol Biol Evol* 2001, **18**:1413-1416.
  48. Howell N: **Evolutionary conservation of protein regions in the protonmotive cytochrome b and their possible roles in redox catalysis.** *J Mol Evol* 1989, **29**:157-169.
  49. Irwin DM, Kocher TD, Wilson AC: **Evolution of the cytochrome b gene of mammals.** *J Mol Evol* 1991, **32**:128-144.
  50. Degli Esposti M, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A: **Mitochondrial cytochrome b: evolution and structure of the protein.** *Biochim Biophys Acta* 1993, **1143**:243-271.
  51. Howell N, Gilbert K: **Mutational analysis of the mouse mitochondrial cytochrome b gene.** *J Mol Biol* 1988, **203**:607-618.
  52. McClellan DA, McCracken KG: **Estimating the influence of selection on the variable amino acid sites of the cytochrome B protein functional domains.** *Mol Biol Evol* 2001, **18**:917-925.
  53. Hu M, Chilton NB, Gasser RB: **The mitochondrial genomes of the human hookworms, Ancylostoma duodenale and Necator americanus (Nematoda: Secernentea).** *Int J Parasitol* 2002, **32**:145-158.
  54. Borner GV, Morl M, Janke A, Paabo S: **RNA editing changes the identity of a mitochondrial tRNA in marsupials.** *Embo J* 1996, **15**:5949-5957.
  55. Dorner M, Altmann M, Paabo S, Morl M: **Evidence for import of a lysyl-tRNA into marsupial mitochondria.** *Mol Biol Cell* 2001, **12**:2688-2698.
  56. Janke A, Paabo S: **Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition.** *Nucleic Acids Res* 1993, **21**:1523-1525.
  57. Kumazawa Y, Nishida M: **Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics.** *J Mol Evol* 1993, **37**:380-398.
  58. Macey JR, Schulte JA 2nd, Larson A: **Evolution and phylogenetic information content of mitochondrial genomic structural**

- features illustrated with acrodont lizards. *Syst Biol* 2000, **49**:257-277.
59. Seutin G, Lang BF, Mindell DP, Morais R: **Evolution of the WANCY region in amniote mitochondrial DNA.** *Mol Biol Evol* 1994, **11**:329-340.
  60. Shao R, Aoki Y, Mitani H, Tabuchi N, Barker SC, Fukunaga M: **The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years.** *Insect Mol Biol* 2004, **13**:219-224.
  61. Campbell NJ, Barker SC: **The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region.** *Mol Biol Evol* 1999, **16**:732-740.
  62. Podsiadlowski L, Carapelli A, Nardi F, Dallai R, Koch M, Boore JL, Frati F: **The mitochondrial genomes of *Campodea fragilis* and *Campodea lubbocki* (Hexapoda: Diplura): High genetic divergence in a morphologically uniform taxon.** *Gene* 2006, **381**:49-61.
  63. Lavrov DV, Lang BF: **Transfer RNA gene recruitment in mitochondrial DNA.** *Trends Genet* 2005, **21**:129-133.
  64. Rawlings TA, Collins TM, Bieler R: **Changing identities: tRNA duplication and remodeling within animal mitochondrial genomes.** *Proc Natl Acad Sci U S A* 2003, **100**:15700-15705.
  65. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**:955-964.
  66. Pesole G, Liuni S, D'Souza M: **PatSearch: a pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance.** *Bioinformatics* 2000, **16**:439-450.
  67. Pesole G, Attimonelli M, Liuni S: **A backtranslation method based on codon usage strategy.** *Nucleic Acids Res* 1988, **16**:1715-1728.
  68. Perna NT, Kocher TD: **Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes.** *J Mol Evol* 1995, **41**:353-358.
  69. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R: **REPuter: the manifold applications of repeat analysis on a genomic scale.** *Nucleic Acids Res* 2001, **29**:4633-4642.
  70. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
  71. Pavesi G, Mauri G, Iannelli F, Gissi C, Pesole G: **GeneSyn: a tool for detecting conserved gene order across genomes.** *Bioinformatics* 2004, **20**:1472-4. Epub 2004 Feb 19..
  72. Blanchette M, Bourque G, Sankoff D: **Breakpoint Phylogenies.** *Genome Inform Ser Workshop Genome Inform* 1997, **8**:25-34.
  73. Goldman N, Yang Z: **A codon-based model of nucleotide substitution for protein-coding DNA sequences.** *Mol Biol Evol* 1994, **11**:725-736.
  74. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**:555-556.
  75. Saccone C, Lanave C, Pesole G, Preparata G: **Influence of base composition on quantitative estimates of gene evolution.** *Methods Enzymol* 1990, **183**:570-583.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

