J Mol Evol (2006) 63:375–392
DOI: 10.1007/s00239-005-0246-5

JOURNAL OF MOLECULAR EVOLUTION

The Relationship Between the Rate of Molecular Evolution and the Rate of Genome Rearrangement in Animal Mitochondrial Genomes

Wei Xu,[1] Daniel Jameson,[2] Bin Tang,[3] Paul G. Higgs[1]

[1] Department of Physics and Astronomy, McMaster University, Main St. West, Hamilton Ontario L8S 4M1, Canada
[2] Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester, UK
[3] Division of Genomics and Proteomics, Ontario Cancer Institute, University of Toronto, Suite 703, 620 University Avenue, Toronto, ON M5G 2M9, Canada

Received: 13 October 2005 / Accepted: 17 April 2006 [*Reviewing Editor:* Dr. David Pollock]

Abstract. Evolution of mitochondrial genes is far from clock-like. The substitution rate varies considerably between species, and there are many species that have a significantly increased rate with respect to their close relatives. There is also considerable variation among species in the rate of gene order rearrangement. Using a set of 55 complete arthropod mitochondrial genomes, we estimate the evolutionary distance from the common ancestor to each species using protein sequences, tRNA sequences, and breakpoint distances (a measure of the degree of genome rearrangement). All these distance measures are correlated. We use relative rate tests to compare pairs of related species in several animal phyla. In the majority of cases, the species with the more highly rearranged genome also has a significantly higher rate of sequence evolution. Species with higher amino acid substitution rates in mitochondria also have more variable amino acid composition in response to mutation pressure. We discuss the possible causes of variation in rates of sequence evolution and gene rearrangement among species and the possible reasons for the observed correlation between the two rates.

Key words: Mitochondrial genome — Genome rearrangement — Molecular clock — Relative rate test — Phylogenetics of arthropods

Correspondence to: Paul G. Higgs; *email:* higgsp@mcmaster.ca

Introduction

There are now hundreds of completely sequenced mitochondrial genomes, and we have therefore built up our own database, known as OGRe (Jameson et al. 2003) to facilitate comparative study of these genomes. Metazoan mitochondrial genomes are useful for phylogenetic studies because they contain a set of well-characterized genes that varies rather little between species. However, there are many sources of potential bias that occur in molecular phylogenetics, both in general and with mitochondrial sequences in particular. The signal for the deeper branches of a tree can be lost due to mutational saturation. There is great heterogeneity in the rates of evolution among species, which leads to substantial problems from long-branch attraction between the rapidly evolving species. Mitochondrial sequences are also heterogeneous in base and amino acid frequencies (Urbina et al. 2006), and this leads to bias in most phylogenetic methods that assume stationary models of evolution.

The problems of phylogenetic analysis at the sequence level do not directly influence analysis at the whole genome level. Gene content and gene order are known for many mitochondrial genomes. This gives information on the types of mechanism and selective forces influencing whole-genome evolution. It is also relevant for phylogenetics because changes in gene content and gene order can be good examples of shared derived characters that denote the common

ancestry of a given group. One example is the translocation of a tRNA-Leu gene that occurred in the common ancestor of hexapods and crustacea. This is an important line of evidence that supports the linking of these two groups to form the Pancrustacea (Boore et al. 1998). This argument was confirmed by Higgs et al. (2003) using a combination of sequence analysis and gene order data. Other examples using changes in mitochondrial gene order to derive phylogenetic information include Scouras and Smith (2001), Boore and Staton (2002), and Lavrov et al. (2004). Of course, there are many branches within a phylogenetic tree where no convenient gene order changes occur, so this type of analysis can only yield partial information.

There have also been several methods developed to infer evolutionary trees from sets of gene orders based on measuring breakpoint distances, inversion distances, or other types of edit distances between gene orders and on finding trees that optimize a predefined criterion related to these distances. Several programs are available that use these methods and some encouraging results have been reported (Blanchette et al. 1999; Sankoff et al. 2000a, b; Cosner et al. 2000; Bourque and Pevzner 2002; Larget et al. 2002; Moret et al. 2002). However, our own fairly extensive attempts to apply these methods to mitochondrial genomes have been disappointing, and we do not report them here.

An aspect of gene order analysis that we wish to emphasize here is that, just as the rates of sequence evolution vary greatly among species, so do the rates of gene order evolution. There are several animal phyla that contain species with conserved gene orders having substantial similarity with the best estimates of the ancestral gene orders, and also contain species with highly rearranged orders having almost no similarity to ancestral orders or to the orders of other extant species. The species with scrambled gene orders are the analogues of the long-branch species in sequenced based phylogenetics. These species will be very hard to position on a tree using gene order evidence. Although there is no obvious direct link between divergent sequences in sequence-based analysis and divergent gene orders in gene order analysis, Shao et al. (2003) have shown that, in practice, in insect genomes there appears to be a correlation between the two, i.e., the species that have highly divergent sequences also tend to have highly rearranged gene orders. In this paper we show that the same result applies in a broader-scale analysis of the arthropods and other animal phyla.

We began with the set of 55 complete arthropod genomes listed in Table 1 plus two nonarthropod outgroup species. This is the full set of arthropod genomes that was available to us at the time we began this analysis, with the exception that we excluded several other insects due to their high similarity with the listed species. Accession numbers of the complete mitochondrial genomes are given.

In the following section, we show that a fairly complete best-estimate tree for the arthropods can be obtained. Using this tree, we then estimate the degree of sequence divergence in each species by measuring the branch length along the tree from the ancestral arthropod to each species. We also measure the amount of gene order rearrangement in each species by comparing each gene order with the ancestral arthropod order. We find that sequence divergence and gene order rearrangement are correlated. We then use relative rate tests to investigate this effect with pairs of closely related species. In the remainder of the paper we show that this also applies in nonarthropod phyla and consider the possible causes of the effect.

A Best-Estimate Tree for the Arthropods

The topology of the tree in Fig. 1 is our best estimate of the arthropod tree derived from a combination of published sources. The relationship among the four principal arthropod groups has been debated for a long time, but evidence is now mounting to support the arrangement ((Chelicerata, Myriapoda), (Crustacea, Hexapoda)). The grouping of Crustacea and Hexapoda is known as Pancrustacea. It is supported by sequence evidence (Shultz and Regier 2000; Giribet et al. 2001) and by the tRNA-Leu translocation (Boore et al. 1998). The pairing of Chelicerata and Myriapoda is less certain but is suggested by the most recent results using combined 18S and 28S rRNA (Mallatt et al. 2004). An alternative possibility that cannot be ruled out is that Myriapoda is a sister group to Pancrustacea and that Chelicerata branches prior to this (Giribet et al. 2001; Pisani 2004).

As there are only two centipedes and two millipedes in our set, the phylogeny within the Myriapoda is not controversial. Within Chelicerata, the basal species is *Limulus* and the split between Acari and Araneae is not controversial. For the species in these latter two groups we take the classification from the NCBI taxonomy.

Although the Pancrustacea group as a whole is well supported, the arrangement of the early branching groups within it is very unclear. Several papers that include crustacean phylogenies are those by Regier and Shultz (1997), Shultz and Regier (2000), Wilson et al. (2000), Richter (2002), Mallatt et al. (2004), Lavrov et al. (2004), and Regier et al. (2005). However, there is no consensus of these results and we do not consider any of these to be definitive. We have therefore left a large number of groups branching simultaneously at this point. The

Table 1. List of species studied and accession numbers for the mitochondrial genomes

Chelicerata			
Xiphosura	*Limulus polyphemus*	NC_003057	L81949
Araneae	*Heptathela hangzhouensis*	NC_005924	AF062954 (*H. kimurae*)
Araneae	*Ornithoctonus huwena*	NC_005925	X13457 (*Aphonopelma* sp.)
Araneae	*Habronattus oregonensis*	NC_005942	AY210445 (*Misumenops asperatus*)
Acari	*Varroa destructor*	NC_004454	AY620940
Acari	*Carios capensis*	NC_005291	L76357 (*C. puertoricensis*)
Acari	*Ornithodoros moubata*	NC_004357	L76355
Acari	*Ornithodoros porcinus*	NC_005820	AF096274 (*O. coriaceus*)
Acari	*Rhipicephalus sanguineus*	NC_002074	AJ003815
Acari	*Amblyomma triguttatum*	NC_005963	AF018641
Acari	*Haemaphysalis flava*	NC_005292	Z74478 (*H. punctata*)
Acari	*Ixodes holocyclus*	NC_005293	AF018650
Acari	*Ixodes hexagonus*	NC_002010	AF115366 (*I. pilosus*)
Acari	*Ixodes persulcatus*	NC_004370	AY274888
Myriapoda			
Chilopoda	*Scutigera coleoptrata*	NC_005870	AF173238
Chilopoda	*Lithobius forficatus*	NC_002629	AF334271 (*L. obscurus*)
Diplopoda	*Thyropygus* sp.	NC_003344	AY210829 (*Orthoporus* sp.)
Diplopoda	*Narceus annularus*	NC_003343	AY288686 (*N. americanus*)
Crustacea			
Remipedia	*Speleonectes tulumensis*	NC_005938	L81936
Ostracoda	*Vargula hilgendorfii*	NC_005306	AB076654
Cephalocarida	*Hutchinsoniella macracantha*	NC_005937	AF370801
Copepoda	*Tigriopus japonicus*	NC_003979	AF363306 (*T. californicus*)
Pentastomida	*Armillifer armillatus*	NC_005934	AY744887 (*Raillietiella* sp.)
Branchiura	*Argulus americanus*	NC_005935	M27187 (*A. nobilis*)
Cirripedia	*Tetraclita japonica*	NC_008974	AY520640
Cirripedia	*Pollicipes polymerus*	NC_005936	AY520651
Malacostraca	*Penaeus monodon*	NC_002184	AF186250 (*P. vannamei*)
Malacostraca	*Cherax destructor*	NC_011243	AF235966 (*C. quadricarinatus*)
Malacostraca	*Portunus trituberculatus*	NC_005037	AY743951 (*Callinectes sapidus*)
Malacostraca	*Panulirus japonicus*	NC_004251	AF498670
Malacostraca	*Pagurus longicarpus*	NC_003058	AF436018
Branchiopoda	*Artemia franciscana*	NC_001620	AFR238061
Branchiopoda	*Triops cancriformis*	NC_004465	AF144219 (*T. longicaudatus*)
Branchiopoda	*Daphnia pulex*	NC_000844	AF014011
Hexapoda			
Collembola	*Tetrodontophora bielanensis*	NC_002735	AY555519
Collembola	*Gomphiocephalus hodgsoni*	NC_005438	AY596362 (*Hypogastrura* sp.)
Thysanura	*Tricholepidion gertschi*	NC_005437	AF370789
Orthoptera	*Locusta migratoria*	NC_001712	AF370793
Paraneoptera	*Aleurodicus dugesii*	NC_005939	ADU06474
Paraneoptera	*Triatoma dimidiata*	NC_002609	AJ243328
Paraneoptera	*Philaenus spumarius*	NC_005944	AY744779
Paraneoptera	*Thrips imaginis*	NC_004371	AY630445 (*Frankliniella* sp.)
Paraneoptera	*Lepidopsocid* RS-2001	NC_004816	AY630450 (*Lepium* sp.)
Paraneoptera	*Heterodoxus macropus*	NC_002651	AY077759 (*H. calabyi*)
Coleoptera	*Pyrocoelia rufa*	NC_003970	U65129 (*Photuris pennsylvanica*)
Coleoptera	*Tribolium castaneum*	NC_003081	AJ878603
Coleoptera	*Crioceris duodecimpunctata*	NC_003372	AJ781621 (*C. asparagi*)
Hymenoptera	*Apis mellifera ligustica*	NC_001566	AY703484
Hymenoptera	*Melipona bicolor*	NC_004529	AY773344 (*M. quinquefasciata*)
Lepidoptera	*Ostrinia furnacalis*	NC_003368	AF286298 (*Galleria mellonella*)
Lepidoptera	*Antheraea pernyi*	NC_004622	AF535029 (*Attacus ricini*)
Lepidoptera	*Bombyx mori*	NC_002355	AF286273 (*Hemileuca* sp.)
Diptera	*Anopheles gambiae*	NC_002084	AF440198 (*A. maculatus*)
Diptera	*Drosophila melanogaster*	NC_001709	M21017
Diptera	*Chrysomya putoria*	NC_002697	AF322424 (*Melinda viridicyanea*)

(continues)

Table 1. Continued

Outgroups			
Brachiopoda	*Terebratulina retusa*	NC_000941	U08324
Mollusca	*Katharina tunicata*	NC_001636	AY377650

Note. Accession numbers for the nuclear 18S rRNA genes are also listed. In cases where the same species is not available, the name of the closely related species that was used as a substitute is listed in parentheses.

subgroups of Pancrustacea that are well supported in these previous papers are the *Armillifer*/*Argulus* pair, Cirripedia, Malacostraca, Branchiopoda, Collembola, and Insecta. The relationship of Collembola and Insecta has been debated in recent papers (Nardi et al. 2003; Delsuc et al. 2003). If these two groups are not sisters, then Hexapoda is paraphyletic. However, we do not consider this matter resolved. For the species within Malacostraca we follow the tree of Morrison et al. (2002), and for the species within Branchiopoda we follow Spears and Abele (2000).

One of the most complete studies of the relationships of the insect orders is that by Wheeler et al. (2001), and we have followed this. Extracting the relevant groups for our data set from the summary Fig. 20 of Wheeler et al. gives (Thysanura, (Orthoptera, (Paraneoptera, (Coleoptera, (Hymenoptera, (Lepidoptera, Diptera)))))). The last four listed orders are holometabolous (insects that go through a full metamorphosis). The relationship between these orders is hard to resolve because of the unusual base composition of the Hymenoptera (*Apis* and *Melipona*). Castro and Dowton (2005) recently addressed this problem with a new genome from the Hymenoptera, *Perga condei*, not contained in our data set. The relationship between the orders depends on the evolutionary model used, but those authors concluded that when the most realistic models were used, Hymenoptera is a sister to (Lepidoptera + Diptera), as above.

The detailed phylogeny of species within the insect orders is largely noncontroversial for the genomes available, with the exception of the six species listed as Paraneoptera (which is a higher level taxon, not a single order). The species in our study are representatives of four different orders: Hemiptera (*Aleurodicus, Triatoma, Philaenus*), Thysanoptera (*Thrips*), Psocoptera (*Lepidosocid*), and Phthiraptera (*Heterodoxus*). We again followed Wheeler et al. (2001) for these orders.

Data and Methods

The amino acid sequences of cytochrome *b* and cytochrome *c* oxidase subunits I, II, and III were used. Each protein was aligned using T-Coffee (Notredame et al. 2002), short sections of poorly aligned sequence were deleted, and the four genes were concatenated. These four proteins are sufficiently well conserved (even for the most divergent species) that alignments covering almost the whole of the sequence length were used. The total length of the concatenated alignment was 1374 amino acids. For the tRNA sequence analysis, all 22 tRNAs on the mitochondrial genome were aligned individually, using the profile alignment facility of ClustalX (Thompson et al. 1997) to align new sequences to seed alignments previously available in our group. The alignments were manually adjusted to be consistent with the cloverleaf secondary structure. The unpaired regions inside the D loop and the TψC loop of the tRNAs were very variable in both sequence and length, and these were deleted. The remainder of the genes were concatenated, producing an alignment of length 1229 nucleotides. We analyzed the two types of sequence separately because we wished to see if similar effects arise in both. Mutational effects at the DNA level should influence the evolution of both proteins and RNAs, but selective effects might act differently on the two. It is particularly relevant to use tRNA sequences in this paper (rather than rRNAs) because we are interested in the relationship between gene order rearrangements and sequence evolution, and tRNAs are the genes that most frequently change position on the genome.

Maximum likelihood (ML) trees were obtained for proteins and tRNAs by determining ML branch lengths on this fixed topology using the PAML package (Yang 2002). For the proteins, we used the mtREV (Adachi and Hasegawa 1996) model and the REVaa model defined in PAML, in both cases using eight gamma-distributed rate categories. The former has fixed parameters, and the latter has variable parameters whose values were optimized using a procedure suggested by Z. Yang (personal communication). First, the REVaa rate matrix parameters were fixed to be equal to the mtREV values. Optimal values of the gamma distribution parameter and the branch lengths were then determined. Second, the gamma parameter and the branch lengths were fixed and new ML values for the rate matrix parameters were obtained. Finally, beginning from these initial values, the rate matrix parameters, the branch lengths, and the gamma parameter were all optimized simultaneously. The likelihoods of the optimal trees with the REVaa and mtREV models were compared using a likelihood ratio test, and the former was found to fit the data very much better. Hence, we show the tree from the REVaa model in Fig. 1, and we use branch lengths calculated with this model in the subsequent calculations.

For the tRNAs, we calculated ML trees using the HKY and general reversible models of evolution using the PAML package (eight gamma-distributed rates in both cases). Using a likelihood ratio test, we found that the general reversible model fit the data significantly better; therefore Fig. 2 shows the result with the general reversible model. We also used the FindModel server (Tao et al. 2005), which carries out the Modeltest program (Posada and Crandall 2001). Using the AIC, the general reversible model + gamma distribution was found to be better than all the simpler rate models tested, which confirms our expectations from the likelihood ratio test.

We measured the evolutionary distance to each arthropod species from the ancestral arthropod by taking the sum of the branch lengths on the path leading to each species from the basal

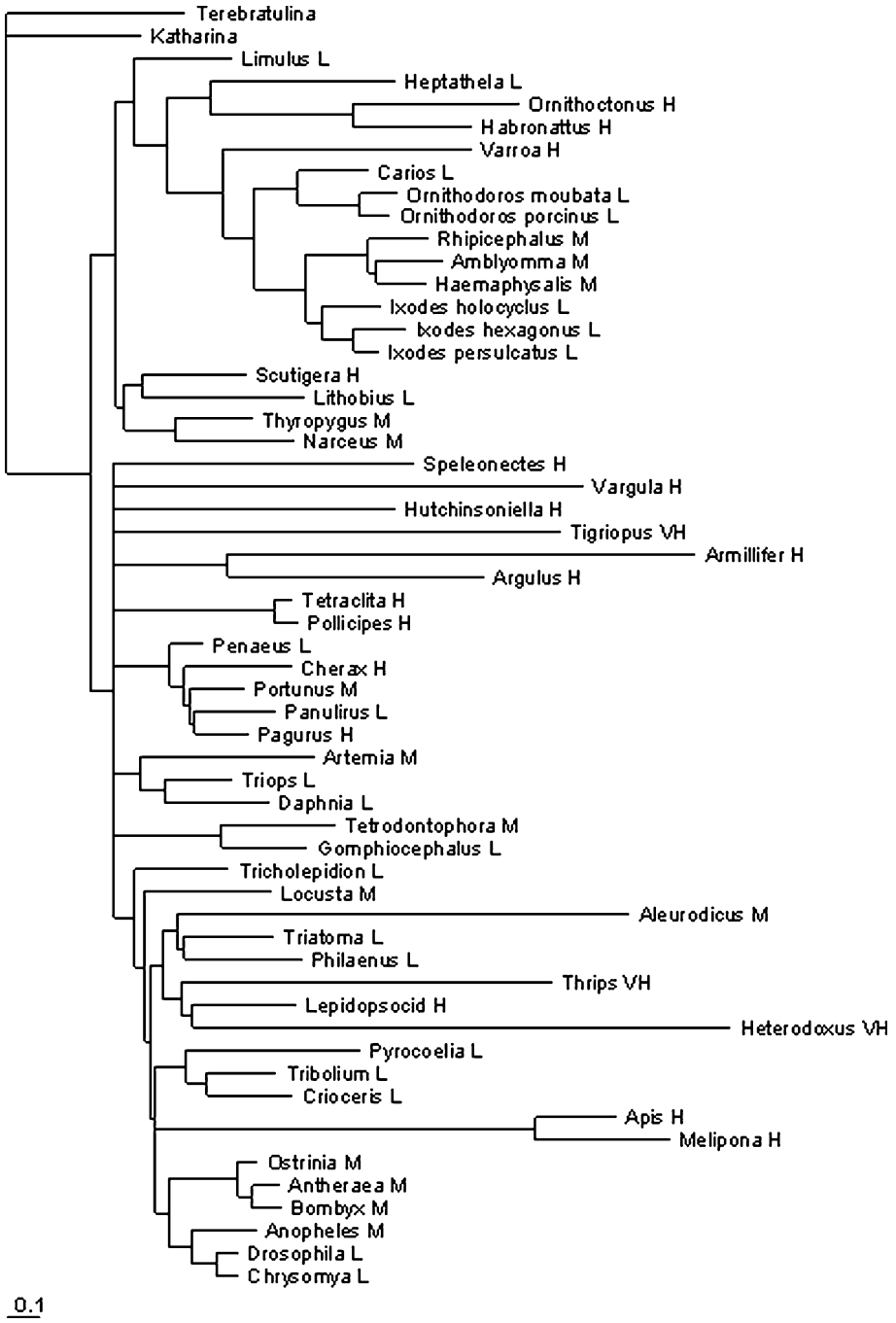

Fig. 1. Best-estimate tree using protein sequences with constrained topology and maximum likelihood branch lengths. VH, H, M, and L indicate the categories of very high, high, medium, and low breakpoint distance defined in Table 2.

split of the arthropods in Figs. 1 and 2. These values are listed as the protein and tRNA distances in Table 2. If evolution were strictly clocklike, all these distances would be equal. It can be seen that there is a wide variation in rates of evolution between species and that evolution is not clocklike. PAML also allows ML trees to be obtained with a global clock. A likelihood ratio test of the no-clock versus global-clock cases showed that the no-clock model fits the data very much better than the global clock.

We have also carried out our own phylogenetic studies with mitochondrial proteins and tRNAs, however, these did not resolve any additional branches on the tree that were not already well supported by the previous evidence used for the best-estimate tree. Therefore we do not show these results. Several of the long-branch species proved extremely difficult to position reliably on the tree using mitochondrial sequences. We therefore consider the best-estimate tree derived above to be more reliable than any of the tree topologies we obtained directly from these mitochondrial sequences.

The ancestral arthropod gene order was almost certainly the same as the present-day order of the horseshoe crab, *Limulus*. This order is also possessed by some of the Acari and Araneae; hence it must be basal to the chelicerate group. The most frequently occurring gene order in the arthropods is possessed by many members of the crustacean and hexapod groups (including *Drosophila*, *Penaeus*, *Daphnia*, etc.). Therefore the *Drosophila* order is almost certainly basal to the pancrustacea. The *Drosophila* order differs from that of *Limulus* by a single tRNA-Leu translocation, which appears to be a derived feature of the pancrustacea (not

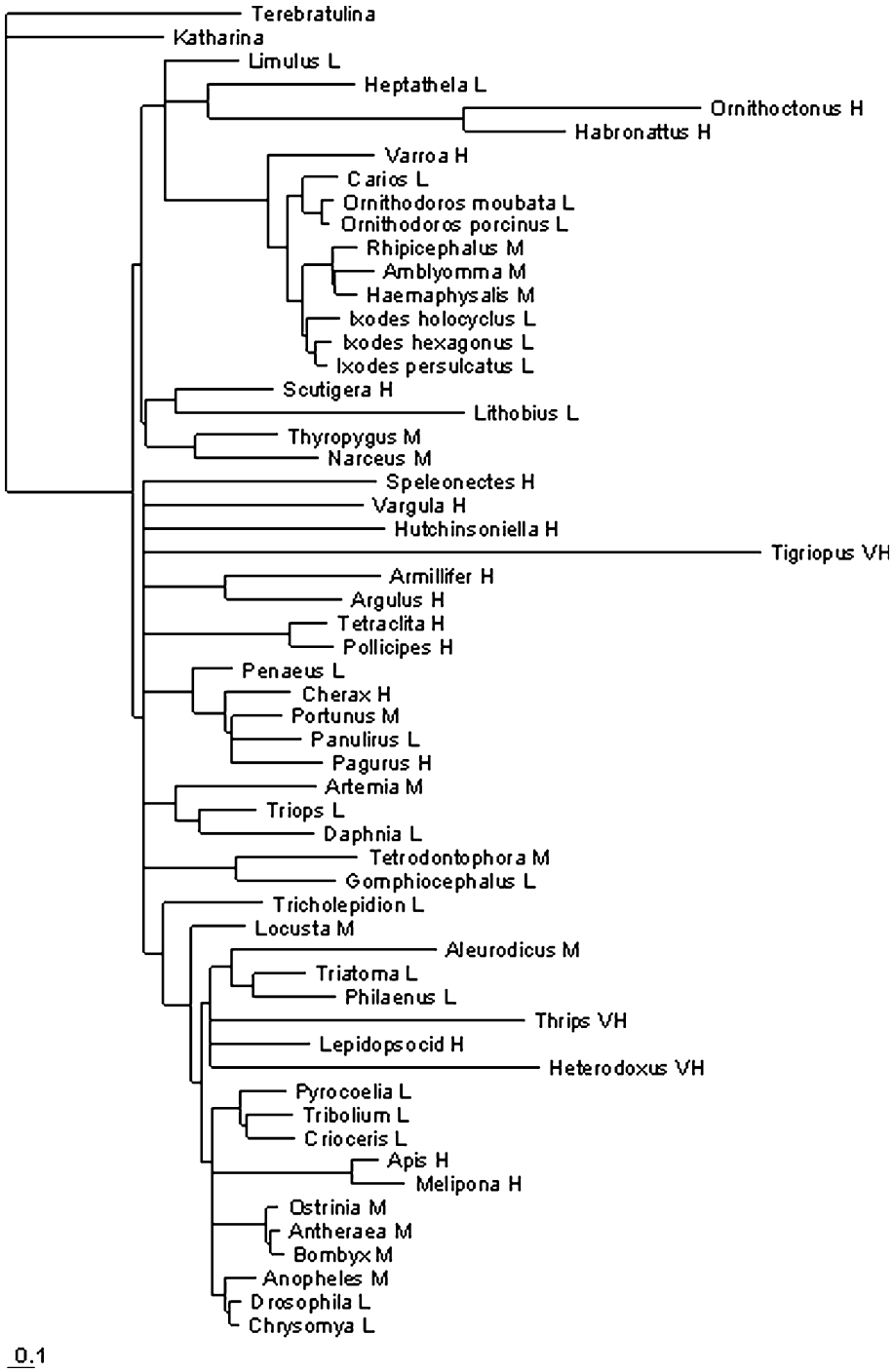

Fig. 2. Best-estimate tree using tRNA sequences with constrained topology and maximum likelihood branch lengths. VH, H, M, and L indicate the categories of very high, high, medium, and low breakpoint distance defined in Table 2.

ancestral to all arthropods). This strongly suggests that the ancestral arthropod order is the same as *Limulus*. This conclusion was also reached by Lavrov et al. (2004).

The simplest measure of the amount of genome rearrangement between two gene orders is the breakpoint distance (Blanchette et al. 1999). The two gene orders are examined for continuous sections where the relative gene order is the same in both. A breakpoint is a boundary between these continuous sections. Since mitochondrial genomes are circular, the number of breakpoints is equal to the number of continuous sections. When the two genomes contain identical sets of genes, the number of breakpoints is the same in both genomes. In these genomes, the gene sets vary slightly from the standard set of 37 genes due to the deletion or duplication of, at most, one or two genes. Where the gene sets are not identical, we define the breakpoint distance as the number of breakpoints in the larger of the two genomes. The breakpoint distances from *Limulus* to each of the arthropods are reported in Table 2.

A second measure of genome rearrangement is the inversion distance. For two gene orders with identical sets of genes it is always possible to transform one order into the other with a series of inversions. The inversion distance is the minimum number of inversions required to do this. In cases where two genomes contained nonidentical gene sets, we removed the additional genes from the larger of the two genomes and then calculated the number of inversions. This was done using the GRAPPA program (Moret

Table 2. Comparison of different distance measures between the ancestral arthropod and each present-day species: breakpoint distance (BP), number of inversions (Inv), number of duplications or deletions (D/D), tRNA sequence distance (tRNA), protein sequence distance (Prot), breakpoint distance with tRNAs excluded (BP*), and nuclear 18S rRNA distance (18S)

		BP	Inv	D/D	tRNA	Prot	BP*	18S
Very high	*Tigriopus japonicus*	35	32	0	2.15	1.34	14	0.30
	Heterodoxus macropus	35	32	0	1.39	1.83	12	0.34
	Thrips imaginis	32	29	1	1.34	1.32	11	0.22
High	*Pollicipes polymerus*	22	16	2	0.69	0.59	0	0.52
	Tetraclita japonica	20	16	0	0.66	0.57	0	0.51
	Argulus americanus	20	18	0	0.72	1.12	5	0.17
	Speleonectes tulumensis	19	16	1	0.83	0.93	3	2.84
	Apis mellifera	19	16	0	0.84	1.50	0	0.19
	Hutchinsoniella macracantha	18	16	0	0.86	0.87	0	0.59
	Pagurus longicarpus	18	12	0	0.65	0.45	5	0.19
	Vargula hilgendorfii	17	15	0	0.79	1.41	5	0.20
	Lepidopsocid RS-2001	17	16	0	0.60	0.59	3	0.23
	Cherax destructor	16	14	0	0.54	0.57	7	0.22
	Habronattus oregonensis	16	14	0	1.48	1.09	0	0.10
	Ornithoctonus huwena	15	13	0	1.95	1.23	0	0.08
	Scutigera coleoptrata	15	15	0	0.48	0.44	7	0.10
	Melipona bicolor	14	8	2	0.93	1.66	0	0.21
	Varroa destructor	14	12	0	0.83	1.09	0	0.49
	Armillifer armillatus	13	12	0	0.85	1.73	0	0.17
Medium	*Narceus annularus*	9	9	0	0.63	0.58	3	0.20
	Thyropygus sp.	9	9	0	0.49	0.46	3	0.11
	Aleurodicus dugesii	8	5	1	1.04	1.54	0	0.46
	Anopheles gambiae	8	6	0	0.41	0.47	0	0.81
	Tetrodontophora bielanensis	8	6	0	0.77	0.70	0	0.18
	Artemia franciscana	7	5	0	0.63	0.64	0	0.18
	Rhipicephalus sanguineus	7	6	0	0.82	0.96	3	0.15
	Amblyomma triguttatum	7	6	0	0.88	1.00	3	0.17
	Haemaphysalis flava	7	6	0	0.82	0.96	3	0.16
	Locusta migratoria	6	5	0	0.38	0.52	0	0.15
	Bombyx mori	6	5	0	0.51	0.54	0	0.27
	Portunus trituberculatus	6	5	0	0.51	0.44	0	0.18
	Ostrinia furnacalis	6	5	0	0.49	0.48	0	0.27
	Antheraea pernyi	6	5	0	0.50	0.54	0	0.30
Low	*Chrysomya putoria*	4	2	1	0.36	0.42	0	0.49
	Tricholepidion gertschi	3	2	0	0.44	0.39	0	0.20
	Daphnia pulex	3	2	0	0.62	0.51	0	0.19
	Pyrocoelia rufa	3	2	0	0.52	0.77	0	0.28
	Tribolium castaneum	3	2	0	0.55	0.53	0	0.18
	Drosophila melanogaster	3	2	0	0.37	0.42	0	0.48
	Panulirus japonicus	3	2	0	0.58	0.53	0	0.18
	Triatoma dimidiata	3	2	0	0.59	0.50	0	0.28
	Lithobius forficatus	3	3	0	1.13	0.61	0	0.08
	Philaenus spumarius	3	2	0	0.69	0.58	0	0.17
	Gomphiocephalus hodgsoni	3	2	0	0.69	0.62	0	0.14
	Penaeus monodon	3	2	0	0.34	0.32	0	0.28
	Crioceris duodecimpunctata	3	2	0	0.55	0.58	0	0.18
	Triops cancriformis	3	2	0	0.42	0.40	0	0.15
	Limulus polyphemus	0	0	0	0.36	0.40	0	0.06
	Heptathela hangzhouensis	0	0	0	0.76	0.87	0	0.07
	Ixodes persulcatus	0	0	0	0.72	0.82	0	0.15
	Ixodes holocyclus	0	0	0	0.76	0.83	0	0.14
	Ixodes hexagonus	0	0	0	0.74	0.90	0	0.15
	Carios capensis	0	0	0	0.70	0.79	0	0.18
	Ornithodoros porcinus	0	0	0	0.67	0.86	0	0.18
	Ornithodoros moubata	0	0	0	0.68	0.88	0	0.23

Note. Species are listed in descending order of breakpoint distance and have been divided into categories of very high, high, medium, and low breakpoint distance.

et al. 2002). The number of duplications and deletions in each species relative to *Limulus* is shown in the D/D column in Table 2, and the number of inversions after removal of any additional genes is also shown. A suitable measure of genome rearrangement accounting for both inversions and duplications/deletions is just the sum of these two columns in the table. In what follows, we

simply call this the inversion distance, since the number of duplications/deletions is always small. Note that the breakpoint distance already includes the effect of duplications/deletions because we defined it as being the number of breakpoints in the larger of the two genomes. Therefore it is not necessary to add the D/D column to the breakpoint column.

In Table 2, the species have been ranked in descending order of breakpoint distance. For convenience, we have also divided the species into four categories according to their breakpoint distances: very high (BP $\geq$ 32), high (13 $\leq$ BP $\leq$ 22), medium (6 $\leq$ BP $\leq$ 9), and low (BP $\leq$ 4). It is apparent from this table that gene order evolution is also nonclocklike. Some species are still identical in gene order to the ancestor, while others are completely scrambled. The highest BP value, 35, corresponds to a break point after almost every gene.

Results

Correlations Between Different Distance Measures

Table 3 lists the Pearson correlation coefficients between the distance measures. The distances are also compared graphically in Fig. 3. There is a very strong correlation between breakpoint distance and inversion distance ($R = 0.99$). This has also been demonstrated in other data sets (Blanchette et al. 1999; Cosner et al. 2000). We prefer breakpoint distance as our principal measure of genome rearrangement in this paper because it is the simplest measure to calculate and it does not presuppose any particular mechanism of rearrangement. If we were sure that inversions were the only rearrangement mechanism, it would make sense to use inversion distance. However, there are many cases of gene rearrangements where genes stay on the same strand, and this suggests that inversions are by no means the dominant mechanism. Calculations of edit distances accounting for both translocations and inversions are possible with heuristic search programs, but these are more complex than is necessary for interpretation of the present data.

There is also a fairly high correlation between the protein and the tRNA distances ($R = 0.69$). This suggests that there has been a speedup in the mutation rate in certain species that has affected both types of genes in a similar way. Figure 3 and Table 3 also show that there is a moderately strong correlation between breakpoint distance and the two measures of sequence distance ($R = 0.60$ and 0.54). Species with elevated rates of sequence evolution also tend to have elevated rates of genome rearrangement. Interpretation of the significance of these correlation coefficients is complicated by the fact that all species are related to one another. Estimates of distances from the common ancestor to each species are partially correlated because the earlier branches on the tree are shared. We consider statistical significance in more detail in the following section using relative rate tests. In this section we want to show the trends in the data in the simplest way. Table 4 reports the minimum, mean, and maximum of the tRNA and protein distances for species in each of the very high, high, medium and low breakpoint distance categories. There is a clear trend of increasing sequence-based distances with increasing breakpoint category.

It is interesting to compare the nature of the correlation between the two sequence-based distances and that between the breakpoint distance and the sequence-based distances. In the former case, the correlation is stronger for the shorter distances. If only the species for which the protein distance is ≤ 1 are included, the correlation between tRNA and protein distances increases to $R = 0.75$, whereas $R = 0.69$ when all species are included. The divergent species create scatter in this plot. The sequence-based distances depend on the many substitutions that occur along the length of the genes. Statistical error in the sequence-based distances should not be too large. Errors in sequence-based distances become larger when distances are larger because the alignments are less reliable, because the sequences may be approaching mutational saturation, and because the distance measures become more sensitive to the details of the evolutionary model used when distances are large. A greater degree of scatter in the long-distance species is therefore to be expected.

In contrast to this, the correlation between the breakpoint distance and the sequence-based distances is stronger for the long-distance species. If only the species in the low and medium breakpoint categories are included, then the correlation of breakpoint distance with protein and tRNA distance disappears altogether ($R < 0.005$ in both cases). This is partly attributable to the greater degree of scatter in breakpoint distances than sequence-based distances. A small number of rearrangement events contribute to the breakpoint distance when the breakpoint distance is small, whereas a large number of point mutations contribute to the sequence-based distances. Hence, if there is an underlying trend, we would expect this to be easier to see when the full range of breakpoint distances is included. However, the fact that the correlation disappears altogether for the less rearranged species suggests that there is really a qualitative difference between the highly rearranged and the less rearranged species. The majority of species have rather infrequent genome rearrangements, and for these species there is little relationship of the genome rearrangement rate to the sequence substitution rate, even though the two measures of sequence substitution rate are correlated for these species. The remaining species seem to have passed through a period of very frequent and complex genome rearrangements, and for these species there is a

Table 3. Correlation coefficients between the distance measures

	Breakpoint	Inversion	tRNA	Protein
Breakpoint	1.00	0.99	0.60	0.54
Inversion	0.99	1.00	0.61	0.54
tRNA	0.60	0.61	1.00	0.69
Protein	0.54	0.54	0.69	1.00

greatly increased rate of sequence substitution as well.

A notable point about genome rearrangement in mitochondrial genomes is that tRNA genes appear to move much more frequently than the "large" genes (proteins and rRNAs). This is easily demonstrated by considering the gene order of the large genes only, after elimination of the tRNAs. The BP* column in Table 2 shows the breakpoint distances from the ancestral order to each species, after exclusion of tRNAs. More than half the species in the high BP category have BP* = 0, i.e., the high numbers of genome rearrangement events in these species involve only the movement of tRNAs. The species with BP* = 0 were divided into two groups: those in the high BP category and those in the medium or low categories (there are no species with BP* = 0 in the very high BP category). Table 4 lists the minimum, mean and maximum of the sequence-based distances for these two groups of species. It can be seen that these distances are substantially larger for the high group than the medium/low group. This means that for the highly rearranged species with BP* = 0, there have been high rates of sequence substitution in both tRNAs and proteins, even though the proteins have not changed position on the genome.

Relative Rate Tests

Relative rate tests are used to determine whether the rates of evolution of two related species (1 and 2) are significantly different. This is done by comparing them both to an outgroup species. Let m_1 be the number of sites where sequence 1 is different but species 2 is the same as the outgroup, and let m_2 be the number of sites where species 2 differs from the other two. The quantity

$$\chi_m^2 = \frac{(m_1 - m_2)^2}{(m_1 + m_2)}$$

is measured, and its value is compared to a chi-square distribution with 1 degree of freedom (df) (Tajima 1993). Dowton (2004) proposed a test for the relative rate of genome rearrangement (RGR) between two species. The quantity

$$\chi_b^2 = \frac{(b_{01} - b_{02})^2}{(b_{01} + b_{02})}$$

is compared to a chi-square distribution, where b_{01} and b_{02} are the breakpoint distances from the outgroup gene order to the gene orders of species 1 and 2. As a concrete example, consider the following species: 1, *Laqueus rubellus*; 2; *Terebratulina retusa*; and 0, *Katharina tunicata*. Here, a mollusk is used as an outgroup to two brachiopods. From the gene orders (see http://ogre.mcmaster.ca) it is found that $b_{01} = 34$ and $b_{02} = 20$. Hence, $\chi_b^2 = 3.63$, and $p = 0.057$. Thus the test says that the rearrangement rate in *Laqueus* is not significantly faster than that in *Terebratulina* (or, at best, marginally so).

However, this test appears too conservative to us. *Katharina* was chosen as an outgroup because it is one of the least rearranged of the invertebrate genomes. Nevertheless, it is likely that there has been some genome rearrangement between *Katharina* and the common ancestor of the brachiopods. The RGR test looses power when the outgroup is too distant (Dowton 2004). We therefore propose the following modified RGR test. We consider pairs of adjacent genes on the genomes, which we call couples. Let n_i be the number of couples that are *not* present in species I but are present in the other species and the outgroup. We can assume that couples shared with the outgroup were present in the common ancestor of the pair. Therefore n_i is the number of couples that were present in the common ancestor and have been broken apart (i.e., a breakpoint has been inserted between them) along the branch to species i. The null assumption is that the probability that a couple is broken apart is the same on branches 1 and 2. Following the same derivation as for the Tajima test, the quantity

$$\chi_n^2 = \frac{(n_1 - n_2)^2}{(n_1 + n_2)}$$

should be compared to a chi-square distribution with 1 df. For the comparison of the two brachiopods given above, we have $n_1 = 14$ and $n_2 = 0$. This gives $\chi_n^2 = 14$ and $p = 1.8 \times 10^{-4}$, thus *Laqueus* is significantly more rearranged than *Terebratulina*, in contrast to the result with the more conservative test. We use the χ_n^2 RGR test in all the examples in this section.

The principal result from the previous section was that species with highly rearranged genomes appeared to have an increased rate of sequence evolution. We now test this using the relative rate tests. We consider each species in the high and very high breakpoint distance categories in turn and treat it as species 1 in a relative rate test. For species 2, we choose the closest relative to species 1 that is in the low breakpoint distance category. We then choose

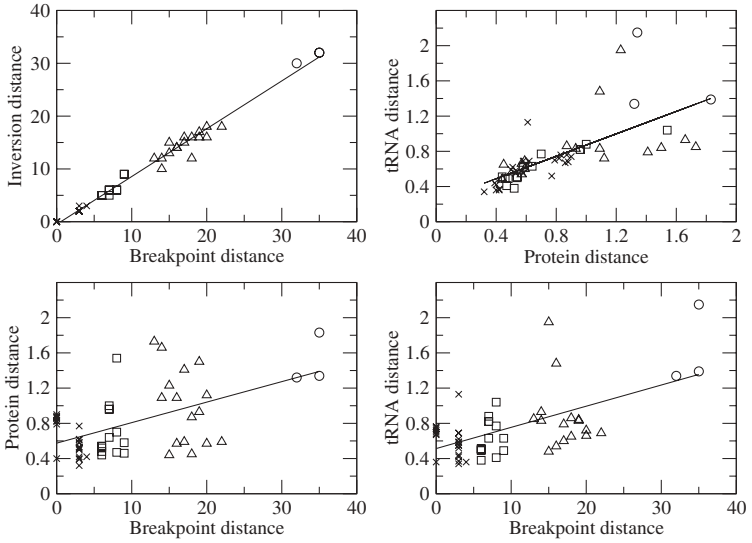

Fig. 3. Graphs showing the correlation between the different distance measures. Symbols indicate the four breakpoint distance categories: circles, very high; triangles, high; squares, medium; crosses, low. Lines are linear regressions through all points.

Table 4. Minimum, maximum, and mean values of the tRNA and protein distances for species in each of the breakpoint categories

	tRNA distance			Protein distance		
Breakpoint category	Min	Mean	Max	Min	Mean	Max
Very high	1.33	1.62	2.14	1.32	1.50	1.83
High	0.48	0.86	1.95	0.44	0.99	1.73
Medium	0.38	0.63	1.04	0.43	0.70	1.54
Low	0.34	0.60	1.13	0.32	0.61	0.90
BP* = 0						
High	0.66	1.01	1.94	0.57	1.15	1.73
Medium/low	0.34	0.60	1.13	0.32	0.63	1.54

the closest outgroup to this pair that is also in the low breakpoint distance category. Since we are deliberately comparing a high rearrangement and a low rearrangement species, we already know that $n_1 >> n_2$, and we expect the RGR test to confirm this. The important issue is to test the relative rates of sequence evolution of these same pairs of species. We therefore used the Tajima test on both the protein and the tRNA sequences for each species pair (see Table 5). Of the 19 highly rearranged species considered, 14 show significantly increased rates of both protein and tRNA evolution, and a further 2 show significantly increased protein rate but no significant increase in the tRNA rate. This confirms the correlation between genome rearrangement rate and sequence evolution rate. However, there are several exceptions that are worth noting. *Pagurus* shows no significant substitution rate increase relative to *Panulirus*, and *Lepidosocid* shows no significant increase relative to *Triatoma* for either protein or tRNA. The most notable exception is the *Scutigera*/*Lithobius* comparison, where the less rearranged species has a significantly higher sequence substitution rate for both proteins and tRNAs. Lavrov et al. (2002) noted that tRNA editing occurs in *Lithobius*, which could be linked to the high evolutionary rate of the tRNAs. Our result shows that there is also an unusually high rate of protein sequence evolution in *Lithobius*. We have not corrected for multiple testing in the results of Table 5, but it would make little difference since many of the p values are extremely low. Correction for multiple testing is an important issue when only a small number of tests are significant, whereas in our case, almost all of the tests are significant.

Table 6 shows a number of additional cases from among the arthropods that involve comparison of species in the medium breakpoint distance category with relatives in the low breakpoint distance category. There seems to be a significant speed-up in sequence substitution rate in *Aleurodicus* and *Artemia* even though the breakpoint distance is only 4 more than their comparison species. There is also a slight speed-up in the protein sequences in *Rhipicephalus*, but no indication of a rate increase in either *Tetrodontophora* or *Ostrinia*. In general, from Table 6 we see that when the breakpoint differences differ less from one another, fewer of the relative rate tests on the sequence evolution give a significant result.

We also looked for the correlation between highly rearranged genomes and high sequence substitution rate in nonarthropod species. According to recent phylogenetic analysis (Halanych 2004), the most important deep-level taxa in the bilaterian animal tree are the deuterostomes, the ecdysozoa, and the lophotrochozoa. Although we do not know exactly what the ancestral gene order was in each of these three taxa, we do have a good idea which of the currently existing gene orders is closest to the ancestral order. This is because there are representatives of each group that share sections of gene order with one another that appear to have been conserved since the time of the earliest bilaterians. We selected

Homo sapiens, *Limulus polyphemus*, and *Katharina tunicata* as conservative species whose gene orders are thought to be close to the ancestral orders of deuterostomes, ecdysozoa and lophotrochozoa, respectively.

As with Table 5, we compared a species that is known to be highly rearranged with a relative that is known to be less rearranged. Species 1 in each triplet was chosen to be one with a large breakpoint distance between it and the most closely related of the three conservative species. Species 2 was chosen to be a related species with a much lower level of gene rearrangement. An outgroup was chosen that also has a low level of genome rearrangement. The amino acid sequences from the three species for the same genes as in the arthropod study (*cox1, cox2, cox3*, and *cob*) were aligned, and the four alignments were concatenated. For all seven examples that we considered, there was a significant speed-up in the protein substitution rate in species 1 relative to species 2 (see Table 7).

We now briefly discuss the interpretation of each of the examples in Table 7. There is a significant speed-up in *Laqueus* relative to *Terebratulina*. The chiton *Katharina* (a mollusk) is a suitable outgroup, since chitons are thought to be the most basal molluscs (Serb and Lydeard 2003). The comparison of *Crassostrea* (a representative bivalve) with *Loligo* (a representative cephalopod) shows a significant increase in rate in bivalves relative to cephalopods. Gastropods are another mollusk group, most of whose genomes are quite highly rearranged. *Cepaea* (a representative gastropod) shows a significant rate increase realtive to *Haliotis*, another gastropod whose genome is less rearranged (unusually for this group). Knudsen et al. (2006) have recently discussed gene orders in mollusks and show that both divergent sequences and divergent gene orders cause problems in phylogenetics. Within the chordates, vertebrates are all quite conserved and the three available urochordates are highly divergent. The comparison of *Halocynthia* (a representative urochordate) with human demonstrates a speed-up in urochordates with respect to vertebrates. Echinoderms are another deuterostome group with relatively derived gene orders. The comparison of *Ophiopholis* with *Balanoglossus* shows a speed-up in echinoderms relative to hemichordates. Finally, nematodes and platyhelminths are two phyla in which all available genomes appear to have highly rearranged gene orders and highly divergent proteins. These phyla can be compared to less divergent phyla. Within ecdysozoa, there is a speed-up in nematodes relative to arthropods. Within lophotrochozoa, there is a speed-up in platyhelminths relative to mollusks.

Although the RGR test we used here seems to be an improvement over that proposed by Dowton (2004), we are still somewhat unsatisfied with it. The problem is that it assumes that the breakup of each of the shared gene couples is an independent event. In reality, an inversion creates two breakpoints and a translocation creates three breakpoints. Thus, up to three shared couples could disappear in a single event. The number of couples broken up by an inversion or translocation will depend on whether the breakpoints fall between the shared couples or elsewhere on the genome. It would be possible to devise a better model for genome rearrangement to use as the null model in the RGR test that would account for these effects. The significance derived using such a null model would depend on the relative rates of inversions and translocations (which is not known accurately). Even this more complicated model would not account for the fact that breakpoints seem to occur preferentially next to tRNA genes and close to the initiation and termination sites of genome replication. For the cases of interest here (Tables 5 and 7) we are deliberately comparing a species that is known to be highly rearranged with a relative that is less rearranged, so the RGR test just confirms what we already know. For the purposes of this paper, it does not seem worth pursuing the development of a more sophisticated RGR test.

Response of Amino Acid Frequencies to Mutational Pressure

The most straightforward explanation for an increase in the rate of substitution in a given lineage is that there has been an increase in the mutation rate. On the other hand, it can also be argued that a rate increase is due to positive selection on new sequence variants. It is difficult to see why positive selection would occur at many sites in many genes (including both RNAs and proteins) simultaneously; therefore it seems more reasonable to attribute the rate increase to mutation. As a way of testing this, we use a method we introduced recently (Urbina et al. 2006) to study the response of base frequencies in coding sequences to mutational pressure.

As mutation rates between the four bases are not equal, the equilibrium frequencies of the bases under mutation are not equal to one another. Substitutions at fourfold degenerate (FFD) sites are synonymous; therefore, neglecting any minor selective effects at the DNA level, the frequencies of bases at FFD sites should be determined by the equilibrium frequencies of the mutational process. Base frequencies at FFD sites vary substantially between mitochondrial genomes of different species. The base frequencies at the first and second positions are observed to vary in response, but the degree of variation is limited by selection at the amino acid level. Urbina et al. (2006) showed that

Table 5. Relative rate tests for all arthropods in the high or very high breakpoint categories

Species 1	Species 2	Outgroup	n_1	n_2	p (RGR)	p (protein)	p (tRNA)
Ornithoctonus	*Heptathela*	*Limulus*	15	0	1.1×10^{-4}	1.1×10^{-8}	9.4×10^{-10}
Habronattus	*Heptathela*	*Limulus*	16	0	6.3×10^{-5}	1.2×10^{-6}	1.1×10^{-5}
Varroa	*Carios*	*Limulus*	14	0	1.8×10^{-4}	3.4×10^{-8}	0.21 (NS)
Scutigera	*Lithobius*	*Limulus*	14	2	2.7×10^{-3}	9.2×10^{-4} (Opp)	2.1×10^{-8} (Opp)
Speleonectes	*Penaeus*	*Limulus*	16	0	6.3×10^{-5}	1.6×10^{-15}	2.1×10^{-7}
Vargula	*Penaeus*	*Limulus*	15	1	4.6×10^{-4}	1.8×10^{-26}	0.005
Hutchinsoniella	*Penaeus*	*Limulus*	16	1	2.7×10^{-4}	1.2×10^{-15}	1.9×10^{-5}
Tigriopus	*Penaeus*	*Limulus*	33	1	4.1×10^{-8}	4.6×10^{-32}	2.9×10^{-18}
Armillifer	*Penaeus*	*Limulus*	10	0	3.9×10^{-3}	3.2×10^{-32}	1.0×10^{-5}
Argulus	*Penaeus*	*Limulus*	17	0	3.7×10^{-5}	5.7×10^{-19}	0.069 (NS)
Tetraclita	*Penaeus*	*Limulus*	17	0	3.7×10^{-5}	5.9×10^{-4}	0.028
Pollicipes	*Penaeus*	*Limulus*	17	0	3.7×10^{-5}	4.3×10^{-4}	0.040
Cherax	*Penaeus*	*Daphnia*	13	0	3.1×10^{-4}	4.6×10^{-11}	2.4×10^{-4}
Pagurus	*Panulirus*	*Penaeus*	18	0	2.2×10^{-5}	0.002 (Opp)	0.60 (NS)
Thrips	*Triatoma*	*Tribolium*	30	0	4.3×10^{-8}	4.3×10^{-18}	1.0×10^{-9}
Lepidopsocid	*Triatoma*	*Tribolium*	15	0	1.1×10^{-4}	0.45 (Opp; NS)	0.12 (Opp; NS)
Heterodoxus	*Triatoma*	*Tribolium*	35	1	1.5×10^{-8}	6.0×10^{-35}	9.6×10^{-11}
Apis	*Drosophila*	*Tribolium*	16	0	6.3×10^{-5}	2.2×10^{-33}	4.4×10^{-6}
Melipona	*Drosophila*	*Tribolium*	11	0	9.1×10^{-4}	9.3×10^{-41}	4.8×10^{-7}

Note. Species 1 is the highly rearranged species in each comparison. n_1 and n_2 are the numbers of gene couples broken in the two branches; p (RGR) is the significance value for the relative genome rearrangement rate test; p (protein) and p (tRNA) are the significance values for the relative rate tests. NS, not significant; Opp, trend in the opposite direction to expectations.

Table 6. Additional relative rate tests among the arthropods

Species 1	Species 2	Outgroup	n_1	n_2	p (RGR)	P (protein)	p (tRNA)
Aleurodicus	*Triatoma*	*Tribolium*	4	0	0.045	1.1×10^{-25}	0.0035
Artemia	*Triops*	*Penaeus*	4	0	0.045	2.8×10^{-7}	8.9×10^{-10}
Rhipicephalus	*I. holocyclus*	*Carios*	7	0	0.008	0.007	0.55 (NS)
Tetrodontophora	*Gomphiocephalus*	*Tricholepidion*	5	0	0.025 (NS)	0.67 (NS)	0.37 (NS)
Ostrinia	*Drosophila*	*Tribolium*	3	0	0.083 (NS)	0.72 (Opp; NS)	0.55 (NS)

Note. For explanation, see Note to Table 5.

first position sites are more responsive than second position sites, which indicates that selection is stronger at second position. Amino acids whose codons differ by a second position mutation tend to more different from one another than those that differ by a first position mutation; therefore selection opposes second position mutations more strongly. Here we use the same model to compare first and second position site frequencies in the arthropods. The arthropod in this study were divided into a group with rapidly evolving proteins (those with a protein distance >1 in Table 2) and those with slowly evolving proteins (those with a protein distance ≤ 1). We show that the rapidly evolving species are more responsive to mutational pressure than the slowly evolving species.

The model is defined as follows. Let $f_{ik}^{(1)}$ and $f_{ik}^{(4)}$ be the frequencies of base k in species i at the first position and FFD sites, respectively. Only genes on the plus strand are considered in this analysis because the strands differ significantly in base frequencies. Suppose that there is a fraction ε_1 of first position sites where selection is negligible and the base is free to vary in the same way as at FFD sites and a fraction $1 - \varepsilon_1$ where selection is very strong and the base is not able to vary at all. Let $\phi_k^{(1)}$ be the frequency of base k at the strongly selected sites. The frequency of the bases in each species at first position should therefore be

$$f_{ik}^{(1)} = (1 - \varepsilon_1)\phi_k^{(1)} + \varepsilon_1 f_{ik}^{(4)}$$

According to the model, the first position frequencies will be a linear function of the FFD frequencies. This is found to apply quite well; see graphs of Urbina et al. (2006). To fit the model to the data it is necessary to perform a simultaneous least-squares fit of the data points for the four bases. The slope of the linear regressions is given by the parameter ε_1, and there are four parameters $\phi_{ik}^{(1)}$ that determine the intercepts. Similarly, let the frequencies at second position be $f_{ik}^{(2)}$. These values can be fitted with the same model:

Table 7. Relative rate tests of nonarthropod species

Species 1	Species 2	Outgroup	n_1	n_2	P (RGR)	p (protein)
Laqueus	*Terebratulina*	*Katharina*	14	0	1.8×10^{-4}	1.5×10^{-6}
Crassostrea	*Loligo*	*Katharina*	11	1	3.9×10^{-3}	1.6×10^{-31}
Cepaea	*Haliotis*	*Katharina*	23	0	1.6×10^{-6}	6.8×10^{-43}
Halocynthia	*Homo*	*Balanoglossus*	22	0	2.7×10^{-6}	5.8×10^{-21}
Ophiopholis	*Balanoglossus*	*Homo*	14	1	7.9×10^{-4}	1.5×10^{-12}
Caenorhabditis	*Limulus*	*Katharina*	17	2	5.8×10^{-4}	8.2×10^{-37}
Schistosoma	*Katharina*	*Limulus*	16	0	6.3×10^{-5}	5.6×10^{-67}

Note. For explanation, see Note to Table 5. Species details as follows: *Laqueus rubellus* (brachiopod), NC_002322; *Terebratulina retusa* (brachiopod), NC_000941; *Katharina tunicata* (chiton mollusk), NC_001636; *Crassostrea gigas* (bivalve mollusk), NC_001276; *Loligo bleekeri* (cephalopod mollusk), NC_002507; *Cepaea nemoralis* (gastropod mollusk), NC_001816; *Haliotis rubra* (gastropod mollusk), NC_005940; *Halocynthia roretzi* (urochordate), NC_002177; *Homo sapiens* (vertebrate), NC_001807; *Balanoglossus carnosus* (hemichordate), NC_001887; *Ophiopholis aculeata* (echinoderm), NC_005334; *Caenorhabditis elegans* (nematode), NC_001328; *Limulus polyphemus* (arthropod), NC_003057; *Schistosoma mansoni* (platyhelminth), NC_002545.

$$f_{ik}^{(2)} = (1 - \varepsilon_2)\phi_k^{(2)} + \varepsilon_2 f_{ik}^{(4)}$$

where the fraction of variable sites at second position is ε_2 and the frequencies of the bases in the strongly selected sites are $\phi_k^{(2)}$.

Table 8 gives the fitted parameter values for both sets of arthropods. All the slope parameters, ε, are positive, meaning that the mutation rate is sufficiently strong to cause variation at both positions, but all the slopes are < 1, meaning that both positions are more constrained by selection than FFD sites. Also, the first position slope is greater than the second position slope in both arthropod groups, meaning that selection is stronger at second position. This is the same effect that was seen with several other sets of species by Urbina et al. (2006). For the present paper, the important point is that the slopes are higher at both positions for the set of rapidly evolving species than for the set of slowly evolving species. The interpretation is that selection is trying to stabilize the amino acid frequencies at the optimal values required for the protein functions. Mutation pressure causes some variation away from this optimum. The fact that the rapidly evolving species vary more than the slowly evolving ones means that there is a higher mutation rate in these species that can more easily overcome stabilizing selection. If the rapidly evolving group was rapid because of large numbers of positively selected amino acid substitutions, there is no reason why the first and second position base frequencies should respond in a systematic way to the mutational frequencies.

Discussion

Here we consider the possible causes of the correlation between high rates of genome rearrangement and high rates of sequence substitution. For a long time it has been thought that the mitochondrial genome is replicated by an asymmetrical mechanism in which the H strand is copied in one direction beginning at an origin site O_H. Replication of the L begins some time later from a different site O_L and proceeds in the reverse direction (Shadel and Clayton 1997; Reyes et al. 1998; Bogenhagen and Clayton 2003). These studies are performed on mammalian genomes, and the same mechanism may not apply in other organisms. There has also been recent counter-evidence proposing an alternative model of replication in mammalian mitochondrial genomes (Yang et al. 2002; Bowmaker et al. 2003). Whatever the mechanism, it is clear that there is an asymmetry between the base compositions of the strands. Variations of base frequencies have also been found along the length of the genome that correlate with the length of time each part of the genome spends in a single-stranded state according to the asymmetric replication model (Reyes et al. 1998; Bielawski and Gold 2002; Faith and Pollock 2003; Krishnan et al. 2004; Raina et al. 2005).

Some of the key enzymes involved in replication are DNA polymerase γ (or POLG), mitochondrial single-strand binding protein, DNA ligase III, and Twinkle (a DNA helicase); see Kaguni (2004) and Korhonen et al. (2004). Amino acid substitutions in these nuclear-encoded proteins can lead to an increase in the mutation rate in the mitochondrial genome (Spelbrink et al. 2000; Del Bo et al. 2003; Wanrooij et al. 2004). Mutations in POLG and Twinkle can also lead to disorders characterized by depletion of mitochondrial genome copy number or by the presence of large deletions within the genome (Van Goethem et al. 2001; Zeviani et al. 2003). It seems likely that variation of the accuracy of the replication process between species is a major cause of the variation in evolutionary rates. Deleterious mutations in the enzymes responsible for DNA replication might lead to an increase in the error rate for both point mutations and genome rearrangements, which would explain the correlation between the two rates.

Table 8. Optimal parameters from fitting mutation pressure model to the two arthropod sets

	ε	ϕ_U	ϕ_C	ϕ_A	ϕ_G
Protein distance > 1					
1st position	0.363	30.8	15.4	29.0	24.8
2nd position	0.240	50.1	20.2	13.3	16.4
Protein distance ≤ 1					
1st position	0.298	28.9	18.0	28.5	24.7
2nd position	0.112	46.8	23.3	16.3	13.6

One important rearrangement mechanism is duplication and deletion of genes (Boore 2000). A tandem duplication of a region of the genome can occur due to slippage during replication. Duplicate copies of genes are likely to be rapidly eliminated or made nonfunctional by small deletions and point mutations. If the duplicated region contains more than one gene, then random deletion of one copy of each of the genes sometimes leads to reshuffling the order. This mechanism leaves all the genes on their original strand. Many of the rearrangements seen in the genomes in OGRe are consistent with this mechanism, although other mechanisms that lead to translocation of genes cannot be ruled out. There have been several recent studies reporting examples of gene rearrangements thought to have arisen by this mechanism (Dowton et al. 2003; Mueller and Boore 2005; Segawa and Aotsuka 2005). The fact that the gene duplication occurs at the time of genome replication again suggests that changes in the organism leading to a decrease in fidelity of genome replication are a major cause of high rates of genome rearrangement as well as point mutations.

A mechanism of inversion is required to explain rearrangements involving switching of genes between strands. Recombination within a circular genome can lead to excision of a smaller circle from a larger one or to inversion of a region of the genome, depending on the way the strands of DNA are reconnected (Lunt and Hyman 1997; Dowton and Campbell 2001). In humans, examples of mitochondrial genome variants containing large deletions have been found. These are known as sublimons (Kajander et al. 2000). Sublimons are found in small numbers in normal individuals but are present at a high frequency in patients with pathological conditions. Recombination of sublimons with one another or with the original genome would be a way of creating rearranged genomes with the full gene complement that might eventually replace the original version of the genome.

There have been several studies that show a relationship between the rate of molecular evolution and physiological properties of the organisms like generation time, metabolic rate, and body size (Li 1993; Martin and Palumbi 1993; Mooers and Harvey 1994; Gillooly et al. 2005). We have not attempted to test these effects with the current species. However, many of the rate increases observed here seem to occur rather sporadically in small groups of species (e.g., the bees versus the other insects or the two spiders, *Habronattus* and *Ornithoctonus*, versus the third), and this makes us doubt that something like generation time or body size has a major influence. To test a generation-time hypothesis in mitochondrial sequences, the replication time and turnover rates of the organelles themselves would be more relevant that the generation times of the organisms, and we do not have this information available.

As an indirect way of looking for correlations between the mitochondrial evolutionary rate and quantities like body size or generation time, we note that if these things were a major influence on evolutionary rates, we might expect them to influence both nuclear and mitochondrial sequences in the same way. It is therefore of interest to compare the mitochondrial sequence distances with those derived from the small subunit rRNA (18S) gene, the nuclear gene for which the most complete sequence information is available. For each of the species in the mitochondrial genome set, we obtained the 18S gene for the same species or a close relative (as detailed in Table 1) and aligned them. Using the same methods as above, we obtained the maximum likelihood branch lengths for these sequences with the same fixed best-estimate tree topology as before. The resulting tree is shown in Fig. 4, and the distances from the ancestral arthropod to each species are reported in Table 2. By far the largest of these distances is that for *Speleonectes*. For clarity, in Fig. 4 the branch leading to *Speleonectes* has been reduced by a factor of 3. If *Speleonectes* is excluded, the typical 18S distances are noticeably shorter than the mitochondrial protein and tRNA distances. Nevertheless, the degree of fluctuation in 18S distances is comparable to that of the mitochondrial sequence distances. Somewhat contrary to our expectations, it seems that the 18S evolution is no more clock-like than the mitochondrial sequences. There is no observable correlation between the 18S distances and the mitochondrial protein and tRNA distances in Table 2: $R = 0.03$ and $R = -0.01$ respectively. The single point from *Speleonectes* affects these numbers noticeably. The correlation coefficients become –0.04 and −0.11 if this species is excluded. Either way, there does not seem to be a relationship between the evolutionary rates in 18S and mitochondrial genes.

As we noted above, the mutational process differs between the two strands and also along each strand. If a gene happens to change position on the genome due to a rearrangement event, then the base frequencies within the gene will be out of equilibrium with the mutational process for the new position. This might

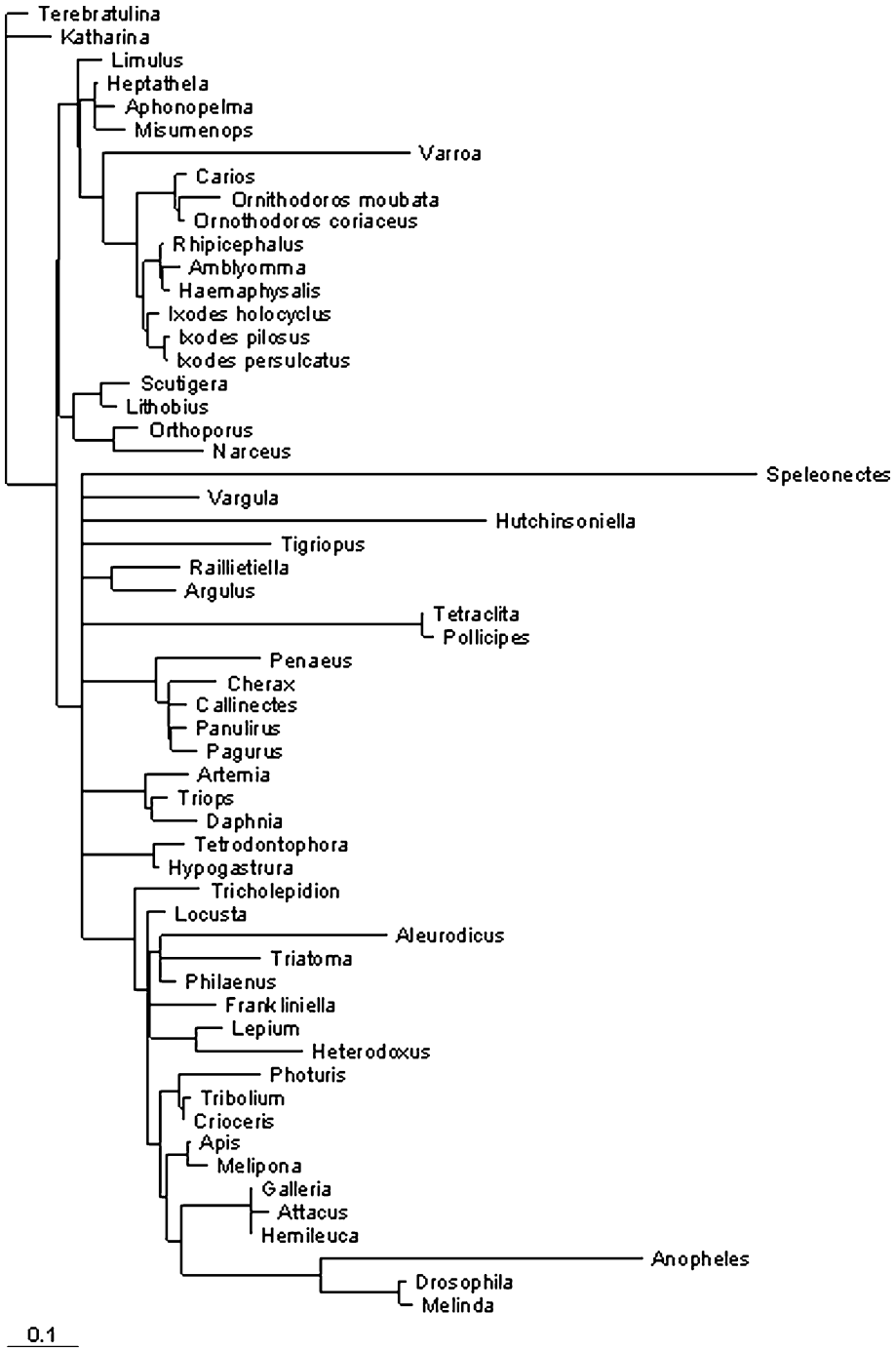

Fig. 4. Best-estimate tree using nuclear small subunit rRNA sequences with constrained topology and maximum likelihood branch lengths. The branch leading to *Speleonectes* is not to scale and is actually three times longer than that shown. The species names differ slightly from those in Figs. 1 and 2 in cases where the sequence from the exact same was not available. However, the species match closely and the topologies are the same. Therefore Figs. 1, 2, and 4 are directly comparable.

lead to a rapid burst of substitutions, particularly at synonymous sites, until equilibrium is reached. According to this argument, an increase in genome rearrangement rate would cause an increase in substitution rate. However, we already noted that for highly rearranged species where only the tRNA genes have moved, there appears to be an increase in substitution rate in both the proteins and the tRNAs. The increase in rate in the proteins cannot be due to them moving to a new position. In contrast to this, it is also possible to think of arguments where the causality goes in the opposite direction, i.e., where the genome rearrangement rate increases as a result of the increase in mutation rate. As mentioned above, although recombination is not a standard part of mitochondrial genome replication, there is some evidence that recombination occurs occasionally, and this would lead to gene reshuffling. It is possible that an increased mutational rate might lead to an increase in the rate of recombination events by creating repeated sequences that are prone to recombination (Samuels et al. 2004) or by creating similarities in gene sequences in different parts of the genome, such as two tRNA genes. It is often found that tRNAs genes occur at the ends of

rearranged fragments of mitochondrial genomes (Stanton et al. 1994), and the ability of these sequences to form stem-loop structures appears to be connected to the mechanism of rearrangement.

If, as we suggested initially, a major cause of the increase in rate in the rapidly evolving species is an increase in the error rate associated with genome replication, then the rate increase is due to mutation not selection. The analysis of the amino acid frequency variation above supports the argument that there is an increase in the point mutation rate in the species with rapidly evolving protein sequences. In a similar way, it is of interest to ask whether highly rearranged genomes arise due to an increase in the rate of random reshuffling events or because of selection for new gene orders. Clearly gene deletions are subject to selection if an essential gene is lost. However, selection can also act on variant gene orders, even when the gene content is the same, due to the mechanism of transcription. In mammalian mitochondria, transcription initiation sites have been identified for the two strands (Tracy and Stern 1995; Fernandez-Silva et al. 2003). Polycistronic RNAs are produced for each strand, which are subsequently processed into mRNAs for individual genes. Cleavage of the primary RNA transcripts occurs at positions either side of tRNA genes (Ojala et al. 1981). According to this model, tRNA genes are required between protein-coding genes in order to ensure proper RNA processing. Gene rearrangements that disrupt this processing mechanism would presumably be selected against. Nevertheless, it is clear that RNA processing is not entirely dependent on tRNAs. For example, the currently available complete genomes from cnidarians and chaetognaths have lost almost all their tRNAs (see diagrams of gene order at ogre.mcmaster.ca), and most genomes contain several positions with consecutive protein coding genes that are not separated by tRNAs.

The position of genes relative to transcription initiation sites can also determine the fate of duplicate gene copies after a gene duplication event. If one duplicate copy is not associated with an appropriate promoter, then this copy automatically becomes a pseudogene and will be lost. Lavrov et al. (2002) explained the rearrangements observed in two millipede genomes in terms of duplication followed by nonrandom loss of genes determined by the transcription direction. This mechanism can give rise to long strings of consecutive genes on the same strand. In fact, there are many species with gene orders where all the genes are on the same strand, including all known examples of acanthocephalans, annelids, brachiopods, cnidarians, echiurans, and platyhelminths, as well as some species of mollusks and nematodes. Many of these groups have arisen independently from ancestral orders that used both strands. It is unlikely that random reshuffling events would place all genes on one strand.

Mechanisms such as this can preferentially create certain gene orders and not others, so in this sense, gene orders are nonrandom. Nevertheless, this does not demonstrate that natural selection favours one gene rearrangement over another. As shown in Table 2, there are species with gene orders having almost no regions in common with the ancestral order. There seems to be no reason why these particular scrambled gene orders should be selected. The picture that emerges is that new gene orders are created by a range of reshuffling processes, and provided they satisfy certain constraints (such as the presence of all necessary genes and the existence of appropriate transcriptional promoters and RNA processing signals), new gene orders may be considered as (nearly) neutral variants of the original order. Selection is therefore acting to weed out inviable variants rather than to select new ones. This is exactly the argument put forward by proponents of neutral evolution theory at the sequence level: many mutations are deleterious, and selection acts to eliminate these, but most of the substitutions that are fixed in populations are due to (nearly) neutral mutations. This parallel seems a fitting point on which to conclude this study of the relationship between gene sequence evolution and gene rearrangement.

Acknowledgments. This work has been supported by Canada Research Chairs, NSERC (Canada), and BBSRC (UK).

References

Adachi J, Hasegawa M (1996) A model of amino acid substitution in proteins encoded by mitochondrial DNA. J Mol Evol 42:459–468

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, Kunisawa T, Sankoff D (1999) Gene order breakpoint evidence in animal mitochondrial phylogeny. J Mol Evol 49:193–203

Bogenhagen DF, Clayton DA (2003) The mitochondrial DNA replication bubble has not burst. Trends Biochem Sci 28:357–360

Bowmaker M, Yang MY, Yasukawa T, Reyes A, Jacobs HT, Huberman JA, Holt IJ (2003) Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. J Biol Chem 278:50961–50960

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. Kluwer Academic, Dordrecht, pp 133–147

Boore JL, Staton JL (2002) The mitochondrial genome of the Sipunculid Phascolopsis gouldii supports its association with Annelida rather than Mollusca. Mol Biol Evol 19:127–137

Boore JL, Lavrov DV, Brown WM (1998) Gene translocation links insects and crustaceans. Nature 392:667–668

Bourque G, Pevzner PA (2002) Genome-scale evolution: reconstructing gene orders in ancestral species. Genome Res 12:26–36

Castro LR, Dowton M (2005) The position of the Hymenoptera within the Holometabola is inferred from the mitochondrial genome of *Perga condei* (Hymenoptera:Symphyta:Pergidae). Mol Phylogenet Evol 34:469–470

Cosner ME, Jansen RK, Moret BME, Raubeson LA, Wang LS, Wanrnow T, Wyman S (2000) An empirical comparison of phylogenetic methods on chloroplast gene order data in Campanulaceae. In: Sankoff D, Nadeau JH (eds.) Comparative genomics. Kluwer Academic, Dordrecht, pp 99–121

Del Bo R, Bordoni A, Sciacco M, Di Fonzo A, Galbiati S, Crimi M, Bresolin N, Comi GP (2003) Remarkable infidelity of polymerase gamma A associated with mutations in POLG1 exonuclease domain. Neurology 61:903–908

Delsuc F, Phillips MJ, Penny D (2003) Comment on ''Hexapod Origins: Monophyletic or Paraphyletic?'' Science 301:1482d

Dowton M (2004) Assessing the relative rate of (mitochondrial) genomic change. Genetics 167:1027–1030

Dowton M, Campbell NJH (2001) Intramitochondrial recombination—Is it why some mitochondrial genes sleep around. Trends Ecol Evol 16:269–271

Dowton M, Castro LR, Campbell SL, Bargon SD, Austin AD (2003) Frequent mitochondrial gene rearrangements at the hymenopteran nad3−nad5 junction. J Mol Evol 56:517–526

Faith JJ, Pollock DD (2003) Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. Genetics 165:735–745

Fernandez-Silva P, Enriquez JA, Montoya J (2003) Replication and transcription of mammalian mitochondrial DNA. Exp Physiol 88:41–56

Gillooly JF, Allen AP, West GB, Brown JH (2005) The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc Natl Acad Sci USA 102:140–145

Giribet G, Edgecombe GD, Wheeler WC (2001) Arthropod phylogeny based on eight molecular loci and morphology. Nature 413:157–161

Halanych KM (2004) The new view of animal phylogeny. Annu Rev Ecol Evol Syst 35:229–256

Higgs PG, Jameson D, Jow H, Rattray M (2003) The evolution of tRNA-Leucine genes in animal mitochondrial genomes. J Mol Evol 57:435–445

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

Kaguni LS (2004) DNA polymerase γ, the mitochondrial replicase. Annu Rev Biochem 73:293–320

Kajander OA, Rovio AT, Majamaa K, Poulton J, Spelbrink JN, Holt IJ, Karhunen PJ, Jacobs HT (2000) Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. Hum Mol Genet 9:2821–2835

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 38:459–460

Korhonen JA, Pham XH, Pellegrini M, Falkenberg M (2004) Reconstruction of a minimal mtDNA replisome *in vitro*. EMBO J 23:2420–2423

Krishnan NM, Seligmann H, Raina SZ, Pollock DD (2004) Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes. DNA Cell Biol 23:707–714

Larget B, Simon DL, Kadane JB (2002) Bayesian phylogenetic inference from animal mitochondrial genome arrangements. J Roy Stat Soc B 64:681–693

Lavrov DV, Boore JL, Brown WM (2002) Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and non-random loss. Mol Biol Evol 19:160–163

Lavrov DV, Brown WM, Boore JL (2004) Phylogenetic position of the Pentastomida and (pan) crustacean relationships. Proc Roy Soc Lond B 271:537–544

Li WH (1993) So what about the molecular clock hypothesis? Curr Opin Genet Dev 3:896–901

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

Mallatt JM, Garey JR, Shultz JW (2004) Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. Mol Phylogenet Evol 31:178–191

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci USA 90:4087–4091

Mooers AO, Harvey PH (1994) Metabolic rate, generation time, and the rate of molecular evolution in birds. Mol Phylogenet Evol 3:344–350

Moret BME, Siepel AC, Tang J, Liu T (2002) Inversion medians outperform breakpoint medians in phylogeny reconstruction from gene order data. Lect Notes Comp Sci 2452:521–536

Morrison CL, Harvey AW, Lavery S, Tieu K, Huang Y, Cunningham CW (2002) Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. Proc R Soc Lond B 269:345–350

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

Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: Monophyletic or paraphyletic? Science 299:1887–1880 (see also Science 301:1482e)

Notredame C, Higgins DG, Heringa J (2000) T–Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Ojala D, Montoya A, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474

Pisani D (2004) Identifying and removing fast-evolving sites using compatibility analysis: an example from the Arthropoda. Syst Biol 53:978–980

Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. Syst Biol 50:580–601

Raina SZ, Faith JJ, Dusotell TR, Seligmann H, Stewart CB, Pollock DD (2005) Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res 15:665–673

Regier JC, Shultz JW (1997) Molecular phylogeny of the major arthorpod groups indicates polyphyly of the crustaceans and a new hypothesis for the origin of hexapods. Mol Biol Evol 14:909–913

Regier JC, Shultz JW, Kambic RE (2005) Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc Roy Soc Lond B 272:395–401

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

Richter S (2002) The Tetraconata concept: hexapod-crustacean relationships and the phylogeny of Crustacea. Organisms Diversity Evol 2:217–237

Samuels DC, Schon EA, Chinnery PF (2004) Two direct repeats cause most human mtDNA deletions. Trends Genet 20:393–398

Sankoff D, Deneault M, Bryant D, Lemieux C, Turmel M (2000a) Chloroplast gene order and the divergence of plants and algae, from the normalized number of induced breakpoints. In: Sankoff D, Nadeau JH (eds.) Comparative genomics. Kluwer Academic, Dordrecht, pp 89–98

Sankoff D, Bryant D, Deneault M, Lang BF, Burger G (2000b) Early eukaryote evolution based on mitochondrial gene order breakpoints. J Comp Biol 7:521–535

Scouras A, Smith MJ (2001) A novel gene order in the crinoid echinoderm *Florometra serratissima*. Mol Biol Evol 18:61–73

Segawa RD, Aotsuka T (2005) The mitochondrial genome of the Japanese freshwater crab, Geothelphusa (Crustacea:Brachyura): evidence for its evolution via gene duplication. Gene 355:28–30

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

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66:409–435

Shao R, Dowton M, Murrell A, Barker SC (2003) Rates of genome rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol Biol Evol 20:1612–1610

Shultz JW, Regier JC (2000) Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. Proc R Soc Lond B 267:1011–1010

Spears T, Abele LG (2000) Branchiopod monophyly and interordinal phylogeny inferred from 18S ribosomal DNA. J Crust Biol 20:1–24

Spelbrink JN, Toivinen JM, Hakkaart GAJ, Kurkela JM, Cooper HM, Lehtinen SK, Lecrenier N, Back JP, Speijer D, Foury F, Jacobs HT (2000) *In vivo* functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. J Biol Chem 275:24818–24828

Stanton DJ, Daehler LL, Moritz CC, Brown WM (1994) Sequences with potential to form stem-and-loop structures are associated with coding region duplications in animal mitochondrial DNA. Genetics 137:233–241

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607

Tao N, Richardson R, Bruno W, Kuiken C (2005) FindModel; http://hcv.lanl.gov/content/hcv–db/findmodel/findmodel.html

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Tracy RL, Stern DB (1995) Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr Genet 28:205–216

Urbina D, Tang B, Higgs PG (2006) The response of amino acid frequencies to directional mutational pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. J Mol Evol 62:340–361

Van Goethem G, Dermaut B, Löfgren A, Martin JJ, Van Broeckhoven C (2001) Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. Nature Genet 28:211–212

Wanrooij S, Luoma P, Van Goethem G, Van Broeckhoven C, Suomalainen A, Spelbrink JN (2004) Twinkle and POLG defects enhance age-dependent accumulation of mutations in the control region of mtDNA. Nucleic Acids Res 32:3053–3064

Wheeler WC, Whiting M, Wheeler QD, Carpenter JM (2001) The phylogeny of the extant hexapod orders. Cladistics 17:113–160 (see also erratum in Cladistics 17:403)

Wilson K, Cahill V, Ballment E, Benzie J (2000) The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: Are malacostracan crustaceans more closely related to insects than to branchiopods? Mol Biol Evol 17:863–874

Yang MY, Bowmaker M, Reyes A, Vergani L, Angeli P, Gringeri E, Jacobs HT, Holt IJ (2002) Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. Cell 111:495–505

Yang Z (2002) Phylogenetic Analysis Using Maximum Likelihood (PAML), version 314; http://abacus.gene.ucl.ac.uk/software/paml.html

Zeviani M, Spinazzola A, Carelli V (2003) Nuclear genes in mitochondrial disorders. Curr Opin Genet Dev 13:262–270