DOI: 10.1534/genetics.103.024885

Note

Assessing the Relative Rate of (Mitochondrial) Genomic Change

Mark Dowton[1]

Institutes of Molecular Biosciences and Conservation Biology, Department of Biology, Wollongong University, Wollongong, New South Wales, 2522, Australia

Manuscript received November 25, 2003
Accepted for publication February 27, 2004

ABSTRACT

I report a framework for assessing whether one mitochondrial genome is significantly more rearranged than another. This relative rate of gene rearrangement test (RGR) behaves according to expectation, distinguishing between highly rearranged and mildly rearranged insect mitochondrial genomes. It may be more broadly applied to assess the relative rate of nuclear gene rearrangement.

THE organization of the mitochondrial genome among Metazoa is remarkably uniform. Of the >300 entirely sequenced mitochondrial genomes, the vast majority contain the same 37 genes, are circular, and have very little noncoding sequence (Boore 1999). Early comparisons suggested that the relative positions of genes were similarly highly conserved, at least when comparisons within phyla were made (Moritz *et al.* 1987). However, more recent studies have exposed many exceptions to this prediction (Dowton and Austin 1999; Hickerson and Cunningham 2000; Shao *et al.* 2001a,b; Morrison *et al.* 2002; Shao and Barker 2002; Dowton *et al.* 2003). Further, rearrangements appear focused in certain taxonomic lineages [*e.g.*, within the Hymenoptera, Apocrita (Dowton and Austin 1999) and the nonhemipteran hemipteroids (Shao *et al.* 2001b), and the Crustacea, Anomura (Morrison *et al.* 2002)], leading to speculation that the rate of gene rearrangement is not uniform, but may sporadically increase or decrease during evolution.

However, there is currently no means of assessing the significance of any rearrangement rate increase or decrease—either lineages have highly rearranged genomes or they do not. It would be extremely useful to identify the branches of the metazoan mitochondrial phylogeny with significantly increased rates of rearrangement, as this would facilitate investigations of the underlying factors associated with such rate changes. Analogously, studies identifying factors associated with increased rates of nucleotide substitution—such as thermal regulation, body size, and generation time (Easteal 1985; Martin and Palumbi 1993; Rand 1994)—have been greatly facilitated by the availability of a test to compare statistically the relative rate of nucleotide substitution.

Here, I present the framework for a relative rate of gene rearrangement test (RGR). The RGR assesses whether one genome is significantly more rearranged than another. Although developed for comparisons of the mitochondrial genome, it may also prove useful for assessing the statistical significance of differences in the relative rate of nuclear gene rearrangement (*e.g.*, Coghlan and Wolfe 2002). Preliminary versions of this test have been used to compare the relative rates of mitochondrial (Shao *et al.* 2003) and nuclear (Burt *et al.* 1999) genome rearrangement, but in both cases no attempts at assessing the significance of any differences were made.

The framework for the RGR follows that developed by Tajima (1993) to assess the relative rate of nucleotide substitution between two sequences. A comparison of two sequences can reveal only the absolute number of differences between those two sequences. To determine whether one sequence is accumulating more substitutions than another, reference to a third sequence is necessary (Figure 1).

m_1 (the number of substitutions between sequence 1 and the reference sequence) and m_2 can be measured directly, and the relative rate of nucleotide divergence is estimated by subtracting these two values. The significance of this value is assessed by performing a chi-square test, with 1 d.f., *i.e.*,

$$\frac{(m_1 - m_2)^2}{m_1 + m_2}. \quad (1)$$

Adopting this framework for the RGR, the relative number of gene rearrangements between two genomes can

[1]*Address for correspondence:* Department of Biology, Northfields Ave., Wollongong University, Wollongong, NSW 2522, Australia. E-mail: mdowton@uow.edu.au

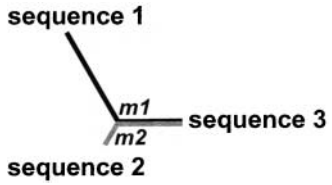

FIGURE 1.—Unrooted phylogram depicting sequence divergence between three sequences. m_1 is the sequence divergence between sequences 1 and 3 and m_2 the sequence divergence between sequences 2 and 3. The relative sequence divergence between sequences 1 and 2 is thus $m_1 - m_2$.

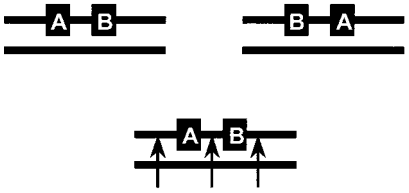

FIGURE 2.—Example calculation of the breakpoint distance. In the simplest case, above, two genomes differ by the relative positions of just two genes, with all other genes in identical positions (not shown). The minimum number of breakpoints that must be introduced into the left genome to convert it into the right genome is three: two breakpoints either side of gene B (to excise B), with a third breakpoint upstream of gene A to permit B to be reinserted back into the genome at this position. By comparison, the number of gene rearrangements calculated in this scenario is one. An example calculation of a more complex gene rearrangement can be found in BLANCHETTE *et al.* (1999).

be estimated by reference to a third genome. WATTERSON *et al.* (1982) and SANKOFF *et al.* (1992) have developed algorithms for assessing the number of gene rearrangements between genome pairs. WATTERSON *et al.* (1982) introduced the notion of a breakpoint distance between genome pairs, defined as the minimum number of breakpoints that must be introduced into one genome to convert it to another, thus applying the parsimony principle. An example of how this number is estimated is shown in Figure 2. In this test, I use breakpoint distances as a means of quantifying the number of gene rearrangements between pairs of genomes. Conveniently, software has been developed to facilitate this calculation (MORET *et al.* 2002).

The RGR can be expressed in a manner analogous to Tajima's relative rate test,

$$\frac{[BP_{(1,3)} - BP_{(2,3)}]^2}{BP_{(1,3)} + BP_{(2,3)}}, \qquad (2)$$

where $BP_{(1,3)}$ refers to the breakpoint distance between genomes 1 and 3, and $BP_{(2,3)}$ refers to the breakpoint distance between genomes 2 and 3. This expression measures the relative rate of gene rearrangement between genomes 1 and 2, with genome 3 the reference genome. As with Tajima's relative rate test, the significance of the calculated value is assessed by reference to chi-square tables, using 1 d.f.. Where multiple tests are performed, Bonferroni corrections are applied. For example, if 10 RGRs are performed, a significant difference is found only where $P < 0.005$ (*i.e.*, $0.05 \div 10$ tests).

To investigate the behavior of the test, I applied the RGR to a range of insect mitochondrial genomes. The insect mitochondrial genome was chosen as it displays a number of properties useful for assessing the behavior of the RGR: (i) the organization of the ancestral insect mitochondrial genome can be straightforwardly inferred, as some insect mitochondrial genomes are organized identically to those of some crustaceans; (ii) the number of gene rearrangements in most insect mitochondrial genomes can be calculated manually; and (iii) insect mitochondrial genomes display a range of organizations, from retention of the ancestral organization to highly rearranged. In this way we can examine whether the test behaves according to expectation. Figure 3 illustrates this expectation. The left side is a phylogeny of the hexapods (from WHEELER *et al.* 2001), while the right side shows the same phylogeny, drawn with branch lengths proportional to the number of gene rearrangements (as measured by the breakpoint distance) that have occurred along each branch. This suggests that Thrips, the lepidopsocid, Heterodoxus, and Apis have increased rates of gene rearrangement compared with other insects.

Table 1 shows the results of RGRs among the various insect mitochondrial genomes. In each case Daphnia (Crustacea) was used as the reference mitochondrial genome (*i.e.*, genome 3), with Drosophila as genome 2. These tests (Table 1, comparison A) indicate that the rate of mitochondrial gene rearrangement is insignificantly different among a number of insect orders, *viz.* the Diptera, Coleoptera, Orthoptera, Hemiptera, and Lepidoptera (*i.e.*, those lineages with relatively short branches in Figure 3), but has significantly increased in the lepidopsocid sp. (Psocoptera), Thrips (Thysanoptera), Heterodoxus (Phthiraptera), and Apis (Hymenoptera; *i.e.*, those lineages with relatively long branches in Figure 3). Thus, the RGR behaves according to expectation.

I used breakpoint distance as an estimate of the number of gene rearrangements, both to automate the calculation and to make it feasible to extend the test to compare highly rearranged mitochondrial and nuclear genomes. However, the use of breakpoint distances may inflate significance when making comparisons between relatively unrearranged genomes. This is due to the breakpoint distance being much greater than the number of gene rearrangements when the degree of genome reorganization is low. For example, a single rearrangement generally produces a breakpoint distance of 3 (see Figure 2), while two rearrangements generally produce a breakpoint distance of 5. If genomes 2 and 3 are identical [such that $BP_{(2,3)} = 0$], genomes 1 and 3 need

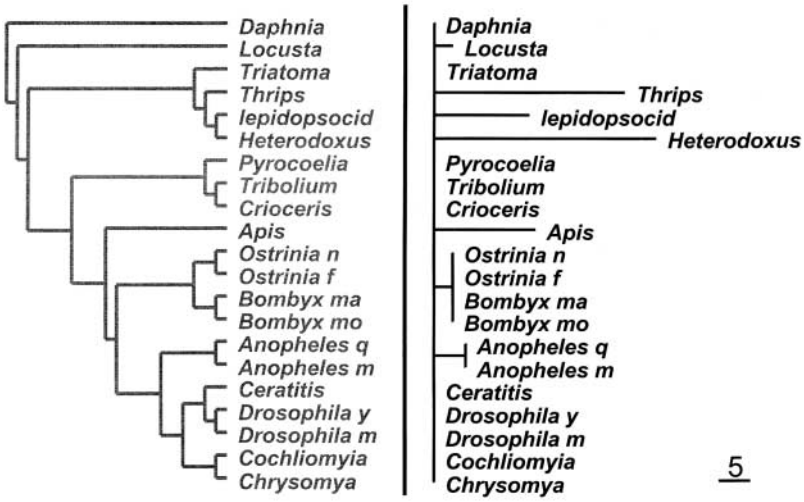

FIGURE 3.—Left, phylogeny of the insects (redrawn from WHEELER *et al.* 2001). Right, the same phylogeny, with branch lengths drawn proportional to the number of breakpoints. Where appropriate, breakpoint distances were assigned to internal branches (where the genome organization of related terminal taxa was identical but different from the ancestral organization), *i.e.*, the branch leading to the four lepidopterans (two Ostrinia and two Bombyx) and the two Anopheles species. In all other cases, breakpoint distances were assigned to terminal branches. This is reasonable, as no deeper-level, synaptomorphic gene rearrangements are evident among these taxa. The bar shows the branch length for a breakpoint distance of 5.

only differ by two rearrangements [$BP_{(1,3)} = 5$] to produce a significant ($P = 0.0253$) RGR. Thus, a comparison of the mitochondrial genomes of Drosophila and Anopheles, using Daphnia as the reference genome, would indicate that Anopheles is significantly more rearranged than Drosophila, even though these two genomes differ by just two gene rearrangements (the reason this comparison produces an insignificant RGR in Table 1, comparison A, is the Bonferroni correction). In such cases, using the number of gene rearrangements rather than the breakpoint distance appears more reasonable. In this case (Table 1, comparison B), all four taxa with highly rearranged genomes are still identified as significantly rearranged, but the level of significance is reduced for Apis and the lepidopsocid from <0.001 to <0.05, after Bonferroni correction.

I expect that both the advantages and shortcomings of Tajima's relative rate test also apply to the RGR. For example, the choice of reference genome may influence the test. Tajima's relative rate test is most powerful when the branch length to the reference sequence is short (ROBINSON *et al.* 1998), and I purposely chose Daphnia and Drosophila with this in mind (they have identical genome organizations, making the branch length to the reference taxon indeed short). Artificially increasing the branch lengths to the reference mitochondrial genomes makes the RGR less able to detect significant increases in the rate of gene rearrangement (Table 1, comparison C; only Heterodoxus and Thrips have significantly increased rates of rearrangement). This is not a criticism of the test, but an indication that it is less able to detect differences when the data are less than ideal.

Indeed, the choice of a close reference genome may

TABLE 1

Assessment of the relative rate of genome rearrangement among insect mitochondrial genomes, using the RGR

	Comparison A		Comparison B		Comparison C	
Genome 1	$BP_{(1,3)}$, $BP_{(2,3)}$	*P*	$GR_{(1,3)}$, $GR_{(2,3)}$	*P*	$GR_{(1,3)}$, $GR_{(2,3)}$	*P*
Orthoptera: Locusta	3, 0	0.0833	1, 0	0.3173	1, 2	0.5673
Hemiptera: Triatoma	0, 0	1.0000	0, 0	1.0000	0, 2	0.1573
Thysanoptera: Thrips	30, 0	4×10^{-8}***	24, 0	9×10^{-7}***	24, 2	1×10^{-4}***
Psocoptera: Lepidopsocid sp.	15, 0	1×10^{-4}***	8, 0	0.0047*	8, 2	0.0578
Phthiraptera: Heterodoxus	35, 0	3×10^{-9}***	31, 0	3×10^{-8}***	31, 2	4×10^{-7}***
Coleoptera	0, 0	1.0000	0, 0	1.0000	0, 2	0.1573
Hymenoptera: Apis	16, 0	1×10^{-4}***	8, 0	0.0047*	8, 2	0.0578
Lepidoptera	3, 0	0.0833	1, 0	0.3173	1, 2	0.5673
Diptera: Anopheles	5, 0	0.0253	2, 0	0.1573	2, 2	1.0000

For all comparisons, genomes 2 and 3 are Drosophila and Daphnia, respectively. However, in comparison C, the number of gene rearrangements between genomes 2 and 3 has been artificially inflated (from 0 to 2), to assess the impact of increasing branch lengths to the reference genomes. The tabulated *P* values are the uncorrected values; asterisks indicate the level of significance after Bonferroni corrections (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). Where genome organization did not differ among members of an order [*viz.* Coleoptera (Crioceris, Pyrocoelia, and Tribolium) and Lepidoptera (Bombyx and Ostrinia)], only a single genome was assessed. Only Anopheles from the Diptera was assessed, as the genome organization of other dipterans (Cochliomyia, Chrysomya, and Ceratitis) does not differ from that of Drosophila.

make the test more statistically valid. The relative rate of nucleotide divergence test attempts to avoid correlation due to shared phylogenetic history (Figure 1, the branch leading from sequence 3 to the internode)—by subtracting m_2 from m_1, the common branch is removed from the analysis. However, Felsenstein (1988) pointed out that changes along this branch may impact on the estimation of m_1 and m_2 simultaneously, with the values becoming correlated. I expect that this is a minor issue with mitochondrial genome organizational data, due to the smaller number of changes assessed (compared with nucleotide divergences), but may become more important with highly rearranged nuclear genome comparisons.

The RGR should thus be useful for distinguishing between stochastic and statistically significant fluctuations in the rate of genome reorganization. We used a preliminary version of this test to assess the degree of correlation between the rate of genomic change and that of nucleotide substitution (Shao *et al.* 2003). We found that, among insects, an increased rate of nucleotide divergence coincided significantly with an increased rate of rearrangement of the mitochondrial genome.

This study demonstrates the utility of the RGR when assessing the relative rate of gene rearrangement between two mitochondrial genomes. Recent analytical advances suggest that it might also be applied straightforwardly to assess the relative rate of gene rearrangement between two nuclear genomes. I used GRAPPA (Moret *et al.* 2002) to calculate breakpoint distances between single chromosomal (*i.e.*, mitochondrial) genomes. Tesler (2002) recently developed a framework for measuring breakpoint distances between multichromosomal genomes. Further, identification of every gene (conceptually, the limiting factor in making between-nuclear genome comparisons) is not necessary—identification of conserved, alignable blocks appears to be a reasonable approximation (Kent *et al.* 2003). Thus, it should be straightforward to assess the relative rate of rearrangement between two nuclear genomes, using the RGR.

LITERATURE CITED

Blanchette, M., T. Kunisawa and D. Sankoff, 1999 Gene order breakpoint evidence in animal mitochondrial phylogeny. J. Mol. Evol **49:** 193–203.

Boore, J. L., 1999 Animal mitochondrial genomes. Nucleic Acids Res. **27:** 1767–1780.

Burt, D. W., C. Bruley, I. C. Dunn, C. T. Jones, A. Ramage *et al.*, 1999 The dynamics of chromosome evolution in birds and mammals. Nature **402:** 411–413.

Coghlan, A., and K. H. Wolfe, 2002 Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. Genome Res. **12:** 857–867.

Dowton, M., and A. D. Austin, 1999 Evolutionary dynamics of a mitochondrial rearrangement "hotspot" in the Hymenoptera. Mol. Biol. Evol. **16:** 298–309.

Dowton, M., L. R. Castro, S. L. Campbell, S. D. Bargon and A. D. Austin, 2003 Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. J. Mol. Evol. **56:** 517–526.

Easteal, S., 1985 Generation time and the rate of molecular evolution. Mol. Biol. Evol. **2:** 450–453.

Felsenstein, J., 1988 Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22:** 521–565.

Hickerson, M. J., and C. W. Cunningham, 2000 Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura). Mol. Biol. Evol. **17:** 639–644.

Kent, W. J., R. Baertsch, A. Hinrichs, W. Miller and D. Haussler, 2003 Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genome. Proc. Natl. Acad. Sci. USA **100:** 11484–11489.

Martin, A. P., and S. R. Palumbi, 1993 Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90:** 4087–4091.

Moret, B. M. E., J. Tang, L.-S. Wang and T. Warnow, 2002 Steps toward accurate reconstructions of phylogenies from gene-order data. J. Comput. Syst. Sci. **65:** 508–525.

Moritz, C., T. E. Dowling and W. M. Brown, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18:** 269–292.

Morrison, C. L., A. W. Harvey, S. Lavery, K. Tieu, Y. Huang *et al.*, 2002 Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. Proc. R. Soc. Lond. Ser. B **269:** 345–350.

Rand, D. M., 1994 Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends Ecol. Evol. **9:** 125–131.

Robinson, M., M. Gouy, C. Gautier and D. Mouchiroud, 1998 Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. **15:** 1091–1098.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B. F. Lang *et al.*, 1992 Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA **89:** 6575–6579.

Shao, R., and S. C. Barker, 2002 The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol. Biol. Evol. **20:** 362–370.

Shao, R., N. J. H. Campbell and S. C. Barker, 2001a Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol. Biol. Evol. **18:** 858–865.

Shao, R., N. J. H. Campbell, E. R. Schmidt and S. C. Barker, 2001b Increased rate of gene rearrangement in the mitochondrial genomes of insects in three hemipteroid orders. Mol. Biol. Evol. **18:** 1828–1832.

Shao, R., M. Dowton, A. Murrell and S. C. Barker, 2003 The rates of gene rearrangement and nucleotide substitution are positively correlated in the mitochondrial genomes of insects. Mol. Biol. Evol. **20:** 1612–1619.

Tajima, F., 1993 Simple methods for testing the molecular clock hypothesis. Genetics **135:** 599–607.

Tesler, G., 2002 GRIMM: genome rearrangements web server. Bioinformatics **18:** 492–493.

Watterson, W. A., W. J. Ewens, T. E. Hall and A. Morgan, 1982 The chromosome inversion problem. J. Theor. Biol. **99:** 1–7.

Wheeler, W. C., M. Whiting, Q. D. Wheeler and J. M. Carpenter, 2001 The phylogeny of the extant hexapod orders. Cladistics **17:** 113–169.

Communicating editor: J. Tamkun