

## REVIEW

# Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species

C Gissi<sup>1</sup>, F Iannelli<sup>1</sup> and G Pesole<sup>2,3</sup>

<sup>1</sup>Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Milano, Italy; <sup>2</sup>Dipartimento di Biochimica e Biologia Molecolare 'E. Quagliariello', Università di Bari, Bari, Italy and <sup>3</sup>Istituto Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Bari, Italy

The mitochondrial genome (mtDNA) of Metazoa is a good model system for evolutionary genomic studies and the availability of more than 1000 sequences provides an almost unique opportunity to decode the mechanisms of genome evolution over a large phylogenetic range. In this paper, we review several structural features of the metazoan mtDNA, such as gene content, genome size, genome architecture and the new parameter of gene strand asymmetry in a phylogenetic framework. The data reviewed here show that: (1) the plasticity of Metazoa mtDNA is higher than previously thought and mainly due to variation in number and location of *tRNA* genes; (2) an exceptional trend towards stabilization of genomic features occurred in deuterostomes and was exacerbated in vertebrates, where gene content, genome architecture and gene strand

asymmetry are almost invariant. Only tunicates exhibit a very high degree of genome variability comparable to that found outside deuterostomes. In order to analyse the genomic evolutionary process at short evolutionary distances, we have also compared mtDNAs of species belonging to the same genus: the variability observed in congeneric species significantly recapitulates the evolutionary dynamics observed at higher taxonomic ranks, especially for taxa showing high levels of genome plasticity and/or fast nucleotide substitution rates. Thus, the analysis of congeneric species promises to be a valuable approach for the assessment of the mtDNA evolutionary trend in poorly or not yet sampled metazoan groups. *Heredity* (2008) **101**, 301–320; doi:10.1038/hdy.2008.62; published online 9 July 2008

**Keywords:** mtDNA; metazoans; evolution; comparative genomics; congeneric species

## Introduction

The mitochondrial genome (mtDNA) of Metazoa is regarded as the marker of choice for the reconstruction of phylogenetic relationships at several taxonomic levels, from populations to phyla, and has been widely used for the resolution of taxonomic controversies. Indeed, the small size of the molecule, its abundance in animal tissues, the strict orthology of encoded genes, the presence of genes/regions evolving at different rates, its uniparental inheritance and the absence (or very low level) of recombination (Elson and Lightowlers, 2006) make this molecule a reliable and easy-to-use phylogenetic marker. Phylogenies have been reconstructed from the sequence of single genes, including the popular *cox1* gene, from the entire genic complement, or from gene arrangement. Indeed, mitochondrial gene order has been very effective in investigating deep-level phylogenetic relationships as it is characterized by a large number of states and generally low rates of changes (Boore *et al.*, 1998; Lavrov and Lang, 2005).

Besides its use as a phylogenetic marker, the mtDNA represents a 'full' genome and the availability of complete sequences for several phyla provides a unique, largely unexplored, opportunity to decode the mechanisms underlying genomic evolution in a phylogenetic framework. Several structural genomic features (for example genome size, gene content, gene order, compositional features, nucleotide substitution rate, repeated sequences, non-coding sequences, secondary structure of the encoded RNA) can be systematically and quite easily investigated in the small mitochondrial genome. These features allow both the description of evolutionary trends in phylogenetically distant organisms and the identification of differences in functional constraints that might account for structural differences. In fact, differences in structural genomic features generally reflect differences in functional and evolutionary constraints. At present, only the rate and pattern of nucleotide substitutions have been widely analysed in mtDNA: in vertebrates, a correlation has been demonstrated between the asymmetrical process of genome replication and parameters as mutation pattern, substitution rate and compositional asymmetry, given that all these parameters are proportional to the time spent by the H strand in single-strand status, and vary with respect to genomic position (Reyes *et al.*, 1998; Bielawski and Gold, 2002; Faith and Pollock, 2003; Raina *et al.*, 2005; Broughton and Reneau, 2006). The evolutionary and functional meaning

Correspondence: Dr C Gissi, Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Via Celoria 26, 20133 Milano, Italy. E-mail: carmela.gissi@unimi.it and Professor G Pesole, Dipartimento di Biochimica e Biologia Molecolare 'E. Quagliariello', Università di Bari, Via Orabona 4, 70126 Bari, Italy. E-mail: graziano.pesole@biologia.uniba.it  
Received 28 January 2008; revised 8 May 2008; accepted 11 June 2008; published online 9 July 2008

of other mt genome-level features awaits more detailed and systematic analyses, although some general hypotheses can be put forward. For instance, gene content is certainly affected by the ability of exchange of genetic material between the mitochondrial and nuclear compartments, the permeability and/or the presence of specific carriers on the mitochondrial membranes, gene dispensability and the difference in multimeric structure of the respiratory chain complexes between organisms. The secondary structure and size of transfer RNAs (tRNAs) and rRNAs are related to peculiarities of the mitochondrial translational apparatus, as demonstrated in nematodes by the concomitant unusual structure of mt tRNA, ribosomal RNAs (rRNA) and elongation factors (Okimoto and Wolstenholme, 1990; Okimoto *et al.*, 1994; Sakurai *et al.*, 2001, 2006; Ohtsuki *et al.*, 2002). The number, size and location of non-coding regions is mostly related to the presence of replication and transcription regulatory signals, although the overlap between regulatory signals and coding sequences cannot be excluded in the compact metazoan mtDNA (Valverde *et al.*, 1994; Peleg *et al.*, 2004). Finally, gene arrangement is affected by the transcription mechanism, for example by the need to co-regulate the expression of some genes or the required stoichiometry of the gene products. Gene order changes can also originate from processes strictly related to replication, such as tandem duplications of genomic segments due to slipped-strand mispairing or imprecise termination of replication (Boore, 2000). Although mtDNA rearrangements can have profound functional implications on gene expression and genome replication, they have mainly been used to extract phylogenetic information, rather than to extrapolate data on the transcription or replication mode: one of the few exceptions is the prediction of the functionality of the duplicated control region (CR) in snake mtDNAs based on the analysis of the variability of compositional skew along the genome (Jiang *et al.*, 2007).

In this review we will first provide a general overview of traditional and newly defined genomic features (that is gene content, genome size, architecture, AR, and gene strand asymmetry, GSA) inferred from the analysis of a carefully-revised data set of more than 1000 complete or almost complete metazoan mtDNAs. Then, we will focus on the comparison of mtDNAs from species belonging to the same genus. Indeed, as emerging from data reported in several recent publications (Fukami and Knowlton, 2005; Mueller and Boore, 2005; Iannelli *et al.*, 2007a), differences observed in congeneric species can provide a good view of the evolutionary trend of the corresponding highest taxonomic groups, and the intra-genus comparison can be a quick and easy strategy to investigate the mtDNA evolutionary trend in previously unexplored taxonomic groups. Our synthesis will surely stimulate further comprehensive analyses of these and other mitochondrial genomic features in a phylogenetic framework.

## The metazoan mtDNA data set

Despite the development of several specialized mitochondrial databases such as the NCBI Organelle Genome Resource, GOBASE, OGRE, AMiGA and MetaAMiGA and Mitome (Wolfsberg *et al.*, 2001; Jameson *et al.*, 2003; Feijao *et al.*, 2006; O'Brien *et al.*, 2006; Lee *et al.*, 2007),

the identification and retrieval of 'all' available complete mtDNAs is far from straightforward. Although these databases collect complete mtDNA sequences, and sometimes allow the retrieval and analysis of basic genomic data (such as gene order and base composition), most are updated rather infrequently and, most crucially, they do not attempt or fail to detect and correct the numerous errors present in gene annotations. In fact, only the manually curated entries of the OGRE database and the NCBI Organelle Genome Resource commonly report and fix mtDNA mis-annotations, although many entries remain uncorrected even in these resources. Most annotation errors consist of: incorrect denomination of a gene; apparent changes in gene position due to incorrect gene boundaries; and apparent lack of a gene due to partial mtDNA sequencing, or lack of annotation. Apart from these expected mis-annotations, there are other more subtle errors that can be identified only through bibliographic analyses. For example, the first published complete mtDNA of *Branchiostoma lanceolatum* (Y16474; Spruyt *et al.*, 1998) has been demonstrated to be derived from *Branchiostoma floridae* (Nohara *et al.*, 2005b) but most mt specialized databases still associate the mtDNA of *B. lanceolatum* to the Y16474 entry, rather than to the correct AB194383 sequence subsequently published (Nohara *et al.*, 2005b). Similarly, most databases report only one of the two highly different male and female mtDNA isoforms transmitted with doubly uniparental inheritance (DUI) in bivalves (Breton *et al.*, 2007), an inconsistency due to the elimination of sequence redundancy (that is additional mtDNAs belonging to the same species) performed without checking for significant differences between the putatively duplicated entries. The identification of the full set of complete mtDNA entries is also complicated by the fact that partial mtDNAs are sometimes labelled as complete sequences (implying apparent gene loss), whereas some complete sequences have an ambiguous description line, very different from the standard one, and can be identified only through specific interrogation of a primary database (see *Lingula anatina* mtDNA, AB178773).

One consequence of inconsistent annotations and/or descriptions of mtDNA entries is that the automated retrieval of the full data set of metazoan mtDNAs is quite problematic, and the same is true for statistics and comparative analyses of basic mitochondrial features such as genome size, gene content and taxonomic distribution of the available genomes. More intense efforts, including genome re-annotation, are required to carry out the analysis of more complex features, such as gene order, genome structure including the position of the CR, tRNA structures and more sophisticated compositional features.

The metazoan mtDNA data set analysed in this review has been obtained by examination of GenEmbl and three specialized mitochondrial databases: NCBI Organelle Genome Resource, OGRE and GOBASE (Ojala *et al.*, 1981; Wolfsberg *et al.*, 2001; Jameson *et al.*, 2003). The final list, updated in March 2008, consists of 1265 sequences, including 59 partial or almost-complete mtDNAs. Some of the analysed partial sequences were erroneously described as complete genomes in the original database, whereas other partial mtDNAs have been included in the data set as they are the only representatives of a phylum, or as members of a congeneric pair. We identified

mis-annotations in number, order and orientation of *mt* genes, in completeness of the mtDNA, or in species definition through different methods, such as reference consultation, software developed *ad hoc* to compare the feature tables of entries, and even sequence similarity analyses in the most enigmatic cases. Thus, 313 entries (24.8% of the entire data set) with mis-annotations in at least one feature (several entries have multiple mis-annotations) have been corrected. The list of analysed mtDNAs and corrected annotations is available on request.

## Statistics on sequenced mtDNAs

The distribution of the completely sequenced mtDNAs between the major metazoan groups is reported in Table 1. As expected, the most represented group is the sub-phylum Vertebrata followed by the phylum Arthropoda (70% and 13% of the total data set, respectively), and the most abundant species are neopterygian fishes and eutherian mammals (31% and 16% of the total data set, respectively). A total of 14 phyla are represented by only 1 or 2 sequences, and 5 of them include only partial mtDNAs (Myzostomida, Phoronida, Nemertea, Pogonophora and Sipuncula; Table 1), indicating that many gaps in the taxonomic sample of metazoan mtDNAs need to be filled.

The size of the 1206 completely sequenced mtDNAs ranges from 32 115 bp of *Placopecten magellanicus* (Mollusca Bivalvia) to 11 423 bp of *Paraspadella gotoi* (Chaetognatha) but one of two circular mt molecules of the rotifer *Brachionus plicatilis* is slightly smaller (11 153 bp; Suga *et al.*, 2008). As shown in Figure 1, the size of the mt genome varies remarkably between the major metazoan groups, and can be highly variable even within the same group (see standard deviation bars). Among the most sampled metazoans (hash mark in Figure 1), Chordata, Echinodermata, Arthropoda and Platyhelminthes are characterized by low standard deviation and stable mtDNA size. On the contrary, Mollusca, enoplean Nematoda and Porifera exhibit a strong heterogeneity in genome size (high standard deviation in Figure 1). By way of contrast, the large standard deviation observed in Brachiopoda and Chelicerata is due to the presence of a single mtDNA with a very large size compared to the almost constant size of remaining mtDNAs (the *L. anatina* mtDNA is 28.8 kb long (Endo *et al.*, 2005) against a mean value of  $14.6 \pm 0.8$  kb for three other brachiopods; the *Metaseiulus occidentalis* mtDNA is 24.9 kb long (Jeyaprakash and Hoy, 2007) against a mean of  $14.8 \pm 0.8$  kb for 27 other chelicerates).

Variation in the length of the CR, where repeated sequences are often present, accounts for much mtDNA size variability, however the largest differences in genome size are due to segmental duplication followed by several rearrangements, with consequent presence of duplicated genes or pseudogenes and increased length of non-coding sequences.

## Gene content

As shown in Table 1, the most frequent *mt* gene content consists of 37 genes, however the gene number varies from 14 genes in two Chaetognatha species to 53 in *Metaseiulus occidentalis* (Arthropoda, Chelicerata; Helfenbein *et al.*, 2004; Papillon *et al.*, 2004; Faure and

Casanova, 2006; Jeyaprakash and Hoy, 2007). This variability in gene content rarely concerns protein-coding and *rRNA* genes (see below), and is mostly due to *tRNA* genes (Table 1).

It is commonly reported that the mtDNA encodes for a full set of tRNAs able to decode the modified mt genetic code and then to perform the mt protein synthesis. This full set of mt tRNAs can vary from one taxon to the another, but it is significantly smaller than the nuclear one, as consequence of the relaxed wobble rules in codon-anticodon decoding, and of the usage of a single tRNA-Met(CAU) for both translation initiation and elongation. Thus, the expected set of mt tRNAs ranges from 24 genes for taxa with one deviation from the universal genetic code (UGA encoding Trp rather than Stop) to 22–23 genes for taxa with four deviations from the universal genetic code (modified meaning of UGA, AUA and AGR codons). However, taxa using the same mt genetic code can also encode for a radically different number of *tRNA* genes. For example, Demospongiae (Porifera) and Cnidaria use the same modified genetic code (one codon deviation) but they encode for a different number of tRNAs. In Demospongiae the number of tRNAs is extremely variable (from 2 to 27 genes), as a consequence of frequent losses or duplications of *tRNA* genes; most species also exhibit distinct initiator and elongator tRNA-Met. At the other extreme, cnidarian mtDNAs encode only one or two tRNAs (*trnM* and *trnW*) and it is believed that most mt tRNAs are imported from the cytoplasm (Beagley *et al.*, 1995). Similarly, the mtDNA of Chaetognatha encodes for 1 tRNA rather than the 22 found in most other invertebrates that employ the same genetic code (Table 1). Furthermore, the number of tRNAs is constant in some higher level taxonomic groups, and highly variable in others. For example, 1.5% of the analysed vertebrate mtDNAs differ from the standard set of 22 *tRNA* genes whereas the figure is 15.6% for arthropod mtDNAs. In vertebrates this variation is almost entirely due to gene acquisition whereas both acquisition and loss are observed in arthropods (data not shown). Gene number of tRNA varies greatly within Bivalvia (Mollusca), Enoplea (Nematoda) and Porifera, where it is difficult or impossible to define a 'standard' *tRNA* gene content (see the high standard deviation of *tRNA* gene number in Table 1). As a cautionary note, we observe that the sporadic absence of a single *tRNA* gene can be caused by errors in tRNA annotations. For example, *trnM* was originally reported as lost in the mtDNA of the chaetognath *Spadella cephaloptera* (Papillon *et al.*, 2004) and identified only in a subsequent study (Faure and Casanova, 2006), although it is still not annotated in the *Spadella* mtDNA entry (NC\_006386).

Thus, we can conclude that tRNAs are the gene category with the highest 'dispensability' in the mitochondrial genome: their number depends not only on the mt genetic code but also on the evolutionary history of the taxa and on unknown factors probably related to the mtDNA's ability to acquire/lose genetic material and to the capacity of the organelle to import and use nuclear-encoded tRNA molecules.

An interesting example of loss/acquisition of *tRNA* genes is represented by the case of tRNA-Met: two distinct genes for an initiator and an elongator tRNA-Met, both with CAU anticodon, have been identified

**Table 1** Taxonomic distribution, gene content, distinct genome architecture and gene strand asymmetry of the available metazoan mtDNAs

Phylum, classification	No. of mtDNA		No. of genes <sup>b</sup>		No. of AR		AR rate <sup>c</sup>			GSA <sup>d</sup>
	C	P <sup>a</sup>	All	tRNA	All	-tRNA	All	-tRNA	Δ%	
Chordata, Tunicata	7		39.3 ± 0.8	24.3 ± 0.8	7	6	<b>100</b>	<b>83.3</b>	14.3	1
Chordata, Cephalochordata	9		37	22	3	2	25	12.5	33.3	0.41–0.51
Chordata, Vertebrata	849	41	37.0 ± 0.3	22 ± 0.2	43	13	4.9	1.4	<b>69.7</b>	0.41–0.62
Hyperotreti	2		37	22	1	1	0	0	—	0.51
Hyperoartia	2		37	22	1	1	0	0	—	0.51
Polypteridae	3		37	22	1	1	0	0	—	0.51
Chondrostei	7		37	22	1	1	0	0	—	0.51
Neopterygii	374	27	37.0 ± 0.1	22.0 ± 0.1	14	7	3.5	1.6	50.0	0.46–0.60
Chondrichthyes	8		37	22	1	1	0	0	—	0.51
Coelacanthiformes	2		37	22	1	1	0	0	—	0.51
Dipnoi	3		37	22	1	1	0	0	—	0.51
Amphibia	82	1	37.1 ± 0.6	22.1 ± 0.5	17	5	19.7	4.9	<b>70.6</b>	0.46–0.53
Archosauria	81	5	37.0 ± 0.4	22.0 ± 0.3	3	3	2.5	2.5	0	0.41–0.51
Lepidosauria	44	1	37.0 ± 0.4	22.0 ± 0.3	11	3	23.26	4.65	<b>72.73</b>	0.49–0.57
Testudines	19		37.1 ± 0.2	22.1 ± 0.2	3	2	11.1	5.6	33.3	0.51–0.62
Prototheria	3		37	22	1	1	0	0	—	0.51
Metatheria	22	7	37	22	1	1	0	0	—	0.51
Eutheria	197		37	22	1	1	0	0	—	0.51
Echinodermata	17	1	37	22	7	6	37.5	31.2	14.3	0.08–0.68
Hemichordata	2		37	22	2	2	<b>100</b>	<b>100</b>	0	0.57–0.68
Xenoturbellida	1		37	22						0.57
Annelida	6	1	37	22	5	1	<b>80</b>	0	<b>80</b>	1
Echiura	1		37	22						1
Myzostomida	0	1	>26	>14						1
Arthropoda	154	7	37.0 ± 1.5	22.0 ± 0.9	70	25	45.1	15.7	<b>64.3</b>	<b>0.03–1</b>
Chelicerata	28		37.5 ± 3.1	22.2 ± 1.6	15	6	51.8	18.5	<b>60</b>	0.14–0.41
Myriapoda	7		37	22	6	3	<b>83.3</b>	33.3	50	0.14–0.24
Crustacea	35	1	37.1 ± 0.4	22.1 ± 0.4	19	11	52.9	29.4	42.1	<b>0.03–1</b>
Hexapoda	84	6	36.9 ± 0.9	21.9 ± 0.9	27	8	38.5	9.6	<b>72.7</b>	<b>0.03–1</b>
Brachiopoda	4		38.5 ± 3.0	23.3 ± 2.5	4	4	<b>100</b>	<b>100</b>	0	1
Chaetognatha	2		14	1	2	2	<b>100</b>	<b>100</b>	0	0.14–0.29
Bryozoa	2		36.5 ± 0.7	21.5 ± 0.7	2	2	<b>100</b>	<b>100</b>	0	0.67–0.73
Entoprocta	2		37.5 ± 0.7	22.5 ± 0.7	2	1	<b>100</b>	0	50	0.35–0.37
Phoronida	0	1	>31	>16						0.10
Nemertea	0	1	>26	>15						0.77
Pogonophora	0	2	>29	>16						1
Sipuncula	0	1	>19	>11						1
Onychophora	1		28	13						0.21
Priapulida	1		37	22						0.51
Rotifera <sup>e</sup>	2		35.5 ± 0.7	21.5 ± 0.7	2	2	<b>100</b>	<b>100</b>	0	1
Mollusca	46		37.5 ± 2.9	22.5 ± 2.4	34	24	<b>73.3</b>	51.1	29.4	<b>0.02–1</b>
Bivalvia	19		37.4 ± 4.0	23.0 ± 3.7	14	12	<b>72.2</b>	61.1	14.3	0.41–1
Cephalopoda	11		38.6 ± 2.8	22.3 ± 0.5	8	6	<b>70</b>	50	25.0	0.02–0.19
Gastropoda	13		36.9 ± 0.3	21.9 ± 0.3	9	5	66.7	33.3	44.4	0.19–0.57
Polyplacophora	1		37	22						0.14
Scaphopoda	2		37	22	2	2	<b>100</b>	<b>100</b>	0	0.08–0.46
Nematoda	26		37.6 ± 2.9	23.2 ± 2.4	17	14	64	52	17.6	<b>0–1</b>
Chromadorea	15		36	22	6	4	35.7	21.4	33.3	1
Enoplea	11		39.8 ± 3.5	24.9 ± 3.0	11	10	<b>100</b>	<b>90.0</b>	9.1	<b>0–0.80</b>
Platyhelminthes	24		36.0 ± 0.2	22.0 ± 0.2	11	3	43.5	8.7	<b>72.7</b>	1
Cnidaria	29	1	17.2 ± 0.4	2.0 ± 0.4	11	10	35.7	32.1	9.1	0.41–1
Porifera	20	2	36.2 ± 8.9	20.1 ± 9.0	18	13	<b>89.4</b>	63.2	27.8	<b>0.02–1</b>

Abbreviations: All, analysis carried out on all *mt* genes; AR, genome architecture; C, complete mtDNA; P, partial mtDNAs; -tRNA, exclusion of *tRNA* genes from the analysis; tRNA, analysis carried out only on *tRNA* genes.

Bold type indicates AR rate ≥ 70, Δ% ≥ 60 and maximum GSA variability (ranging from about zero to one).

The number and percentage of distinct genome architectures has been calculated for the entire set of *mt* genes (All) and excluding *tRNA* genes (-tRNA). The difference between AR percentage for 'All' and '-tRNA' dataset is also reported (Δ%).

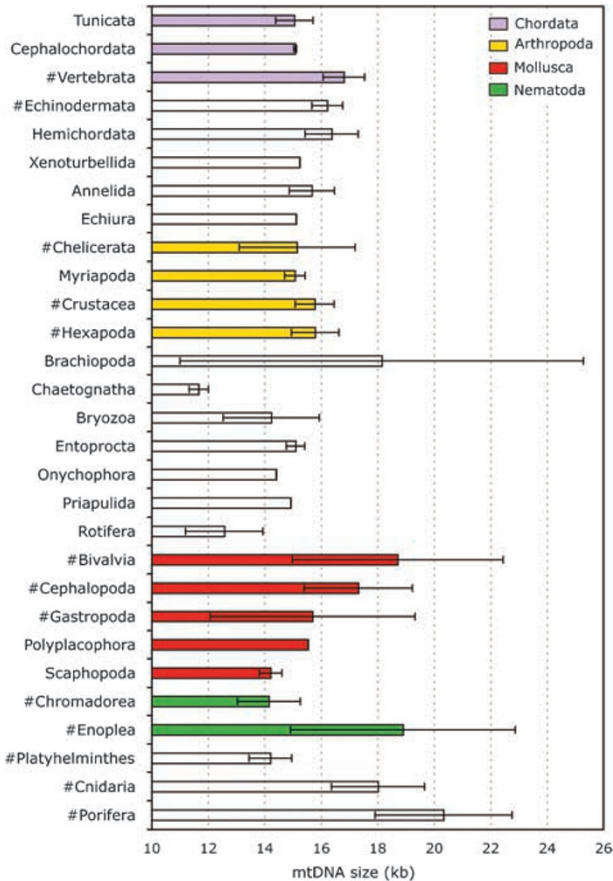
<sup>a</sup>Only partial mtDNAs (length higher than 7 kb) belonging to the following categories are indicated: mtDNAs erroneously reported as complete genomes; unique mtDNA representatives of a taxonomic group; mtDNAs representative of a congeneric pair.

<sup>b</sup>Mean ± standard deviation calculated for complete mtDNAs. The gene number of partial mtDNAs is highlighted by the '>' symbol.

<sup>c</sup>The variability of genome arrangement (AR rate) has been calculated as  $(N_{AR-All} - N_{AR-tRNA}) / (N_{mtDNA} - 1) \times 100$ . The value is reported only for taxa with more than one complete genome. Δ% has been calculated as  $(N_{AR-All} - N_{AR-tRNA}) / N_{AR-All} \times 100$ .

<sup>d</sup>The gene strand asymmetry has been calculated as  $|N_{major-strand} - N_{minor-strand}| / N_{total\ genes}$ . The range for each taxon is reported.

<sup>e</sup>The two circular molecules of *Brachionus plicatilis* (Suga et al., 2008) has been considered as a single genome.



**Figure 1** Mean length (kb  $\pm$  standard deviation bars) of complete mtDNAs calculated for the major metazoan groups. Hash mark indicates groups including more than 10 complete mtDNAs. The size of both circular mt chromosomes of *Brachionus plicatilis* (Suga et al., 2008) has been used to calculate the mean value in Rotifera.

only in the phylum Placozoa, the basal metazoan lineage (Dellaporta et al., 2006; Signorovitch et al., 2007) and in Porifera Demospongiae. However, the existence of an elongator and an initiator tRNA-Met have been also hypothesized in tunicates and some bivalves, where an unusual tRNA-Met(UAU) has been recognized as an elongator tRNA-Met additional to the common tRNA-Met(CAU) (Hoffmann et al., 1992; Beagley et al., 1999; Yokobori et al., 1999, 2003, 2005; Gissi et al., 2004). This peculiar phylogenetic distribution suggests that the elongator and initiator tRNAs-Met have been lost early in metazoan diversification, and have been re-acquired independently in the two distant lineages of mollusc bivalves and tunicates. Our revised mtDNA annotations highlighted that the unusual tRNA-Met(UAU) is encoded by all available mtDNAs of tunicates and by six bivalve species, whereas the common tRNA-Met(CAU) gene is duplicated in five additional bivalve species (based on our annotation, *trnM*(UAU) is present in *Mizuhopecten yessoensis*, *P. magellanicus*, both M and F types of three *Mytilus* species, and the M type of *Venerupis philippinarum*; *trnM*(CAU) is duplicated in *Acanthocardia tuberculata*, two *Crassostrea* species, *Hiattella arctica*, and the F type of *V. philippinarum*). Thus, it should be further investigated whether the sporadic distribution of duplicated tRNA-Met(CAU) in Bivalvia is correlated

to the absence of the additional tRNA-Met(UAU), as it is possible that the mt translational system of bivalves has acquired distinct elongator and initiator tRNAs-Met by duplication of an ancestral tRNA-Met gene or editing of a duplicated gene.

With respect to the other mt-encoded genes, the common content consists of 2 rRNA genes (*rrnS* and *rrnL*) and 13 protein-coding genes for subunits of the respiratory chain complexes (*nad1-nad6*; *nad4L*; *cox1-cox3*; *cob*; *atp6*; *atp8*). The two rRNA genes are always mitochondrially encoded, and their duplication is very rare, having been observed in only four species (*rrnS* is duplicated in the bivalve *Crassostrea gigas* and in distinct haplotypes of the pillbug nematode *Thaumamermis cosgrovei* (Milbury and Gaffney, 2005; Tang and Hyman, 2007); *rrnL* is duplicated in the chigger mite *Leptotrombidium pallidum* (Shao et al., 2005b); both *rrnS* and *rrnL* are duplicated in the large mt haplotype of the nematode *Strelkovimermis spiculatus*). Loss or acquisition of the protein-coding genes are also rather infrequent (Table 2), and the actual loss of genes remains sometimes ambiguous due to uncertainty in gene annotation or to the availability of incomplete mtDNA sequences (see the absence of *nad3* and *nad6* in the mite *Metaseiulus occidentalis*, Jeyaparakash and Hoy, 2007, and the absence of *atp8* and *nad6* in two Hexactinellida sponges, Haen et al., 2007, respectively). Few cases of sporadic gene loss have been reported in vertebrates (Table 2), and only the loss of *nad6* in the neopterygian *Choinodraco rastrispinosus* has been described as a feature common to all Antarctic fishes of the suborder Nothothenioidei, due to the real loss of the protein function (Papetti et al., 2007). Notably, genes encoding for subunits of the adenosine-triphosphate (ATP) synthase complex show a significant tendency to be lost or acquired by the metazoan mtDNA (Table 2). Indeed, *atp9* is encoded by most Porifera mtDNAs (see below) and *atp6* is absent in all Chaetognatha. *atp8* is absent or highly modified in several phylogenetically distant groups, that is in Chaetognatha, Rotifera, most Mollusca Bivalvia (13 of the 19 sampled mtDNAs), Nematoda and Platyhelminthes. Indeed, putative *atp8* genes have been annotated in one rotifer (Steinauer et al., 2005) and in the nematode *Trichinella spiralis* (Lavrov and Brown, 2001a), but we note that none of these putative proteins contain the MPQL amino acid signature conserved at the N-terminal of other metazoan ATP8 (Gissi et al., 2004). It should be stressed that the ATP8 protein is characterized by short and variable length, and by a higher conservation of the secondary structure compared to the primary sequence (Papakonstantinou et al., 1996a, b; Gray, 1999), and that both these features have hampered the annotation of *atp8* in several cases. Indeed, among bivalves *atp8* has been annotated only in *Hiattella* (Heterochonchia; Dreyer and Steiner, 2006) and *Lampsilis* (Palaeoheterodonta; Serb and Lydeard, 2003) but we have identified this gene also in the M and F types of both *Venerupis* and *Inversidens* species. Similarly, *atp8* was not annotated in the first published tunicate mtDNA of *Halocynthia roretzi* (Yokobori et al., 1999), although a very short form of ATP8 is actually mitochondrially encoded in all tunicates (Gissi and Pesole, 2003; Yokobori et al., 2005; Iannelli et al., 2007a). In spite of these annotation difficulties, *atp8* could be considered a real ‘dispensable’ mitochondrially encoded gene, as it is absent in several distant metazoan lineages.

**Table 2** Taxa with a non-standard content of mitochondrial protein-coding genes with respect to the usual presence of 13 genes encoding for subunits of the respiratory chain

	Missing protein genes	Additional protein genes/introns
Vertebrata	<i>atp8</i> and <i>nad5</i> in <i>Polypedates megacephalus</i> (Amphibia; Zhang <i>et al.</i> , 2005) <i>nad5</i> in <i>Sphenodon punctatus</i> (Lepidosauria; Rest <i>et al.</i> , 2003) <i>nad6</i> in <i>Chionodraco rastrospinosus</i> (Neopterygii; Papetti <i>et al.</i> , 2007)	
Arthropoda	<i>nad3</i> and <i>nad6</i> in <i>Metaseiulus occidentalis</i> (Jeyaprakash and Hoy, 2007)	
Annelida		Group II intron interrupting <i>cox1</i> gene in <i>Nephtys</i> sp. (Valles <i>et al.</i> , 2008)
Chaetognatha	<i>atp8</i> and <i>atp6</i> <sup>a</sup>	
Rotifera	<i>atp8</i> in <i>Brachionus plicatilis</i> (Suga <i>et al.</i> , 2008) Two un-orthodox putative <i>atp8</i> in <i>Leptorhynchoides thecatus</i> (Steinauer <i>et al.</i> , 2005)	
Mollusca	<i>atp8</i> in most Bivalvia <sup>b</sup>	
Nematoda	<i>atp8</i> ; putative <i>atp8</i> in <i>Trichinella spiralis</i> (Lavrov and Brown, 2001b)	
Platyhelminthes	<i>atp8</i> <sup>a</sup>	
Cnidaria		<i>mutS</i> (mismatch-repair protein) in Octocorallia (Pont-Kingdon <i>et al.</i> , 1995; Pont-Kingdon <i>et al.</i> , 1998; Medina <i>et al.</i> , 2006) <i>dnaB</i> (family B DNA polymerase) in <i>Aurelia aurita</i> (Scyphozoa; Shao <i>et al.</i> , 2006b) Homing-endonuclease in the group I intron interrupting the <i>cox1</i> gene in the two hexacorallians <i>Metridium senile</i> (Beagley <i>et al.</i> , 1998) and <i>Savalia savaglia</i> (Sinniger <i>et al.</i> , 2007) One unknown ORF in <i>Aurelia aurita</i> (Scyphozoa; Shao <i>et al.</i> , 2006b) and <i>Pocillopora</i> genus (Hexacorallia; Fukami and Knowlton, 2005; Flot and Tillier, 2007) Group I introns in Hexacorallia (Medina <i>et al.</i> , 2006; Chen <i>et al.</i> , 2008)
Porifera	<i>atp8</i> and <i>nad6</i> in <i>Sympagella nux</i> <sup>c</sup> (Haen <i>et al.</i> , 2007)	<i>atp9</i> , except in <i>Amphimedon queenslandica</i> (Erpenbeck <i>et al.</i> , 2007) Two unknown ORFs in <i>Iphitheon panicea</i> (Haen <i>et al.</i> , 2007) <i>tatC</i> (C subunit of the twin-arginine translocase) in <i>Oscarella carmela</i> (Wang and Lavrov, 2007) LAGLIDAGDG ORF in the group I intron interrupting the <i>cox1</i> gene in <i>Plakortis angulospiculatus</i> (NC_010217)

Abbreviation: ORF, open reading frame.

<sup>a</sup>Feature shared by all sequenced mtDNAs of a taxa.

<sup>b</sup>*atp8* is absent in 9 of the 13 bivalve species: *Acanthocardia tuberculata* (NC\_008452), *Argopecten irradians* (NC\_009687), *Crassostrea gigas* (NC\_001276), *Crassostrea virginica* (NC\_007175), *Mizuhopecten yessoensis* (NC\_009081), F and M types of *Mytilus edulis* (NC\_006161 and AY823623), F and M types of *Mytilus galloprovincialis* (NC\_006886 and AY363687); F, M and recently masculinized types of *Mytilus trossulus* (NC\_007687, DQ198231 and DQ198225) and *Placopecten magellanicus* (NC\_007234).

<sup>c</sup>Partial mtDNA.

It needs to be further investigated whether this loss is related to the transfer of the gene to the nuclear genome, or to the actual loss of the ATP8 protein due to differences in the ATP synthase structure between taxa.

Extra protein-coding genes are present only in the mtDNA of non-bilaterian metazoans, that is Cnidaria and Porifera (Table 2). Although some of these genes are open reading frames (ORFs) with no similarity to known proteins, it has been observed that the additional proteins of Cnidaria are mostly involved in DNA interaction, whereas those of Porifera have metabolic roles (Flot and Tillier, 2007) and are also usually encoded by the mtDNA of closest relatives of metazoans. In fact, among the additional proteins, the *atp9* and *tatC* genes identified in Porifera are also encoded by the mtDNA of protists and unicellular relatives of metazoans (Yen *et al.*, 2002; Burger *et al.*, 2003; Dellaporta *et al.*, 2006), indicating that this is

one of the several primitive non-metazoan features retained by Porifera (Haen *et al.*, 2007). The sporadic presence/absence of distinct additional genes in non-bilaterian species indicates a general propensity of these mtDNAs to acquire or lose genetic material, sometimes through the activity of mobile elements. Indeed, Hexacorallia possess one or two introns of group I, probably still active as shown by the tendency of the *nad5* intron to gain genes from the rest of the genome (Medina *et al.*, 2006; Brugler and France, 2007; Sinniger *et al.*, 2007), whereas in the sponge *Amphimedon queenslandica* the *atp9* gene has been transferred to the nuclear genome, probably through a transposon-mediated translocation, as the functional nuclear *atp9* gene is flanked by inverted repeats (Erpenbeck *et al.*, 2007).

Finally, introns of group I have been identified only in basal metazoans belonging to Cnidaria and Porifera,

whereas surprisingly a group II intron has been found in the annelid *Nephtys* sp., probably as result of a recent horizontal gene transfer from a bacterial or viral vector (Valles *et al.*, 2008).

### Genome architecture

The mtDNA AR is defined as the order of the entire set of functional mt-encoded genes, included duplicated and unusual *mt* genes. Therefore, this feature takes into account both gene content and gene order, and a difference in mtDNA AR means a difference in gene content and/or gene order. We have estimated the variability in mtDNA AR for each major taxonomic group by the 'AR rate', given by  $(N_{AR}-1)/(N_{mtDNA}-1)$ , where  $N_{AR}$  and  $N_{mtDNA}$  are the number of different ARs and the number of completely sequenced mtDNAs of that taxa, respectively (Table 1). Thus, the AR rate ranges between 0 (no AR variability) and 100 (all mtDNAs have a different AR).

As shown in Table 1, AR rate values (*All genes*) higher than 70, indicatives of a strong variability in genome AR, are observed in nine phyla (Hemichordata, Annelida, Brachiopoda, Chaetognatha, Bryozoa, Entoprocta, Rotifera, Mollusca and Porifera) as well as in subgroups belonging to phyla with good conservation of mtDNA AR: Tunicata within Chordata, Myriapoda within Arthropoda, and Enoplea within Nematoda. Although some of these groups are represented by a few sequences (Table 1), the AR variability seems to be real, as in many cases the AR differences have been found in the most closely related species within that taxonomic group (data not shown).

Vertebrates are the sub-phylum with the lowest variability in mtDNA AR compared to other groups of similar taxonomic rank (4.9 AR rate against a range of 35.7–100 for other sub-phyla/classes), suggesting that the almost frozen *mtDNA* gene order and content ascribed to vertebrates is an exception in metazoan evolution, rather than a rule. Moreover, the frequent changes in genome AR found in vertebrate lepidosaurians and amphibians (Table 1) suggest that the mtDNA AR within vertebrates is not so invariant as previously believed, and that the observed stability could be the consequence, to some extent, of the uneven sampling among vertebrates.

Excluding tRNAs, often described as highly 'mobile' mitochondrial genes, the number of different genome ARs decreases only slightly in taxa showing the highest AR variability (see AR-tRNA and  $\Delta\%$  of Tunicata, Hemichordata, Brachiopoda, Chaetognatha, Bryozoa, Rotifera, Mollusca, enoplean Nematoda and Porifera in Table 1), with the exception of the phylum Annelida, where the extensive rearrangements appear exclusively associated with the translocation of *tRNA* genes. However, the exclusion of *tRNA* genes leads to a large decrease in AR rate ( $\Delta\% \geq 60$ ) in Vertebrata, Arthropoda and Platyhelminthes, all taxa with moderate variability of genome AR (see AR-tRNA and  $\Delta\%$  in Table 1). These data underline that changes in the number and location of *tRNA* genes are the main cause of differences in mtDNA AR only in case of a moderate AR variability, and suggest that different rearrangement mechanisms, characterized by a differential involvement of *tRNA* genes, could occur in taxa with unstable versus stable mtDNA AR. Alternatively, a 'saturation effect' in gene

order changes can also account for the apparent lack of strong tRNA mobility in taxa with the highest AR rate (see also the paragraph on mtDNA variability in congeneric species).

In general, the observation of high variability in mtDNA AR in several phylogenetically distant groups indicates that acceleration of the rate of genomic rearrangements has occurred independently several times in the evolutionary history of metazoans, and suggests that the causes of this increased rate could be investigated through the analysis of physiological and metabolic features common to distant taxa sharing the same evolutionary pattern.

### Gene strand asymmetry

The mtDNA of most metazoans is characterized by an asymmetric distribution of the genes between the two strands, allowing the identification of a major and a minor strand depending on the number of encoded genes. Here, the GSA has been quantified as the absolute value of the difference in gene number between the two strands, divided by the total number of genes. The resulting GSA value ranges 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 total genes on the major strand.

Quantitative data on the GSA point out to a large variability of GSA, ranging from almost 0 to 1, in Arthropoda, Mollusca, enoplean Nematoda and Porifera (Table 1), indicating that frequent exchanges of genes between the two mt strands have occurred during the evolution of these genomes. On the contrary, the GSA is almost invariant and very close to 0.5 in all vertebrates and cephalochordates, in accordance with the observation of low variability in genome AR (see AR rate in Table 1) and rare gene inversions in these two groups (see below).

Mitochondrial genomes showing a symmetric gene distribution (GSA  $\leq 0.1$ , thus 50–55% of the total genes on each strand) are very rare, indeed this feature has been observed only in 17 phylogenetically distant species belonging to four phyla (a few echinoderms, cephalopods, scaphopods, crustaceans, hexapods, nematodes and sponges; Table 1). On the contrary, about 10% (122 sequences) of the analysed complete mtDNAs are characterized by the most extreme GSA, that is by all genes encoded on the same strand and therefore transcribed in the same direction. This peculiar feature is shared by all, or almost all, mtDNAs of five phyla (Annelida, Brachiopoda, Platyhelminthes, Cnidaria and Porifera) and three sub-phyla/classes (Tunicata, Bivalvia and Chromadorea; Table 3). These mtDNAs can be also regarded as containing only one cluster of co-oriented genes, and it is notable that the few exceptions to the 'one-cluster structure' found in Porifera, Cnidaria and Bivalvia consist of mtDNAs with two clusters of co-oriented genes, one for each strand (Table 3). Similarly, the mtDNAs of non-bivalve molluscs are characterized by a structure with one-, two- or four-exact or defective clusters of co-oriented genes (Table 3), where defective clusters are defined as co-oriented gene cluster interrupted by the presence of one or two adjacent genes located on the opposite strand. This feature suggests that



**Table 3** Taxa with mtDNA containing up to four clusters of co-oriented genes

Phylum	Total <sup>a</sup>	One-cluster mtDNA <sup>b</sup> (interrupting genes)	Two-cluster mtDNA <sup>b</sup> (interrupting genes)	Four-cluster mtDNA <sup>b</sup> (interrupting genes)
Annelida <sup>c</sup>	6	6		
Brachiopoda	4	4		
Platyhelminthes	24	24		
Cnidaria	29	26	3 <i>Aurelia aurita</i> ; <i>Briareum asbestinum</i> ; <i>Pseudopterogorgia bipinnata</i>	
Porifera	20	18	2 <i>Aplysina fulva</i> ; <i>Oscarella carmela</i>	
Subphylum/class				
Tunicata	7	7		
Chromadorea	15	15		
Bivalvia	19	16	3 <i>Lampsilis ornata</i> ( <i>nad5</i> ); <i>Inversidens japonensis</i> F–M ( <i>nad5</i> )	
Cephalopoda	11	0	4 <i>Nautilus macromphalus</i> ( <i>trnT</i> ; <i>atp6</i> ); <i>Octopus ocellatus</i> ( <i>trnT</i> ); <i>Octopus vulgaris</i> ( <i>trnT</i> ); <i>Vampyroteuthis infernalis</i> ( <i>trnT</i> )	5 <i>Dosidicus gigas</i> ( <i>trnT</i> ); <i>Sepia esculenta</i> ( <i>trnT</i> ); <i>Sepia officinalis</i> ( <i>trnT</i> ); <i>Todarodes pacificus</i> ( <i>trnT</i> ); <i>Watasenia scintillans</i> ( <i>trnT</i> )
Gastropoda	13	0	12 <i>Albinaria caerulea</i> ( <i>trnT-cox3</i> ); <i>Aplysia californica</i> ( <i>trnT-cox3</i> ; <i>trnC</i> ); <i>Biomphalaria glabrata</i> ( <i>trnT-cox3</i> ); <i>Biomphalaria tenagophila</i> ( <i>trnT-cox3</i> ); <i>Cepaea nemoralis</i> ; <i>Conus textile</i> ( <i>trnT</i> , <i>trnL</i> (UUR)); <i>Haliotis rubra</i> ( <i>trnT</i> ); <i>Ilyanassa obsoleta</i> ( <i>trnT</i> ); <i>Lophiotoma cerithiformis</i> ( <i>trnT</i> ); <i>Pupa strigosa</i> ( <i>trnT-cox3</i> ; <i>trnC</i> ); <i>Roboastra europaea</i> ( <i>trnT-cox3</i> ); <i>Thais clavigera</i> ( <i>trnT</i> )	
Crustacea	35	2	2 <i>Tigriopus</i> genus	
Hexapoda	84	2 <i>Campanulotes bidentatus</i> ( <i>trnQ</i> ); <i>Bothriometopus macrocnemis</i>	2 <i>Eriocheir sinensis</i> ; <i>Megabalanus volcano</i>	
Low-sampled taxa				
Polyplacophora	1	0	1 <i>Katharina tunicata</i> ( <i>trnT</i> ; <i>trnP</i> )	
Scaphopoda	2	0	1 <i>Siphonodentalium lobatum</i> ( <i>rrnS</i> )	
Rotifera <sup>d</sup>	2	2		
Bryozoa	2	0	1 <i>Flustrellidra hispida</i>	
Entoprocta	2	0	2 <i>Loxocorone allax</i> ( <i>trnT</i> ; <i>trnP</i> ); <i>Loxosomella aloxiata</i> ( <i>trnT</i> ; <i>trnP</i> )	
Chaetognatha	2	0	2 <i>Paraspadella gotoi</i> ; <i>Spadella cephaloptera</i> ( <i>nad4L</i> )	
Echiuria	1	1		
Myzostomida	1 <sup>e</sup>	1		
Phoronida	1 <sup>e</sup>	0	1 <i>Phoronis psammophila</i> <sup>f</sup> ( <i>trnT</i> ; <i>trnP</i> )	
Pogonophora	2 <sup>e</sup>	2		
Sipuncula	1 <sup>e</sup>	1		

<sup>a</sup>Total number of available complete mtDNAs for each taxon.

<sup>b</sup>A cluster consists of at least three co-oriented genes and can be interrupted by one or two adjacent genes located on the opposite strand. Genes or gene pairs interrupting a cluster are listed in parenthesis separated by a semicolon. The species name is also indicated.

<sup>c</sup>All genes are transcribed by the same strand even in the partial mtDNAs of the polychaete *Scoloplos armiger* (Bleidorn et al., 2006) and the hirudinid *Helobdella robusta* (Boore and Brown, 2000).

<sup>d</sup>Two putative *trnS* genes are encoded by the minus strand in the Acanthocephala *Leptorhynchoides thecatus* (Steinauer et al., 2005), whereas all genes are transcribed from the same strand in the two circular mitochondrial chromosomes of *Brachionus plicatilis* (Suga et al., 2008).

<sup>e</sup>Only partial mtDNAs.

<sup>f</sup>Indicated as *Phoronis architecta* in the publication (Helfenbein and Boore, 2004).

inversions of large mitochondrial regions have occurred frequently during the evolution of some taxonomic groups, such as Mollusca, Porifera and Cnidaria. Finally, it is intriguing that five of the eight previously described taxa with only one coding strand are also characterized

by a strong variability of genome AR (see Tunicata, Annelida, Brachiopoda, Bivalvia and Porifera in Tables 1 and 3), suggesting a possible correlation between these two features and that the same mechanism could be responsible for both the presence of a single cluster of



co-oriented genes and a strong variability in gene number/order.

It should be stressed that the distribution of genes between the two strands is directly related to the mechanism of mtDNA transcription. In principle, only one strand should be transcribed in genomes with a single-coding strand, and the existence of one or multiple transcription units per strand could be predicted based on the distribution of the co-oriented genes as clustered or scattered along a given strand, respectively. However, the data on mt transcription in mammals demonstrate that we can also expect the transcription of the entire minor or non-coding strand, against the principle of the highest cellular economy, probably in response to the requirement of gene regulation by antisense transcripts.

### Benefits of congeneric comparisons

The best way to carry out evolutionary analyses is to compare a given feature in organisms divergent enough to observe significant differences, but sufficiently similar to exclude the saturation in the occurred differences. Congeneric species, that is species belonging to the same genus, can be regarded as a special case of closely related organisms for which the phylogenetic distance is based on traditional morphological characters and is the lowest that can be measured between clearly distinguishable organisms. Of course, different congeneric pairs are not expected to have the same divergence time neither the same evolutionary rate and pattern but their evolutionary divergence may be considered as minimized and standardized within a given phylum on the basis of classical taxonomic criteria.

The current availability of a large number of congeneric mtDNAs is partly the consequence of fortuitous coincidences related to the interest of several authors on the same subject (Yokobori *et al.*, 2004; Cook, 2005; Akasaki *et al.*, 2006), whereas in other cases the mtDNA has been intentionally sequenced and compared in congeneric species, subspecies, strains or members of a species complex with the aim to reconstruct phylogenetic relationships (Minegishi *et al.*, 2005; Nishibori *et al.*, 2005; Goios *et al.*, 2007), to study a peculiar evolutionary dynamics (Mueller and Boore, 2005), to develop reliable molecular markers for species identification and monitoring (Guo *et al.*, 2006; Nakao *et al.*, 2007; Park *et al.*, 2007a), or to solve controversies on the taxonomic status (Fukami and Knowlton, 2005; Iannelli *et al.*, 2007b). Beyond being the result of exotic interest, these studies expanded our knowledge on genomic evolution and can be applied in field as different as human health (Le *et al.*, 2002; Nakao *et al.*, 2007; Park *et al.*, 2007a), aquaculture activities (Milbury and Gaffney, 2005; Guo *et al.*, 2006, 2007), genetic improvement of commercially relevant species (Nishibori *et al.*, 2005) and monitoring of natural populations or endangered species (Milbury and Gaffney, 2005; Nishibori *et al.*, 2005). For instance, the recent evolutionary history of mice has been reconstructed through the analysis of wild-type and inbred *Mus* strains with a well-documented history, revealing a faster substitution rate and a higher accumulation of non-synonymous substitutions in inbred mice compared to wild types (Goios *et al.*, 2007); the mitochondrial phylogeny of chicken breeds and wild-type species of the genus *Gallus* has revealed that several inter-species

hybridizations have contributed to the establishment of contemporary domesticated chicken, also providing useful data for a genetic improvement of chicken breeds (Nishibori *et al.*, 2005). As regards genomic evolution, the comparisons of Plethodontidae salamanders have further elucidated the mechanisms of gene duplication and modification of gene order, given that these amphibians are characterized by unusually frequent mtDNA rearrangements compared to other vertebrates (Mueller and Boore, 2005). Similarly, the high incidence of tandem duplications in several lineages of parthenogenic lizards, such as the geckos *Heteronotia binoei*, has been studied in detail to investigate the origin and evolution of large duplicated mt segments (Fujita *et al.*, 2007).

The capacity of mtDNA to discriminate between closely related species, thus its potential as a 'DNA barcode', has stimulated the sequencing of the entire mtDNA in several species, in order to identify the mt region most useful as a species-specific marker. Thus, the mtDNA has been entirely sequenced for the two human intestinal parasites *Diphyllobothrium latum* and *Diphyllobothrium nihonkaiense* (Platyhelminthes), sibling species morphologically almost indistinguishable at the larval stage, with the aim of identifying reliable molecular markers for diagnostic purposes (Nakao *et al.*, 2007; Park *et al.*, 2007a). Similarly, the problematic morphological identification of two monogeneans belonging to the genus *Gyrodactylus*, both parasites of commercially relevant fishes, has led to the sequencing of their entire mt genome (Huysse *et al.*, 2007; Plaisance *et al.*, 2007). The mtDNA has been also used for monitoring the progeny of inter-species crosses of carp, providing evidence of mtDNA paternal leakage and recombination in triploid crucian carp (Guo *et al.*, 2006), and of strictly maternal inheritance in allotetraploid carp (Guo *et al.*, 2007).

Recently, the comparative mitogenomics approach has been applied to clarify the taxonomic status of some species, as in the case of members of the coral species complex *Montastraea annularis* (Cnidaria; Fukami and Knowlton, 2005), coral morphospecies of the genus *Pocillopora* (Cnidaria; Flot and Tillier, 2007), distinct genotypes of the cestode *Echinococcus granulosus* (Platyhelminthes; Le *et al.*, 2002), and putative cryptic species of the cosmopolitan ascidian *Ciona intestinalis* (Tunicata; Iannelli *et al.*, 2007b). These studies demonstrate that the capacity of mtDNA to solve taxonomic controversies at species-level depends on the overall mt evolutionary dynamics, that is on the existence of a fast rate in the nucleotide substitutions and/or in the changes of genome AR. Thus, the low variability of cnidarian mtDNA, characterized by slow substitution rate and rare genomic rearrangements (Lavrov *et al.*, 2005; Medina *et al.*, 2006; Wang and Lavrov, 2007), has made the mtDNA useless for species definition in both *Montastraea* and *Pocillopora* genera (Fukami and Knowlton, 2005; Flot and Tillier, 2007). In fact, the gene order is identical within both genera, whereas the sequence divergence ranges from 0 to 0.12% between six *Montastraea annularis* individuals (Fukami and Knowlton, 2005) and is 0.18% between two *Pocillopora* species (Flot and Tillier, 2007). These differences are too small to permit a clear discrimination between species. Only the hypervariable mtDNA regions (*atp6* and the putative CR) of needle corals belonging to the *Seriatopora* genus appear to have

enough genetic variability to be useful for species phylogeny and population genetic studies (Chen *et al.*, 2008). On the contrary, the fast evolutionary dynamics of Platyhelminthes and Tunicata make the mtDNA of these taxa a reliable molecular marker for species discrimination. Indeed, the fast substitution rate of Platyhelminthes mtDNA has allowed the recognition of different genotypes of the parasite *E. granulosus* as distinct species (Le *et al.*, 2002): these 'genotypes' were defined based on adult morphology, host specificity, life cycle and molecular features (Thompson and McManus, 2002), but their taxonomic status is highly debated. The comparative analysis of the complete mtDNAs of G<sub>4</sub> (horse-dog strain) and G<sub>1</sub> (sheep-dog strain) genotypes of *E. granulosus* revealed that the G<sub>4</sub> and G<sub>1</sub> genotypes are almost as distant from each other, as each one is distant from the congeneric species *Echinococcus multilocularis*, thus the sequence divergence is compatible with an interspecies comparison and suggests that G<sub>1</sub> and G<sub>4</sub> merit the status of separate species (Le *et al.*, 2002). The most extreme evolutionary dynamics has been found in tunicate mtDNA, characterized by high substitution rate and extensive gene rearrangements (Yokobori *et al.*, 1999, 2005; Gissi *et al.*, 2004). These peculiarities have enabled the unequivocal demonstration of the existence of two cryptic species in the cosmopolitan ascidian *C. intestinalis*: individuals of two putative species show strong mtDNA differences incompatible with an intra-species variability at level of gene order, base composition, non-coding regions and sequence divergence (Iannelli *et al.*, 2007b).

In conclusion, the mtDNA appears a powerful marker to discriminate between closely related taxa, provided that the taxa demonstrate a relatively rapid evolutionary dynamics.

### mtDNA variability in congeneric species

In this review, we have systematically collected and detailed the mtDNA AR of the available congeneric metazoan species, as comparisons at low evolutionary distance facilitate the detection of genomic differences commonly obscured over long evolutionary distances by multiple or reverse changes. Moreover, the variability observed in closely related organisms can be regarded as an indication of the genomic plasticity present at higher taxonomic levels.

As reported in Table 4, 428 mtDNA sequences can be included in intra-genus comparative analyses, and at least two species/subspecies have been sequenced for 144 metazoan genera. The taxonomic distribution of these genera reflects the uneven mtDNA sampling already discussed: most genera belong to Vertebrata and there are no congeneric species for Annelida, Myriapoda and any poorly sampled phyla (Table 1). The most widely sampled genera include mice (genus *Mus*), for which there are 20 sampled species/strains, and freshwater eels of the genus *Anguilla* (Neopterygii), for which all extant 15 species and 3 subspecies have been sampled (Supplementary Material Table S1). Intra-genus differences in genome AR are observed in all analysed genera of Tunicata, Bivalvia, enoplea Nematoda and Porifera, whereas a constancy of genome AR is observed in almost all vertebrates and in all comparisons

of the phyla Cnidaria and Echinodermata (compare 'No Genera' to 'NG<sub>diff-AR</sub>' columns in Table 4).

The type and extent of congeneric differences are summarized in Table 5, which also shows some relevant comparisons between different mtDNA isoforms belonging to the same species such as M and F mtDNA type of bivalves and distinct haplotypes of some enoplean nematodes. In 10 of the 19 intra-genus comparisons, the *tRNA* genes specifically contribute to increase the differences in genome AR, as the number of AR decreases when tRNAs are excluded from the data set (compare AR-All and AR-tRNA in Table 5). Thus, we have analysed separately the effect of differences in gene content from gene order on genome AR. In particular differences in gene content involve mostly tRNA and rarely concern *rRNA* genes (Table 5). Similarly, the gene order differences, estimated by the normalized breakpoint distance (BDn), involve almost exclusively *tRNA* genes, as the exclusion of tRNAs decreases or set at zero the BDn in most comparisons (Table 5). Strikingly, tunicate and enoplean comparisons (that is *Phallusia* and *Romanomermis*) exhibit the most extensive differences in gene order (highest values of BDn-All) and no BDn decrease when tRNAs are excluded from the data set, suggesting saturation of gene order rearrangements. In general, data on congeneric comparisons confirm the widespread belief of high tRNA mobility in metazoan mtDNA, and highlight frequent losses/acquisitions of *tRNA* genes.

It can be also noticed that in several genera there is a concurrent variation of gene content and gene order (Table 3), indeed 8 of 19 genera show differences both in gene content and gene order. This observation is consistent with a genome rearrangement mechanism by duplication of large mtDNA segments and subsequent degeneration/loss of most additional sequences, with functionally surviving genes mostly encoding for tRNAs. According to this hypothesis, most genera with changes in genome AR are also characterized by variation of genome size (Table 5).

### A discontinuous pattern of gene rearrangement in Chordata

Gene order rearrangements at intra-genus level are very frequent in ascidians (Tunicata), moderate in lancelets (Cephalochordata) and rare in vertebrates (Tables 4 and 5). Similarly, the intensity of these rearrangements is again higher in protochordates (that is Cephalochordata and Tunicata) than in vertebrates.

As shown in Table 5, congeneric comparisons of ascidians show extensive gene order rearrangements ranging from the translocation of one or six genes in the *Ciona* genus (depending on the species pair; Yokobori *et al.*, 2003; Gissi *et al.*, 2004; Iannelli *et al.*, 2007b), to the translocation of half of the genes in the *Phallusia* genus, that shows a BDn close to the random value (Iannelli *et al.*, 2007a). Moreover, additional intra-genus differences concern the length and number of non-coding regions, and the presence/absence of short duplicated sequences including non-coding regions or small portions of *RNA* genes (Gissi *et al.*, 2004; Iannelli *et al.*, 2007a).

In the sub-phylum of Cephalochordata, the mtDNA has been analysed in all three existing genera of

**Table 4** Metazoan mtDNAs of congeneric species, with number of genera showing differences in genomic architecture (NG<sub>diffAR</sub>)

Phylum, classification	No. of congeneric mtDNAs	No. of Genera	NG <sub>diffAR</sub>
<i>Chordata</i>			
Tunicata	5	2	2
Cephalochordata	8	2	1
Vertebrata, Polypteridae	2	1	0
Vertebrata, Chondrostei	3	1	0
Vertebrata, Neopterygii	115	41	0
Vertebrata, Chondrichthyes	2	1	0
Vertebrata, Coelacanthiformes	2	1	0
Vertebrata, Amphibia	46	15	3
Vertebrata, Archosauria	28	11	0
Vertebrata, Lepidosauria	7	3	0
Vertebrata, Testudines	6	2	0
Vertebrata, Metatheria	2	1	0
Vertebrata, Eutheria	91	22	0
Echinodermata	5	2	0
<i>Arthropoda</i>			
Chelicerata	11	4	1
Crustacea	8	4	2
Hexapoda	27	9	1
<i>Mollusca</i>			
Bivalvia	13	4	4
Cephalopoda	4	2	0
Gastropoda	2	1	0
<i>Nematoda</i>			
Chromadorea	4	2	0
Enoplea	7	3	3
Platyhelminthes	19	5	1
Cnidaria	9	4	0
Porifera	2	1	1
Total	428	144	19

*Branchiostoma*, *Epigonichthys* and *Asymmetron*: an intra-genus rearrangement has been initially ascribed to the genus *Epigonichthys* (Nohara *et al.*, 2005a) and successively to the genus *Asymmetron* (Kon *et al.*, 2007), as the taxonomic status of *Epigonichthys lucayanus* has been revised from *E. lucayanus* species to *Asymmetron lucayanus* species complex (Nishikawa, 2004; Kon *et al.*, 2006). The rearrangement within the genus *Asymmetron* is fairly simple, consisting of the translocation of *trnN*, and the transposition plus inversion of the *trnQ* gene (Kon *et al.*, 2007). However, this rearrangement is rather peculiar as it includes a gene inversion. Indeed, gene inversions are observed in Chordata only within Cephalochordata, given that the *Asymmetron* genus shows the inversion of a large mt region, including four genes, with respect to other lancelet genera (Kon *et al.*, 2007). These data suggest the occurrence of mtDNA recombination in Cephalochordata, as gene inversions can be easily explained by a recombination mechanism.

Among vertebrates, only salamanders of the family Plethodontidae (Amphibia) exhibit gene rearrangements at the intra-genus level (Table 5) and represent an interesting group where it is possible to observe the mechanism of duplication-random gene loss, commonly hypothesized for mitochondrial gene order rearrangements 'at work' (Moritz and Brown, 1986; Boore, 2000; Mueller and Boore, 2005). In these animals a modified

gene order is often associated with the presence of pseudogenes and/or additional gene copies, which could be vestiges of ancestrally duplicated regions (Mueller and Boore, 2005). The detailed analyses of the combined pattern of sequence remnants and variable gene order in different plethodontid genera have suggested the occurrence of a single tandem duplication in *Batrachoseps attenuatus*, two independent duplications in *Plethodon elongatus* and one or two independent tandem duplications in the *Aneides* genus (Mueller and Boore, 2005). Moreover, the observed duplications are compatible with events of imprecise initiation-termination of DNA replication, slipped-strand mispairing or intra-molecular recombination, that is with mechanisms commonly hypothesized to explain the occurrence of sequence duplications (Mueller and Boore, 2005). Other peculiar cases of intra-genus variability in vertebrates can be observed in some lizards and geckos (Lepidosauria; Macey *et al.*, 2004; Fujita *et al.*, 2007). In amphisbaenian reptiles of the genus *Bipes*, the gene order is identical between three congeneric species, but in *Bipes baporus* there is a tandem duplication of the gene cluster *trnT-trnP* followed by pseudogene formation so that the final gene order is equal to the ancestral one (Macey *et al.*, 2004). Surprisingly, the pattern of pseudogene formation differs between different populations of *Bipes baporus*, thus different rearrangements can be identified when functional and non-functional genes are taken into

**Table 5** Congeneric species showing differences in genome architecture, gene order (breakpoint distance) and gene content

Classification	Genus	No. of sp. (mtDNA) <sup>a</sup>	No. of AR		BDn range		Intra-genus differences in gene content	Size <sup>b</sup>
			All	-tRNA	All	-tRNA		
Tunicata	<i>Ciona</i>	3	3	2	0.08–0.41	0–0.20	None	14 706 ± 103
	<i>Phallusia</i>	2	2	2	0.71	0.80	Additional <i>trnI</i> , and a <i>trnX</i> with uncertain amino acid specificity in <i>P. fumigata</i>	15 057 ± 676
Cephalochordata	<i>Asymmetron</i>	2 (4) <sup>c</sup>	2	2	0–0.16	0	None	15 084 ± 24
Amphibia	<i>Aneides</i>	2	2	2	0.08	0.20	Additional <i>trnE-trnP-trnF-trnL(UUR)</i> gene cluster in <i>A. hardii</i>	21 191 ± 1405
Chelicerata	<i>Batrachoseps</i>	2	2	1	0.08	0	None	18 674 ± 1577
	<i>Plethodon</i>	3	2	1	0.08	0	Additional <i>trnT</i> in <i>P. elongatus</i>	19 334 ± 623
Crustacea	<i>Leptotrombidium</i>	3	2	2	0–0.08	0	Duplicated <i>rrnL</i> and a pseudogene of <i>rrnS</i> in <i>L. pallidum</i>	14 736 ± 1769
Hexapoda	<i>Pollicipes</i>	2	2	1	0	0	Two additional copies of <i>trnC</i> repeated in tandem in <i>P. polymerus</i> .	15 275 ± 508
	<i>Tigriopus</i>	2	2	1	0.08	0	None	14 587 ± 58
	<i>Rhagophthalmus</i>	2	2	1	0	0	The cluster <i>trnI-trnQ-trnM</i> is lost in <i>R. ohbai</i>	15 843 ± 197
	<i>Crassostrea</i>	2	2	2	0.32	0	Duplicated <i>rrnS</i> , <i>trnK</i> , and <i>trnQ</i> genes in <i>C. gigas</i>	17 734 ± 694
	<i>Inversidens</i> <sup>d</sup>	1 (2)	2	2	0.27	0.20	None	16 897 ± 100
	<i>Mytilus</i> <sup>d,e</sup>	3 (7)	2	1	0	0	Additional <i>trnG</i> , <i>trnN</i> , <i>trnI</i> , <i>trnQ</i> genes in the ‘recently masculinized’ type of <i>M. trossulus</i> .	17 561 ± 1021
Enoplea	<i>Venerupis</i> <sup>d</sup>	1 (2)	2	1	0	0	Additional <i>trnM(UAU)</i> and duplicated <i>trnK</i> in the M type; duplicated <i>trnV</i> in the F type.	22 059 ± 873
	<i>Romanomermis</i>	3	3	3	0.47–0.78	0.50–0.79	Several gene duplications: the number of annotated genes is 48 in <i>R. culicivora</i> , 40 in <i>R. iyengari</i> and 39 in <i>R. nielseni</i>	20 220 ± 5442
	<i>Strelkovimermis</i> <sup>f</sup>	1 (2)	2	2	0	0	Duplicated <i>rrnS</i> and <i>rrnL</i> in the large mtDNA haplotype	17 574 ± 644
	<i>Thaumamermis</i> <sup>f</sup>	1 (2)	2	1	0 <sup>g</sup>	0	A hypervariable segment differentiates haplotypes: it contains several pseudo- and functional genes ( <i>rrnS</i> , and <i>tRNA</i> genes) organized in a variable gene order	20 760 ± 1056
Platyhelminthes	<i>Schistosoma</i>	5	5	2	0–0.44	0–0.36	None	14 895 ± 1183
Porifera	<i>Amphimedon</i>	2	2	2	0.26	0	<i>A. queenslandica</i> lacks <i>atp9</i> and 10 tRNA genes, whereas <i>A. compressa</i> lacks the elongator <i>trnM</i> gene	19 262 ± 987

Abbreviations: All: analysis carried out on all *mt* genes; -tRNA: exclusion of *tRNA* genes from the analysis.

The normalized breakpoint distance (BDn) is the breakpoint distance (Blanchette *et al.*, 1997) calculated on the order of the genes shared between congeneric species, and divided by the number of genes.

<sup>a</sup>The number of mtDNAs is reported only when different from the number of species (indicated as sp.).

<sup>b</sup>Average mtDNA length (bp) ± standard deviation.

<sup>c</sup>Including three representatives of the *Asymmetron lucayanus* species complex.

<sup>d</sup>Male (M) and female (F) isoforms of the same species. The recently masculinized mtDNA isoform is available only for *Mytilus trossulus*.

<sup>e</sup>The existence of a single *Mytilus edulis* species complex, including *M. edulis*, *M. galloprovincialis* and *M. trossulus*, has also been hypothesized (Gosling, 1992).

<sup>f</sup>Two haplotypes of the same species.

<sup>g</sup>Gene order outside the hypervariable segment.

account (Macey *et al.*, 2004). Several complete mtDNAs of hybrid parthenogenic geckos of the *H. binoei* species complex contain tandem duplications ranging in size from 5.7 to 9.4 kb, and including both the CR and all gene categories (tRNA, rRNA and proteins; Fujita *et al.*, 2007). The small direct repeats located at the duplication endpoints have suggested a slipped-strand mispairing during DNA replication as the mechanism generating these duplications, moreover the relatively common indels in the duplicated segments point to deletions as the mechanism disrupting gene function (Fujita *et al.*, 2007).

The pattern and frequency of gene order rearrangement observed in congeneric comparisons of Chordata (Table 5) well exemplifies the data derived from the analysis of the entire data set of 865 chordate mtDNAs (Table 1). Indeed, the intra-genus comparisons confirm the extreme variability in gene order observed in tunicates, which show different gene orders from one species to the other, and no gene block shared by all available species (Iannelli *et al.*, 2007a). On the contrary, cephalochordates are characterized by moderate rearrangements, including gene inversions, and by a gene order that can be easily traced back to that of vertebrates (Nohara *et al.*, 2005a; Kon *et al.*, 2007). Finally, the gene order of vertebrates is not so invariant as previously reported, and the mtDNA AR appears remarkably dynamic especially in Amphibia and Lepidosauria (see respective AR rate in Table 1), as also indicated by the few cases of vertebrate congeneric comparisons here described. In fact, the dogma of a frozen gene order in vertebrates has been broken by the identification of several deviations from the 'typical' gene order in marsupials, birds, crocodiles, reptiles, amphibians, bony fishes and lampreys (Boore, 1999; Pereira, 2000; Inoue *et al.*, 2003). Many of these alternative gene arrangements involve genes close to the CR or surrounding the L-strand replication origin (Boore, 1999), and are sometimes associated with duplication of the CR, leading to the hypothesis that errors in mtDNA replication have given rise to these rearrangements. The observation of these 'hot spots' strongly suggests the need to analyse the evolution of vertebrate gene order including data on the location of the replication origins and their eventual duplications/losses, in order to better discriminate between alternative mechanisms of gene order rearrangements. Although an updated review of the vertebrate mtDNA organization will be noteworthy and timely, it is outside the scope of this review.

#### Arthropoda: loss and acquisition of *tRNA* genes and control regions

Among 17 arthropod genera, congeneric differences in mtDNA AR have been found in three sub-phyta, and are less frequent in Hexapoda compared to Crustacea and Chelicerata (Tables 4 and 5).

In Hexapoda, differences in gene content have been found in luminous beetles of the genus *Rhagophthalmus* (Coleoptera), where a cluster of three *tRNA* genes is present in *Rhagophthalmus lufengensis* and absent in the congeneric beetle *R. ohbai* (Li *et al.*, 2007). Although the authors suggest further investigations, both mtDNA sequences are complete, and unusual tRNA structures that could confuse tRNA identification have not been

observed in coleopteran mtDNAs, suggesting that this is a real mtDNA difference.

In Crustacea, the observed intra-genus rearrangements are always related to *tRNA* genes, indeed the gene order differences between two *Tigriopus* species are due to the translocation of the single *trnW* gene (Machida *et al.*, 2002; Burton *et al.*, 2007), whereas there are three almost identical copies of *trnC* repeated in tandem in *Pollicipes polymerus* (Lavrov *et al.*, 2004) against only one *trnC* gene in *Pollicipes mitella* (Lim and Hwang, 2006; this paper has several gene mis-annotations).

Congeneric differences are much more extensive in Chelicerata, at least within Acari, and differences in genome size, gene content and gene arrangement have been found in chigger mites of the genus *Leptotrombidium* (Table 5). In particular, *L. pallidum* differs from two congeneric species for the presence of a *rrnS* pseudogene, two identical and oppositely oriented *rrnL* genes, and four nearly identical extra non-coding regions. However, the arrangement of the shared *mt* genes is almost identical within the genus, except for the translocation of *trnQ* (Shao *et al.*, 2005b, 2006a). These peculiarities, combined with a gene arrangement drastically different from the ancestral arthropod gene order, have led to the hypothesis that the mtDNA AR of *L. pallidum* arose by a mechanism of 'duplication-random gene loss' followed by inter-genome non-homologous recombination (Shao *et al.*, 2006a). Interestingly, clues of mtDNA recombination have been found even in other congeneric comparisons of Acari. For instance, four prostriate ticks of the genus *Ixodes* share the same gene arrangement but differ in the number of CRs: non-Australasian *Ixodes* have a single CR, whereas Australasian *Ixodes* have two CRs undergoing concerted evolution, probably by the recombination mechanism of gene conversion (Shao *et al.*, 2005a).

Within Chelicerata, differences in gene content have also been reported for scorpions belonging to the genus *Mesobuthus*, as a *trnD* gene has been identified and annotated in *Mesobuthus gibbosus* (Jones *et al.*, 2007) but not in *Mesobuthus martensii* (Choi *et al.*, 2007). However, our comparative analyses demonstrate that *trnD* is absent in both *Mesobuthus* species, and that the reported difference is only apparent and related to the atypical cloverleaf structure of scorpion *tRNA* genes, which have D- and T-replacement loops and short anticodon stems (Davila *et al.*, 2005; Choi *et al.*, 2007; Jones *et al.*, 2007). In fact, we noted that the predicted *trnD* of *M. gibbosus* is quite bizarre: it has only three stems, a few paired nucleotides, and it is fully embedded inside (about 220 bp from the beginning of) the *rrnS* gene, whose 5' end is well conserved in scorpions and other arthropods (data not shown). Moreover, *trnD* is lost also in the other sequenced scorpion mtDNA (*Centruroides limpidus*; Davila *et al.*, 2005). Therefore, our re-annotation of scorpion mtDNAs strongly suggests that *trnD* is absent in all available scorpion species, and the gene content and genome AR is conserved in the *Mesobuthus* genus. Thus, the only intra-genus difference between *Mesobuthus* scorpions concerns genome size, and is due to the presence of repeated sequences in the CR.

In summary, the mtDNA AR variability observed in Arthropoda at the intra-genus level is mostly related to translocations, losses or acquisitions of *tRNA* genes. This observation reflects the general evolutionary trend of

arthropod mtDNA, where the most common rearrangements are translocations of *tRNA* genes (see decrease of  $\Delta\%$  in Table 1), and involve mostly genes adjacent to the CR or surrounding a *tRNA* gene cluster (*trnA-trnR-trnS(AGN)-trnE-trnF*) considered as the replication origin of the mtDNA minor strand (Boore, 1999).

#### The extraordinary variability of mollusc Bivalvia

Bivalve molluscs show frequent and extensive mtDNA variability at intra-genus level comparable to that observed in tunicates (Tables 4 and 5). In fact, two oyster species (*Crassostrea* genus) show broad differences in gene content and gene order (with relocation of most *tRNA* genes), together with other unusual mt features such as the presence of two *trnM(CAU)* genes and the split of *rrnL* into two fragments in both species, and the presence of a *cob* translational frameshift only in *Crassostrea virginica* (Milbury and Gaffney, 2005). These features are also associated with a high mt sequence divergence, comparable only to that observed between congeneric ascidians of the *Ciona* genus (Milbury and Gaffney, 2005).

Similar peculiarities have been also found in bivalves possessing two distinct mtDNA types, that is a female (F) and a male (M) type differentially transmitted to the progeny through DUI (Breton et al., 2007). In DUI species, the extent of genome rearrangement between the two gender-specific mtDNAs is quite variable, and, based on our mtDNA re-annotations, it ranges from the translocation of several genes in *Inversidens japonensis* (no gene inversion observed in our analysis), to differences in gene content in *V. philippinarum* and in some isoforms of the *Mytilus* genus (Table 5; Mizi et al., 2005; Breton et al., 2006; Zbawicka et al., 2007). The gene order is identical in the three analysed species of *Mytilus*, also described as members of the same species complex (Gossling, 1992), and duplication of the CR and some adjacent genes has been sporadically found only in some M and 'recently masculinized' mtDNA types of *Mytilus* (Mizi et al., 2005; Breton et al., 2006). In addition to these features, species with DUI show an accelerated nucleotide substitution rate compared to other animals, that is also much pronounced in the F than in M type (reviewed in Breton et al., 2007).

The differences here summarized for congeneric comparisons of Bivalvia are in accordance with the extraordinary variability previously described in this taxon, characterized by strong differences in genome AR, loss of *atp8*, presence of additional *trnM* genes, and broad variability in *tRNA* gene number (Tables 1 and 2). Moreover, the mt genomes of Bivalvia show a strong size variability (Figure 1).

#### Two distinct evolutionary trends in Nematoda

Within the phylum Nematoda, a surprising intra-genus variability in genome AR has been found in the class Enoplea (Table 5).

The unpublished mtDNAs of three *Romanomermis* species, enoplean parasites, reveal many gene duplications and extensive rearrangements, in accordance with the previous identification in *Romanomermis culicivora* of a 3-kb-long repeat encompassing protein-coding genes and organized in tandem copies with several deletions and inversions (Azevedo and Hyman, 1993). An even

more atypical AR variability has been found between distinct individuals of two enoplean obligate parasites, *T. cosgrovei* and *S. spiculatus*. In *T. cosgrovei* the mtDNAs of different individuals range in size from 19 to 34 kb, and greatly differ in the so-called hypervariable region of the genome (Tang and Hyman, 2007). In fact, the hypervariable segment is characterized by duplicated regions arranged in direct or inverted orientation, and contains both functional and non-functional genes organized in a shuffled gene order. The remaining portion of the genome, defined as 'constant', has an identical gene order in all analysed *T. cosgrovei* individuals (Tang and Hyman, 2007). In *S. spiculatus*, the unpublished small and large mtDNA haplotypes differs at least for the duplication of both *rRNA* genes in the large haplotype.

Compared to Enoplea, the mtDNA AR is quite stable in nematodes belonging to the Chromadorea class. In fact, we have observed no gene order differences in two chromadorean congeneric species pairs, that is between two members of the putative *Necator americanus* complex species (Hu et al., 2002, 2003a), and between two *Caenorhabditis elegans* and *Caenorhabditis briggsae* (the mtDNA of *C. briggsae* is a mitochondrial scaffold derived from the nuclear genome project; see Supplementary Material Table S1). Our comparative analyses have identified only the presence of additional non-coding regions (NC) in the mtDNA of *C. briggsae* compared to *C. elegans*, in accordance with the observed variability in number and size of non-coding regions found in nematodes. As consequence, there is a quite large difference in genome size between the two *Caenorhabditis* species (mean  $\pm$  s.d.: 14 107  $\pm$  443 bp).

In summary, congeneric comparisons in Nematoda suggest that gene order rearrangements and size variations are remarkably higher in Enoplea than in Chromadorea, an observation also confirmed by the analysis of the whole data set of nematodes (Table 1). The mtDNA of Nematoda is characterized by evolutionary dynamics strongly different from that of other metazoans, as it shows a fast nucleotide substitution rate (Lavrov and Lang, 2005), a very high AT content, and unusual characteristics related to the mitochondrial translational machinery, such as unique initiation codons (Wolstenholme, 1992; Hu et al., 2003b), unconventional cloverleaf *tRNA* structures with a TV-replacement loop (Okimoto and Wolstenholme, 1990) and small *rRNAs* whose reduced size is probably correlated with the unconventional *tRNA* structure (Okimoto et al., 1994). In addition, intra-genome recombination, a process likely absent in most metazoan mtDNAs, has been experimentally demonstrated in the enoplean *Meloidogynes javanica* (Lunt and Hyman, 1997), and is suggested to be the mechanism responsible of the intricate mtDNA AR found in the enoplean *Romanomermis* and *Thaumamermis* species (Azevedo and Hyman, 1993; Tang and Hyman, 2007). Recombination is also the more simple and plausible mechanism to explain mitochondrial gene inversions, which is the frequent exchange of coding strand observed in nematodes. Indeed, all Chromadorea nematodes are characterized by a single-coding strand (GSA = 1 in Table 1), whereas in Enoplea the *mt* genes are encoded by both strands following a distribution pattern highly different from one species to the other (so the GSA of Enoplea ranges from 0 to 0.80). Thus, it can be

hypothesized that gene inversions and recombination events have frequently occurred during the evolution of nematode mtDNAs.

#### Platyhelminthes: gene translocations with no inversions

Among the five intra-genus comparisons within Platyhelminthes (Table 4), only the genus *Schistosoma* shows gene order rearrangements (Table 5). In the six available *Schistosoma* species it has been observed the translocation of several tRNA and a few protein genes, the variability of genome size due to the presence of repeated sequences in the main non-coding region (Despres *et al.*, 1993; Littlewood *et al.*, 2006), and the conservation of gene transcriptional polarity (that is no gene inversions; Littlewood *et al.*, 2006). This intra-genus variability perfectly reflects the general evolutionary dynamics of the phylum Platyhelminthes, characterized by variable genome AR due to frequent translocation of tRNA genes (see  $\Delta\%$  in Table 1), presence of a single-coding strand in all species (Tables 1 and 3) and loss of *atp8* (Table 2). These characteristics have to be considered in the context of other peculiarities distinguishing the mtDNA of Platyhelminthes from other metazoans, such as the use of a unique genetic code (Bessho *et al.*, 1992; Telford *et al.*, 2000) and a very fast nucleotide substitution rate (Lavrov and Lang, 2005).

## Conclusions

The phylogenetic tree of Metazoa in Figure 2 shows the mitochondrial genomic features analysed here in an evolutionary context, and also depicts taxa with extremely fast or slow mt nucleotide substitution rates with long and short branches, respectively (see legend of Figure 2 for references).

Although the current view of metazoan phylogeny is partially resolved and the mtDNA sequences are extremely limited for many taxa (dashed lines in Figure 2), it is clear that the overall mtDNA plasticity in Metazoa is higher than previously thought. The gene category mainly responsible for this plasticity consists of the tRNA genes, whose number is highly variable both among and within taxa (Table 1), suggesting that metazoan mitochondria represent a quite flexible system with respect to their ability to import tRNAs from the cytoplasm and the tolerance of the mtDNA towards the acquisition or loss of functional tRNA genes. In contrast, the number of protein-coding genes is stable in metazoans, as a few additional proteins have been found in non-bilaterians (green boxes in Figure 2) and the *atp8* gene is lost in five evolutionary-distant taxa (grey boxes in Figure 2). The presence of additional proteins is considered as a primitive mtDNA character, shared also by non-metazoans, whereas the loss of *atp8* (as well as the presence of putative and highly modified isoforms of *atp8* in some rotifers and nematodes) is observed only in fast-evolving taxa. Thus, we can hypothesize that *atp8* is a 'dispensable' gene in the mtDNA, and that its loss has occurred several times during evolution, usually concomitant with an increase of the mt substitution rate. The genome AR is highly variable in several taxa scattered along the phylogenetic tree (yellow boxes in Figure 2), and most of these taxa do not show a fast nucleotide substitution rate, suggesting that the rate of genome rearrangements is independent from the substitution

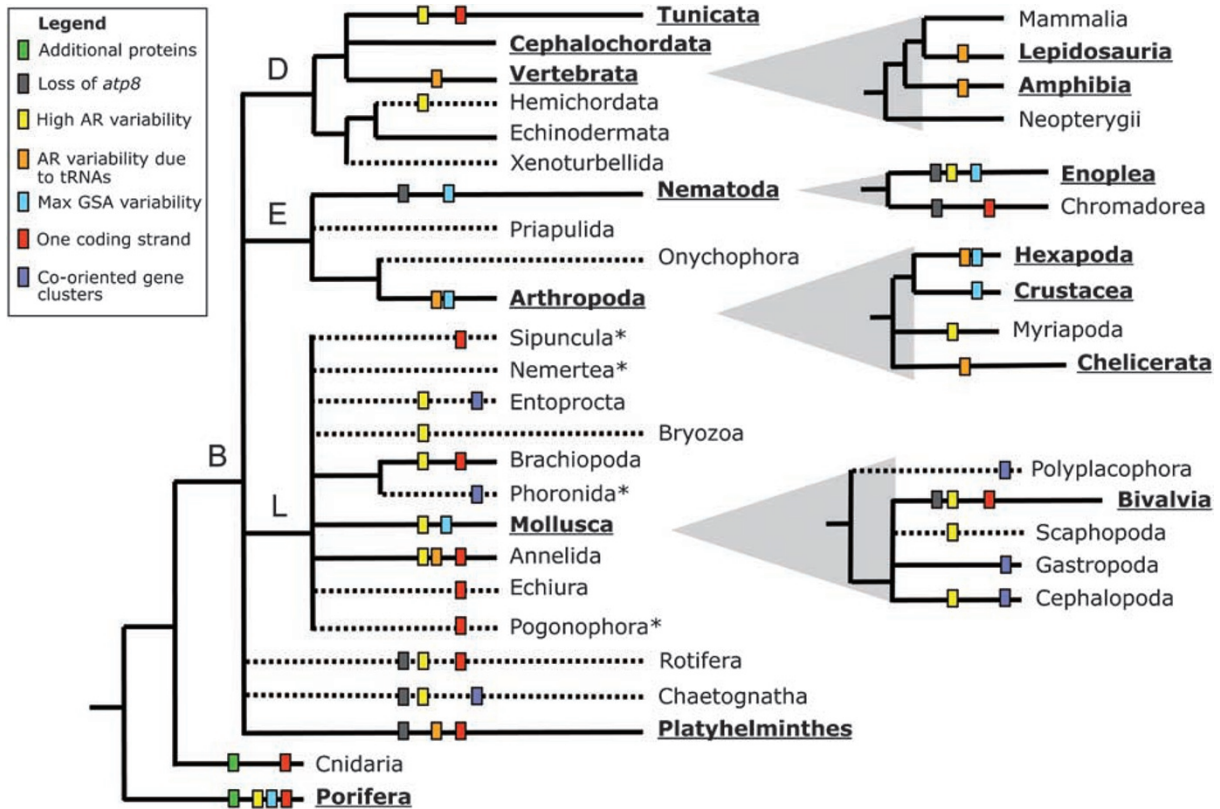
rate. Moreover, the AR variability is mainly due to the variation of number/position of tRNAs (orange boxes in Figure 2) only in taxa with moderate/low AR rate (absence of yellow boxes in Figure 2): only in Annelida we observe the concurrent presence of orange and yellow boxes in Figure 2. Similar data emerge also from the congeneric comparisons, as in the highest rearranged *Phallusia* and *Romanomermis* species, the BDN does not decrease excluding the tRNA genes (Table 5). This situation suggests that different rearrangement mechanisms, characterized by a differential involvement of the tRNA genes, could occur in taxa with a variable versus stable mtDNA AR, or that a 'saturation effect' in the gene order changes has obscured the involvement of tRNA in the highly rearranged mtDNAs.

The asymmetric distribution of the genes between the two strands is a feature common to almost all metazoans (Table 1). The most extreme GSA, characterized by all genes encoded on the same strand, has been observed in a wide taxonomic range including Tunicata, chromadorean Nematoda, Lophotrochozoa, Rotifera, Platyhelminthes and non-bilaterians (red boxes in Figure 2), leading to the hypothesis that this gene organization was present in the metazoan ancestor. It is also noteworthy that the exceptions to the rule of 'one coding-strand' found in non-bilaterians and lophotrochozoans (especially molluscs) consist mostly of mtDNAs with a few perfect or defective clusters of co-oriented genes (violet boxes in Figure 2). This further suggests that the ancestral gene organization was initially modified by inversion of large mtDNA segments, and then by strand translocation of single genes (the interrupting genes in Table 3). The observation of the maximum variability in GSA (GSA ranging from 0 to about 1; light blue in Figure 2) in two large ecdysozoan phyla is also in accordance with the theory of an ancestral 'one coding-strand' mtDNA structure that was slowly modified during metazoan diversification.

The data summarized here highlight that the mtDNA evolutionary trend has dramatically changed in vertebrates compared with other metazoans. Indeed, an almost invariant gene content, mtDNA AR and GSA characterizes the sub-phylum Vertebrata, and a general trend towards stabilization of these structural features is also present in remaining deuterostomes (see the few coloured boxes marking deuterostome branches in Figure 2, compared to the boxes abundance observed in other taxa). The most notable exception to the stability of deuterostome mtDNA concerns Tunicata, which exhibit variability in mtDNA AR and GSA similar to some lophotrochozoans and non-bilaterians. This similarity could suggest that the peculiar evolutionary trend of tunicate mtDNA is a primitive, rather than a derived mitochondrial trait of Chordata.

The differences in gene order and/or gene content found in congeneric comparisons species have always confirmed the genome plasticity observed at the highest taxonomic levels (Table 5 and Figure 2). As a result of these comparisons, we have confirmed that differences in mtDNA AR are essentially due to variation in the number and location of tRNA genes, whereas in the fast-evolving tunicates and enoplean species the observed differences remain high even when tRNA genes are excluded from analyses (compare BDN-All to BDN-tRNA in Table 5).





**Figure 2** Structural genomic features of metazoan mtDNA drawn into a phylogenetic context. Detailed explanation of coloured boxes: 'high architecture (AR) variability', AR rate calculated on all genes  $\geq 70$  (Table 1); 'AR variability due to tRNAs', percentage difference in AR variability  $\geq 60$  (see  $\Delta\%$  in Table 1); 'max gene strand asymmetry (GSA) variability', GSA ranging from 0 to almost 1 in most mtDNA of the metazoan group (see GSA in Table 1); 'one coding strand', presence of one coding strand (GSA = 1) in all, or almost all, mtDNAs of a taxa (Table 3); 'co-oriented gene clusters', genome AR with two or four clusters of co-oriented genes in most mtDNA of the metazoan group (Table 3). Low-sampled groups (only 1–2 mtDNAs) are represented by dashed lines. Asterisk indicates partial mtDNA sequences. Names in bold and underlined indicate the availability of congeneric species showing differences in mtDNA AR (Table 5). Long/short branches indicate fast/slow mt nucleotide substitution rates according to the following references: Tunicata (Yokobori *et al.*, 1999, 2005); Platyhelminthes (Lavrov and Lang, 2005); Rotifera Acanthocephala (Steinauer *et al.*, 2005); Chaetognatha (Papillon *et al.*, 2004); Onychophora (Podsiadlowski *et al.*, 2008); Nematoda (Lavrov and Lang, 2005); Bryozoa (Waeschenbach *et al.*, 2006); Chelicerata (Park *et al.*, 2007b); Bivalvia (Knudsen *et al.*, 2006); Cnidaria (Medina *et al.*, 2006) and Porifera (Lavrov and Lang, 2005). The metazoan phylogeny has been obtained from the 'Tree of life web project' (<http://tolweb.org/tree/phylogeny.html>), with the following exceptions: the deuterostome phylogeny was modified according to Zeng and Swalla (2005) and Bourlat *et al.* (2006); the chordate phylogeny is represented as unresolved, based on the controversy between traditional and molecular data (Delsuc *et al.*, 2006); Echiura and Pogonophora are represented as distinct phyla, although they are often included in Annelida. Myzostomida are not represented, as commonly included in Annelida. Abbreviations: B, Bilateria; D, Deuterostomia; E, Ecdysozoa; L, Lophotrochozoa.

Congeneric comparisons have often revealed more subtle genomic differences that are not easily identifiable in distant species: some studies have allowed the detailed reconstruction of evolutionary processes such as the pattern of gene duplication and pseudogene formation (in Plethodontidae salamanders and *Bipes* reptiles; Macey *et al.*, 2004; Mueller and Boore, 2005), and the origin of new non-coding or regulatory regions (that is in tunicates and *Ixodes* ticks; Shao *et al.*, 2005a; Iannelli *et al.*, 2007a,b). Thus, congeneric comparisons are very promising for the study of non-coding regions, the identification of regulatory sequences, and the investigation of rearrangement mechanisms. The advantages for evolutionary genomic studies, as well as for the setting of diagnostic test for species identification-monitoring, make the strategy of analysing the mtDNA in congeneric species a valuable approach, especially for phyla such as Ctenophora, Tardigrada, Kinorhyncha, Gastrotricha and Gnathostomulida for which complete mtDNA sequences are still absent.

## Acknowledgements

This work was supported by the 'Italian Ministry of University and Research' and the University of Milano (COFIN-2006 and FIRST projects to CG; FIRB project 'Laboratorio internazionale di Bioinformatica' to GP). We thank David Horner, Giulio Pavesi, Ernesto Picardi and Francesca Griggio for help in mtDNA analyses, and for valuable suggestions on the manuscript.

## References

- Akasaki T, Nikaido M, Tsuchiya K, Segawa S, Hasegawa M, Okada N (2006). Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. *Mol Phylogenet Evol* 38: 648–658.
- Azevedo JL, Hyman BC (1993). Molecular characterization of lengthy mitochondrial DNA duplications from the parasitic nematode *Romanomermis culicivorax*. *Genetics* 133: 933–942.

- Beagley CT, Macfarlane JL, Pont-Kingdon A, Okimoto R, Okada NA, Wolstenholme DR (1995). Mitochondrial genomes of Anthozoa (Cnidaria). In: Palmieri F, Papa S, Saccone C, Gadaleta MN (eds). *Progress in Cell Research*, 1st edn, vol. 5. Elsevier: Amsterdam. pp 149–153.
- Beagley CT, Okimoto R, Wolstenholme DR (1998). The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* **148**: 1091–1108.
- Beagley CT, Okimoto R, Wolstenholme DR (1999). *Mytilus* mitochondrial DNA contains a functional gene for a tRNA<sup>Ser</sup>(UCN) with a dihydrouridine arm-replacement loop and a pseudo-tRNA<sup>Ser</sup>(UCN) gene. *Genetics* **152**: 641–652.
- Bessho Y, Ohama T, Osawa S (1992). Planarian mitochondria. II. The unique genetic code as deduced from cytochrome c oxidase subunit I gene sequences. *J Mol Evol* **34**: 331–335.
- Bielawski JP, Gold JR (2002). Mutation patterns of mitochondrial H- and L-strand DNA in closely related Cyprinid fishes. *Genetics* **161**: 1589–1597.
- Blanchette M, Bourque G, Sankoff D (1997). Breakpoint phylogenies. *Genome Inform Ser Workshop Genome Inform* **8**: 25–34.
- Bleidorn C, Kruse I, Albrecht S, Bartolomeus T (2006). Mitochondrial sequence data expose the putative cosmopolitan polychaete *Scoloplos armiger* (Annelida, Orbiniidae) as a species complex. *BMC Evol Biol* **6**: 47.
- Boore JL (1999). Animal mitochondrial genomes. *Nucleic Acids Res* **27**: 1767–1780.
- Boore JL (2000). The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff D, Nadeau JH (eds). *Comparative Genomics*, 1st edn, vol. 1. Kluwer Academic: Dordrecht/Boston/London. pp 133–147.
- Boore JL, Brown WM (2000). Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol Biol Evol* **17**: 87–106.
- Boore JL, Lavrov DV, Brown WM (1998). Gene translocation links insects and crustaceans. *Nature* **392**: 667–668.
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M et al. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**: 85–88.
- Breton S, Beaupre HD, Stewart DT, Hoeh WR, Blier PU (2007). The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends Genet* **23**: 465–474.
- Breton S, Burger G, Stewart DT, Blier PU (2006). Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (*Mytilus* spp.). *Genetics* **172**: 1107–1119.
- Broughton RE, Reneau PC (2006). Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. *Mol Biol Evol* **23**: 1516–1524.
- Brugler MR, France SC (2007). The complete mitochondrial genome of the black coral *Chrysopathes formosa* (Cnidaria: Anthozoa: Antipatharia) supports classification of antipatharians within the subclass Hexacorallia. *Mol Phylogenet Evol* **42**: 776–788.
- Burger G, Forget L, Zhu Y, Gray MW, Lang BF (2003). Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci USA* **100**: 892–897.
- Burton RS, Byrne RJ, Rawson PD (2007). Three divergent mitochondrial genomes from California populations of the copepod *Tigriopus californicus*. *Gene* **403**: 53–59.
- Chen C, Dai CF, Plathong S, Chiou CY, Chen CA (2008). The complete mitochondrial genomes of needle corals, *Seriatopora* spp. (Scleractinia: Pocilloporidae): an idiosyncratic *atp8*, duplicated *trnW* gene, and hypervariable regions used to determine species phylogenies and recently diverged populations. *Mol Phylogenet Evol* **46**: 19–33. pub 21 September 2007.
- Choi EH, Park SJ, Jang KH, Hwang W (2007). Complete mitochondrial genome of a Chinese scorpion *Mesobuthus martensii* (Chelicerata, Scorpiones, Buthidae). *DNA Seq* **18**: 461–473.
- Cook CE (2005). The complete mitochondrial genome of the stomatopod crustacean *Squilla mantis*. *BMC Genomics* **6**: 105.
- Davila S, Pinero D, Bustos P, Cevallos MA, Davila G (2005). The mitochondrial genome sequence of the scorpion *Centruroides limpidus* (Karsch 1879) (Chelicerata; Arachnida). *Gene* **360**: 92–102.
- Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW et al. (2006). Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci USA* **103**: 8751–8756.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**: 965–968.
- Despres L, Imbert-Establet D, Monnerot M (1993). Molecular characterization of mitochondrial DNA provides evidence for the recent introduction of *Schistosoma mansoni* into America. *Mol Biochem Parasitol* **60**: 221–229.
- Dreyer H, Steiner G (2006). The complete sequences and gene organisation of the mitochondrial genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiattella arctica*—and the first record for a putative Atpase subunit 8 gene in marine bivalves. *Front Zool* **3**: 13.
- Elson JL, Lightowlers RN (2006). Mitochondrial DNA clonality in the dock: can surveillance swing the case? *Trends Genet* **22**: 603–607.
- Endo K, Noguchi Y, Ueshima R, Jacobs HT (2005). Novel repetitive structures, deviant protein-encoding sequences and unidentified ORFs in the mitochondrial genome of the brachiopod *Lingula anatina*. *J Mol Evol* **61**: 36–53.
- Erpenbeck D, Voigt O, Adamski M, Adamska M, Hooper JN, Worheide G et al. (2007). Mitochondrial diversity of early-branching metazoa is revealed by the complete mt genome of a haplosclerid demosponge. *Mol Biol Evol* **24**: 19–22.
- Faith JJ, Pollock DD (2003). Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics* **165**: 735–745.
- Faure E, Casanova JP (2006). Comparison of chaetognath mitochondrial genomes and phylogenetical implications. *Mitochondrion* **6**: 258–262.
- Feijao PC, Neiva LS, de Azeredo-Espin AM, Lessinger AC (2006). AMiGA: the arthropodan mitochondrial genomes accessible database. *Bioinformatics* **22**: 902–903.
- Flot JF, Tillier S (2007). The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: the putative D-loop and a novel ORF of unknown function. *Gene* **401**: 80–87.
- Fujita MK, Boore JL, Moritz C (2007). Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). *Mol Biol Evol* **24**: 2775–2786.
- Fukami H, Knowlton N (2005). Analysis of complete mitochondrial DNA sequences of three members of the *Montastraea annularis* coral species complex (Cnidaria, Anthozoa, Scleractinia). *Coral Reefs* **24**: 410–417.
- Gissi C, Iannelli F, Pesole G (2004). 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.
- Gissi C, Pesole G (2003). Transcript mapping and genome annotation of ascidian mtDNA using EST data. *Genome Res* **13**: 2203–2212.
- Goios A, Pereira L, Bogue M, Macaulay V, Amorim A (2007). mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Res* **17**: 293–298.

- Gossling E (1992). Systematic and geographic distribution of *Mytilus*. In: Gossling E (ed). *The Mussel Mytilus: Ecology, Physiology, Genetics and Aquaculture*, 1st edn, vol. 1. Elsevier: Amsterdam. pp 1–20.
- Gray MW (1999). Evolution of organellar genomes. *Curr Opin Genet Dev* 9: 678–687.
- Guo X, Liu S, Liu Y (2006). Evidence for recombination of mitochondrial DNA in triploid crucian carp. *Genetics* 172: 1745–1749.
- Guo X, Liu S, Liu Y (2007). Evidence for maternal inheritance of mitochondrial DNA in allotetraploid. *DNA Seq* 18: 247–256.
- Haen KM, Lang BF, Pomponi SA, Lavrov DV (2007). Glass sponges and bilaterian animals share derived mitochondrial genomic features: a common ancestry or parallel evolution? *Mol Biol Evol* 24: 1518–1527.
- Helfenbein KG, Boore JL (2004). The mitochondrial genome of *Phoronis architecta*—comparisons demonstrate that phoronids are lophotrochozoan protostomes. *Mol Biol Evol* 21: 153–157.
- Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL (2004). The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad Sci USA* 101: 10639–10643.
- Hoffmann RJ, Boore JL, Brown WM (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* 131: 397–412.
- Hu M, Chilton NB, Abs El-Osta YG, Gasser RB (2003a). Comparative analysis of mitochondrial genome data for *Necator americanus* from two endemic regions reveals substantial genetic variation. *Int J Parasitol* 33: 955–963.
- Hu M, Chilton NB, Gasser RB (2002). The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). *Int J Parasitol* 32: 145–158.
- Hu M, Chilton NB, Gasser RB (2003b). The mitochondrial genome of *Strongyloides stercoralis* (Nematoda)—idiosyncratic gene order and evolutionary implications. *Int J Parasitol* 33: 1393–1408.
- Huysse T, Plaisance L, Webster BL, Mo TA, Bakke TA, Bachmann L et al. (2007). The mitochondrial genome of *Gyrodactylus salaris* (Platyhelminthes: Monogenea), a pathogen of Atlantic salmon (*Salmo salar*). *Parasitology* 134: 739–747.
- Iannelli F, Griggio F, Pesole G, Gissi C (2007a). The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. *BMC Evol Biol* 7: 155.
- Iannelli F, Pesole G, Sordino P, Gissi C (2007b). Mitogenomics reveals two cryptic species in *Ciona intestinalis*. *Trends Genet* 23: 417–422.
- Inoue JG, Miya M, Tsukamoto K, Nishida M (2003). Evolution of the deep-sea gulper eel mitochondrial genomes: large-scale gene rearrangements originated within the eels. *Mol Biol Evol* 20: 1917–1924.
- Jameson D, Gibson AP, Hudelot C, Higgs PG (2003). OGRE: a relational database for comparative analysis of mitochondrial genomes. *Nucleic Acids Res* 31: 202–206.
- Jeyaprakash A, Hoy MA (2007). The mitochondrial genome of the predatory mite *Metaseiulus occidentalis* (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. *Gene* 391: 264–274.
- Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA et al. (2007). Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol Biol* 7: 123.
- Jones M, Gantenbein B, Fet V, Blaxter M (2007). The effect of model choice on phylogenetic inference using mitochondrial sequence data: lessons from the scorpions. *Mol Phylogenet Evol* 43: 583–595.
- Knudsen B, Kohn AB, Nahir B, McFadden CS, Moroz LL (2006). Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: conservation of the gene order in Euthyneura. *Mol Phylogenet Evol* 38: 459–469.
- Kon T, Nohara M, Nishida M, Sterrer W, Nishikawa T (2006). Hidden ancient diversification in the circumtropical lancelet *Asymmetron lucayanum* complex. *Mar Biol* 149.
- Kon T, Nohara M, Yamanoue Y, Fujiwara Y, Nishida M, Nishikawa T (2007). Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences. *BMC Evol Biol* 7: 127.
- Lavrov DV, Brown WM (2001a). *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics* 157: 621–637.
- Lavrov DV, Brown WM (2001b). *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics* 157: 621–637.
- Lavrov DV, Brown WM, Boore JL (2004). Phylogenetic position of the Pentastomida and (pan)crustacean relationships. *Proc Biol Sci* 271: 537–544.
- Lavrov DV, Forget L, Kelly M, Lang BF (2005). Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Mol Biol Evol* 22: 1231–1239.
- Lavrov DV, Lang BF (2005). Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. *Syst Biol* 54: 651–659.
- Le TH, Pearson MS, Blair D, Dai N, Zhang LH, McManus DP (2002). Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*. *Parasitology* 124: 97–112.
- Lee YS, Oh J, Kim YU, Kim N, Yang S, Hwang UW (2007). Mitome: dynamic and interactive database for comparative mitochondrial genomics in metazoan animals. *Nucleic Acids Res* 16: 16.
- Li X, Ogoh K, Ohba N, Liang X, Ohmiya Y (2007). Mitochondrial genomes of two luminous beetles, *Rhagophthalmus lufengensis* and *R. ohbai* (Arthropoda, Insecta, Coleoptera). *Gene* 392: 196–205.
- Lim JT, Hwang UW (2006). The complete mitochondrial genome of *Pollicipes mitella* (Crustacea, Maxillopoda, Cirripedia): non-monophyly of maxillopoda and crustacea. *Mol Cells* 22: 314–322.
- Littlewood DT, Lockyer AE, Webster BL, Johnston DA, Le TH (2006). The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. *Mol Phylogenet Evol* 39: 452–467.
- Lunt DH, Hyman BC (1997). Animal mitochondrial DNA recombination. *Nature* 387: 247.
- Macey JR, Papenfuss TJ, Kuehl JV, Fourcade HM, Boore JL (2004). Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. *Mol Phylogenet Evol* 33: 22–31.
- Machida RJ, Miya M, Nishida M, Nishida S (2002). Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). *Mar Biotechnol (NY)* 4: 406–417.
- Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL (2006). Naked corals: skeleton loss in Scleractinia. *Proc Natl Acad Sci USA* 103: 9096–9100.
- Milbury CA, Gaffney PM (2005). Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. *Mar Biotechnol (NY)* 7: 697–712.
- Minegishi Y, Aoyama J, Inoue JG, Miya M, Nishida M, Tsukamoto K (2005). Molecular phylogeny and evolution of the freshwater eels genus *Anguilla* based on the whole mitochondrial genome sequences. *Mol Phylogenet Evol* 34: 134–146.
- Mizi A, Zouros E, Moschonas N, Rodakis GC (2005). 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* **22**: 952–967.
- Moritz C, Brown WM (1986). Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. *Science* **233**: 1425–1427.
- Mueller RL, Boore JL (2005). Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Mol Biol Evol* **22**: 2104–2112.
- Nakao M, Abmed D, Yamasaki H, Ito A (2007). Mitochondrial genomes of the human broad tapeworms *Diphyllobothrium latum* and *Diphyllobothrium nihonkaiense* (Cestoda: Diphyllobothriidae). *Parasitol Res* **101**: 233–236.
- Nishibori M, Shimogiri T, Hayashi T, Yasue H (2005). Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. *Anim Genet* **36**: 367–375.
- Nishikawa T (2004). A new deep-water lancelet (Cephalochordata) from off Cape Nomamisaki, SW Japan, with a proposal of the revised system recovering the genus *Asymmetron*. *Zool Sci* **21**: 1131–1136.
- Nohara M, Nishida M, Miya M, Nishikawa T (2005a). Evolution of the mitochondrial genome in cephalochordata as inferred from complete nucleotide sequences from two *Epigonichthys* species. *J Mol Evol* **60**: 526–537.
- Nohara M, Nishida M, Nishikawa T (2005b). New complete mitochondrial DNA sequence of the lancelet *Branchiostoma lanceolatum* (Cephalochordata) and the identity of this species' sequences. *Zoolog Sci* **22**: 671–674.
- O'Brien EA, Zhang Y, Yang L, Wang E, Marie V, Lang BF et al. (2006). GOBASE—a database of organelle and bacterial genome information. *Nucleic Acids Res* **34**: D697–D699.
- Ohtsuki T, Sato A, Watanabe Y, Watanabe K (2002). A unique serine-specific elongation factor Tu found in nematode mitochondria. *Nat Struct Biol* **9**: 669–673.
- Ojala D, Montoya J, Attardi G (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**: 470–474.
- Okimoto R, Macfarlane JL, Wolstenholme DR (1994). The mitochondrial ribosomal RNA genes of the nematodes *Caenorhabditis elegans* and *Ascaris suum*: consensus secondary-structure models and conserved nucleotide sets for phylogenetic analysis. *J Mol Evol* **39**: 598–613.
- Okimoto R, Wolstenholme DR (1990). A set of tRNAs that lack either the T psi C arm or the dihydrouridine arm: towards a minimal tRNA adaptor. *EMBO J* **9**: 3405–3411.
- Papakonstantinou T, Law RH, Nagley P, Devenish RJ (1996a). Non-functional variants of yeast mitochondrial ATP synthase subunit 8 that assemble into the complex. *Biochem Mol Biol Int* **39**: 253–260.
- Papakonstantinou T, Law RH, Nesbitt WS, Nagley P, Devenish RJ (1996b). Molecular genetic analysis of the central hydrophobic domain of subunit 8 of yeast mitochondrial ATP synthase. *Curr Genet* **30**: 12–18.
- Papetti C, Lio P, Ruber L, Patarnello T, Zardoya R (2007). Antarctic fish mitochondrial genomes lack ND6 gene. *J Mol Evol* **65**: 519–528.
- Papillon D, Perez Y, Caubit X, Le Parco Y (2004). Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. *Mol Biol Evol* **21**: 2122–2129.
- Park JK, Kim KH, Kang S, Jeon HK, Kim JH, Littlewood DT et al. (2007a). Characterization of the mitochondrial genome of *Diphyllobothrium latum* (Cestoda: Pseudophyllidea)—implications for the phylogeny of eucestodes. *Parasitology* **134**: 749–759.
- Park SJ, Lee YS, Hwang UW (2007b). The complete mitochondrial genome of the sea spider *Achelia bituberculata* (Pycnogonida, Ammotheidae): arthropod ground pattern of gene arrangement. *BMC Genomics* **8**: 343.
- Peleg O, Kirzhner V, Trifonov E, Bolshoy A (2004). Overlapping messages and survivability. *J Mol Evol* **59**: 520–527.
- Pereira SL (2000). Mitochondrial genome organization and vertebrate phylogenetics. *Genet Mol Biol* **234**: 745–752.
- Plaisance L, Huysse T, Littlewood DT, Bakke TA, Bachmann L (2007). The complete mitochondrial DNA sequence of the monogenean *Gyrodactylus thymalli* (Platyhelminthes: Monogenea), a parasite of grayling (*Thymallus thymallus*). *Mol Biochem Parasitol* **154**: 190–194.
- Podsiadlowski L, Braband A, Mayer G (2008). The complete mitochondrial genome of the onychophoran *Epiperipatus billeyi* reveals a unique transfer RNA set and provides further support for the ecdysozoa hypothesis. *Mol Biol Evol* **25**: 42–51.
- Pont-Kingdon GA, Okada NA, Macfarlane JL, Beagley CT, Watkins-Sims CD, Cavalier-Smith T et al. (1998). Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue bacterial *MutS*: a possible case of gene transfer from the nucleus to the mitochondrion. *J Mol Evol* **46**: 419–431.
- Pont-Kingdon GA, Okada NA, Macfarlane JL, Beagley CT, Wolstenholme DR, Cavalier-Smith T et al. (1995). A coral mitochondrial *MutS* gene. *Nature* **375**: 109–111.
- Raina SZ, Faith JJ, Disotell TR, Seligmann H, Stewart CB, Pollock DD (2005). Evolution of base-substitution gradients in primate mitochondrial genomes. *Genome Res* **15**: 665–673.
- Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM et al. (2003). Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* **29**: 289–297.
- Reyes A, Gissi C, Pesole G, Saccone C (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol Biol Evol* **15**: 957–966.
- Sakurai M, Ohtsuki T, Watanabe Y, Watanabe K (2001). Requirement of modified residue m1A9 for EF-Tu binding to nematode mitochondrial tRNA lacking the T arm. *Nucleic Acids Res Suppl* **1**: 237–238.
- Sakurai M, Watanabe Y, Watanabe K, Ohtsuki T (2006). A protein extension to shorten RNA: elongated elongation factor-Tu recognizes the D-arm of T-armless tRNAs in nematode mitochondria. *Biochem J* **399**: 249–256.
- Serb JM, Lydeard C (2003). 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* **20**: 1854–1866.
- Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M (2005a). Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. *Mol Biol Evol* **22**: 620–629.
- Shao R, Barker SC, Mitani H, Takahashi M, Fukunaga M (2006a). Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium* (Acari: Acariformes). *J Mol Evol* **63**: 251–261.
- Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M (2005b). Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *J Mol Evol* **60**: 764–773.
- Shao Z, Graf S, Chaga OY, Lavrov DV (2006b). Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* **381**: 92–101.
- Signorovitch AY, Buss LW, Dellaporta SL (2007). Comparative genomics of large mitochondria in placozoans. *PLoS Genet* **3**: e13.
- Sinniger F, Chevallon P, Pawlowski J (2007). Mitochondrial genome of *Savalia savaglia* (Cnidaria, Hexacorallia) and early metazoan phylogeny. *J Mol Evol* **64**: 196–203.
- Spruyt N, Delarbre C, Gachelin G, Laudet V (1998). Complete sequence of the amphioxus (*Branchiostoma lanceolatum*)

- mitochondrial genome: relations to vertebrates. *Nucleic Acids Res* **26**: 3279–3285.
- Steinauer ML, Nickol BB, Broughton R, Orti G (2005). First sequenced mitochondrial genome from the phylum Acanthocephala (*Leptorhynchoides thecatus*) and its phylogenetic position within Metazoa. *J Mol Evol* **60**: 706–715.
- Suga K, Mark Welch DB, Tanaka Y, Sakakura Y, Hagiwara A (2008). Two circular chromosomes of unequal copy number make up the mitochondrial genome of the rotifer *Brachionus plicatilis*. *Mol Biol Evol* **7**: 7.
- Tang S, Hyman BC (2007). Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei*. *Genetics* **176**: 1139–1150.
- Telford MJ, Herniou EA, Russell RB, Littlewood DT (2000). Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc Natl Acad Sci USA* **97**: 11359–11364.
- Thompson RC, McManus DP (2002). Towards a taxonomic revision of the genus *Echinococcus*. *Trends Parasitol* **18**: 452–457.
- Valles Y, Halanych KM, Boore JL (2008). Group II introns break new boundaries: presence in a Bilaterian's genome. *PLoS ONE* **3**: e1488.
- Valverde JR, Marco R, Garesse R (1994). A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. *Proc Natl Acad Sci USA* **91**: 5368–5371.
- Waeschenbach A, Telford MJ, Porter JS, Littlewood DT (2006). The complete mitochondrial genome of *Flustrellidra hispida* and the phylogenetic position of Bryozoa among the Metazoa. *Mol Phylogenet Evol* **40**: 195–207.
- Wang X, Lavrov DV (2007). Mitochondrial genome of the homoscleromorph *Oscarella carmela* (Porifera, Demospongiae) reveals unexpected complexity in the common ancestor of sponges and other animals. *Mol Biol Evol* **24**: 363–373.
- Wolfsberg TG, Schafer S, Tatusov RL, Tatusov TA (2001). Organelle genome resource at NCBI. *Trends Biochem Sci* **26**: 199–203.
- Wolstenholme DR (1992). Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* **141**: 173–216.
- Yen MR, Tseng YH, Nguyen EH, Wu LF, Saier Jr MH (2002). Sequence and phylogenetic analyses of the twin-arginine targeting (*Tat*) protein export system. *Arch Microbiol* **177**: 441–450.
- Yokobori S, Fukuda N, Nakamura M, Aoyama T, Oshima T (2004). Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. *Mol Biol Evol* **21**: 2034–2046.
- Yokobori S, Oshima T, Wada H (2005). Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. *Mol Phylogenet Evol* **34**: 273–283.
- Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A *et al.* (1999). Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). *Genetics* **153**: 1851–1862.
- Yokobori S, Watanabe Y, Oshima T (2003). 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.
- Zbawicka M, Burzynski A, Wenne R (2007). Complete sequences of mitochondrial genomes from the Baltic mussel *Mytilus trossulus*. *Gene* **406**: 191–198.
- Zeng L, Swalla BJ (2005). Molecular phylogeny of protochordates: chordate evolution. *Can J Zool* **83**: 24–33.
- Zhang P, Zhou H, Liang D, Liu YF, Chen YQ, Qu LH (2005). 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.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)