EFFICIENCIES OF GENES AND ACCURACY OF TREE-BUILDING METHODS IN RECOVERING A KNOWN *DROSOPHILA* GENEALOGY

J.E. STEINBACHS[1], N.V. SCHIZAS[1,2], and J.W.O. BALLARD[1]
[1] *The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL, 60605-2496, USA*
[2] *University of Chicago, Department of Ecology and Evolution, 1101 E. 57th Street, Chicago, IL, 60637, USA*

Phylogenetic hypotheses generated from seven *Drosophila* mitochondrial genomes support a well-corroborated genealogy with a single evolutionary history. These mitochondrial data form a model system for investigating the efficiency of genes and accuracy of different tree-building methods in recovering a well-supported genealogy. We consider 15 genes (13 protein-coding and 2 rRNAs) and 83 tree-building methods (27 distance, 4 parsimony, 50 maximum likelihood, and 2 Bayesian). Among the 15 genes examined, ND4 recovered the true genealogy most efficiently (82 out of 83 methods). Generally, maximum likelihood models enforcing a clock most accurately reclaim the true genealogy. Surprisingly, however, this method fails to recover the well-supported topology for more than half of the genes. Additional studies are required to test the generality of the results presented here.

1 Introduction

In this study, we compare the robustness of distinct genes and of different tree-building methods in resolving a well-corroborated genealogy. We also aim to provide a set of standard methods for analyzing whole genomes, and to provide a data set for benchmarking new methods or methods we do not examine. We do not attempt to optimize the search algorithm for each gene in this study. However, we acknowledge that systematists have different philosophical approaches to analyzing data and constructing topologies. Here, we treat all methods equally without taking into consideration the statistical and philosophical differences that these methods imply (see Swofford *et al.* [1] for a comprehensive treatment).

New phylogenetic methods and models may be first examined through simulation analysis in the four taxon case[2,3], followed by tests with simulated data for multiple taxa[4]. These methods should then be tested with experimentally derived data sets from closely related taxa so as to minimize the potentially confounding problem of multiple substitutions on any given branch. Ideally, there should be no recombination among these sequences so that a single tree topology accurately depicts the genealogy of linked genes. In this study, we use the complete mitochondrial genomes of seven closely-related *Drosophila*[5,6]

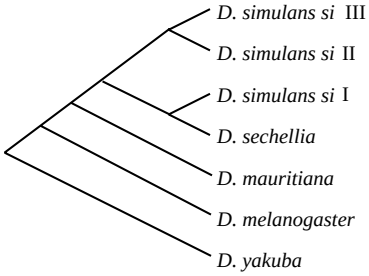

Figure 1: The whole mitochondrial genome data set recovered this topology regardless of the method/model. MP with equal weighting supports each node with greater than 95% of bootstrap pseudoreplicates. Thus, we assume that this is the known genealogy.

to examine the accuracy of a number of different methods. We assume this is the true/known mitochondrial genealogy but we do not claim this reflects the phylogenetic relations of these taxa.

The *Drosophila* mitochondrial genome contains 22 tRNAs, 13 protein-coding genes and two rRNAs, with whole genome nucleotide divergence ranging from 1.12% to 6.43%. There is no evidence of recombination and the mitochondrial genome is inherited as a single completely linked molecule [5]. These features simplify our analysis, providing us with a single genealogy for the molecule.

In this study, we investigate the reliability of all protein-coding genes, both rRNAs, and 83 different phylogenetic methods in recovering the mitochondrial genealogy of this group of organisms (Fig. 1). All methods and models included in this study recovered the same topology when the complete data set was employed. However, distinct selective pressures are acting on the mitochondrial and autosomal genes of *D. simulans* [7], so we do not have an independent method of corroborating this genealogy.

2 Materials and Methods

2.1 Organisms and Sequence Data

We used the complete mitochondrial genomes from seven isofemale lines of *Drosophila*. The *D. melanogaster* subgroup consists of six major mitochondrial lineages, of which we used *D. melanogaster* Oregon R (AF200828), *D. sechellia* (AF200832), *D. mauritiana ma*II (AF200830), and *D. simulans si*I, -II, and -III (AF200834, AF200841, AF200852, respectively) [5]. We included a representative of each *D. simulans* haplotype because they are not mono-

phyletic with regard to *D. sechellia* [8]. We did not include the *D. mauritiana ma*I haplotype as it differs by a single substitution from *D. simulans si*III and is likely the result of introgression [8]. *D. yakuba* (NC001322) is the outgroup of the *D. melanogaster* subgroup [6].

We followed the alignment of Ballard [5], deleting 76bp from the 15,034 bp sequence. The region between 5,535-5,584 forms the majority of an intervening sequence between COIII and the glycine tRNA. The region between positions 6,022-6,047 forms an intervening sequence between ND3 and the alanine tRNA [7]. We imported the aligned sequences into DnaSP 3.14 [9] to define the protein-coding regions, rRNAs and intervening sequences.

For each protein-coding gene we exclude start and stop codons. Furthermore, the Atp8 gene is shortened by one codon so that it does not overlap with Atp6; this codon is invariant in all taxa included in this study. The *D. melanogaster* ND5 locus has a single amino acid deletion. An additional 41 insertion/deletion events were parsimoniously scored to minimize the number of events, placing this information into a stepmatrix at the end of the data set (characters 15036-15078).

2.2 Phylogenetic Methods

We examined 83 nucleotide substitution models. We tested UPGMA, neighbor-joining [10](NJ), minimum evolution [4] (ME), with negative branches allowed for the ME tests, using nine basic distance models: p distances, Jukes-Cantor [11] (JC) and JC $+\alpha$, the gamma shape parameter (denoted hereafter as g); Kimura's two-parameter [12] (K2P) and K2P+g; HKY [13] and HKY+g; Generalized Time Reversible [14, 15] (GTR) and GTR+g. For these analyses, we used the method of Gu and Zhang [16], as implemented in GAMMA, to estimate the gamma shape parameter.

We examined four parsimony (MP) methods: equally weighted, and three non-equal weighting schemes (transition/transversion ratios of 1:2, 1:4 and 1:8).

We tested 25 maximum likelihood (ML) models with and without enforcing the molecular clock (for a total of 50 ML models). We initially tested five basic ML models [JC, F81 [17], K2P, HKY, and GTR] with equal rates. We then tested each of these models with gamma-distributed variable sites (g), with the proportion of variable sites estimated by likelihood (I), both gamma shape and the proportion of variable sites estimated (g+I), and site-specific rates for protein-coding loci (g+I+codons). Each codon position was allowed to have a different substitution rate for models with site-specific rates. When the clock was not enforced, we estimated the gamma shape parameter following a NJ

Log-Det search.

We implemented all distance, parsimony, and ML methods in a series of scripts for use with PAUP* [18]. For ME, MP and ML, we used an exhaustive search method to explore tree space. When more than one equally parsimonious tree was found, a strict consensus tree was constructed. These scripts and the complete aligned data set are available at http://cb.fmnh.org/.

Because of recent developments in the application of the Bayesian approach to phylogenetic inference [19], we included this type of analysis for comparison to more traditional methods. We used the most general Bayesian tree estimation program available, BAMBE 2.02b [19], following the procedure on the BAMBE web site [20] for parameter tuning. Inference runs consisted of 5,000 burn-in cycles, with 500,000 inference cycles, using the HKY model enforcing a clock (UPGMA search for the initial tree). We repeated the analysis without enforcing the clock assumption, using NJ to determine the initial tree. On the resulting genealogies, we collapsed nodes with a posterior probability less than 50%, then used the collapsed topology for subsequent analyses.

2.3 Accuracy of the Topology

To evaluate the accuracy of the topology obtained from each gene, we calculated the topological distance between each genealogy to that derived from the complete mitochondrial genome. This distance (dT) is based on the work by Robinson and Foulds [21] and Penny and Hendy [22]. Another possible tree comparison method uses the agreement metric [23].

dT reflects the number of internal branches present in one tree but not in the other. In our data set, dT ranges from 0 to 8, with 0 reflecting identical topologies. The maximum score of dT is 8 because the known topology contains four internal branches. While we can calculate dT easily, we must exercise caution interpreting the results of those calculations: dT only indicates *how many* branches differ between the topologies in the comparison, not *which* branches differ.

To facilitate the comparison of methods, we compare unrooted trees. This comparison does not consider the position of the root. Thus, a method may generate a topologically correct tree but place the root on an incorrect branch. When a phylogenetic method produced multiple topologies, we compared the consensus topology to the known genealogy; we could instead have conducted a pairwise comparison, and then taken the average of these multiple dT calculations.

To determine the most accurate method, we devised a ranking scheme based on the topological distance measure. For each method, we counted

the number of genes for which a given model returned the correct topology (*nc*); this value monotonically increases with increasing levels of accuracy. For each model, we summed together all the dT values for all genes (sum of dT); this sum monotonically decreases with increasing levels of accuracy. For each method, then, the best model will be one that returns the correct topology most often (highest *nc*) and, when it does not, it recovers topologies that are at least close to the known genealogy (small sum of dT). For models that have the same value of nc and sum of dT, the simpler model with fewer parameters is preferable. Note that there is more than one way to arrive at a particular combination of nc and sum of dT, so this ranking method is not perfect.

To rank the genes from most efficient ("good") to least efficient ("poor"), we followed a similar strategy. For each gene, we counted the number of methods/models for which the known topology was successfully recovered (*nc-gene*). "Good" genes will have a higher *nc-gene* than "poor" ones. For each gene, we also summed the dT values across all methods/models to determine the sum of dT for each gene. As with the methods/models, this sum will be lower for "good" genes than for "poor" ones. From an empirical point of view, one might want to minimize the sequencing effort, so we then ranked according to gene length. Combining all criteria for genes, "good" genes will be characterized by a high *nc-gene*, a low sum of dT across all methods/models, and small gene length.

3 Results

We rank genes from most efficient to least efficient (Fig. 2). ND4 outperformed all other genes, recovering the correct topology for almost 99% of methods/models. Equally weighted parsimony was the only case for which ND4 failed to recover the correct topology. For more than 80% of methods/models, ND3 and lrRNA recovered the known genealogy. Atp6, Atp8, cytochrome *b* and ND6 never recovered the correct topology. Two genes (COI, ND5) out of the three longest (see Table 1 for lengths) performed poorly.

In phylogenetic studies, longer genes are expected to perform better than shorter genes[24, 25, 26]. In our study, there lacks a strong correlation between the frequency with which a particular gene recovers the known genealogy (% *nc-gene*) and gene length (Fig. 3). Based on previous studies [24, 25], we were surprised by the performance of three genes: ND5, ND3 and cytochrome *b*. Generally, ND5 and cytochrome *b* are expected to perform reasonably well as they are the longer genes in the mitochondrial genome. In contrast, ND3, a short gene, has not been identified as a "good" gene.

Table 1 shows the topological distances for the most accurate models for

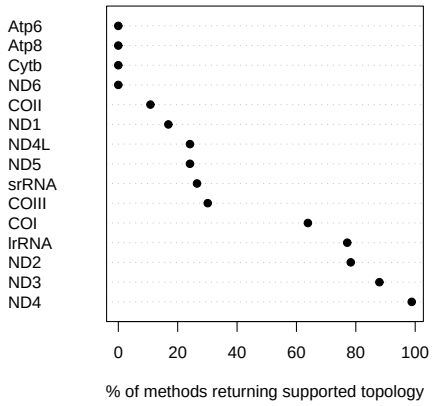

Figure 2: Efficiency of genes in recovering the well-supported topology. Cytb: cytochrome b.

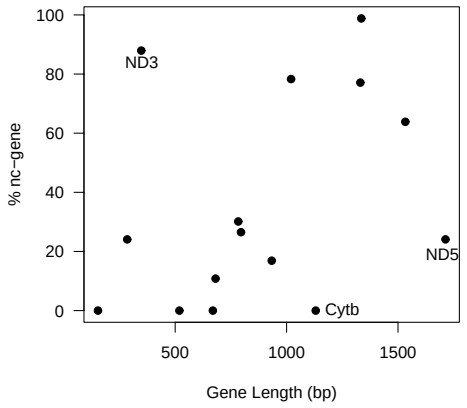

Figure 3: Relationship between % *nc-gene* and gene length. With all genes, $r^2 = 0.13$. When we exclude ND5, ND3 and cytochrome b, $r^2 = 0.65$. We detected a similar pattern between %*nc-gene* and the number of variable sites per gene (data not shown). Cytb: cytochrome b

each method. The *nc* values range from 5 to 7, indicating that no method/model combination recovered the known topology combination recovered the known topology for more than 50% of the genes. The sum of *d*T range from 18 to 32. ML and Bayesian models with clock enforced outperform models without clock. Of these highest scoring methods, ML with clock (HKY+I) had the smallest sum of *d*T, making it the best method overall. ME and NJ performed better than UPGMA. Furthermore, in ME and NJ methods, models with the least number of parameters (*e.g., p*) performed as well as models with the greatest number of parameters (*e.g.,* GTR). Unequally weighted MP performed better than MP with equal weights.

4 Discussion

One of our primary goals was to investigate the phylogenetic efficiency of mitochondrial genes in a closely-related group of species, and to compare our results with other studies conducted with distantly related taxa[24, 25, 26]. The generality of the results presented here needs to be tested in other species assemblages.

Overall, ND4 performed well in this study and in previous studies investigating the performance of mitochondrial protein-coding genes in vertebrate taxa[24, 25]. On the contrary, ND5 and cytochrome *b* were not accurate in our study, but recovered the hypothesized vertebrate tree[24, 25]. All previous studies that considered the efficiency of mitochondrial genes suggest that the length of the molecular marker is crucial to ensure that the correct topology is recovered. In our study, gene length is not an important factor in recovering the known topology. In fact, ND5, our longest gene, recovered the correct topology for fewer than 30% of the methods/models while ND3, our third shortest gene, was over 80% efficient. Though gene length is not a significant factor in our data set (Fig. 3), it is clear that sequence length is important because the whole genome data consistently recovers the well-supported topology.

Given its length and its performance in past studies[24, 25], we expected ND5 to recover the known genealogy. This was not the case. For this gene only, distance methods consistently outperform ML and MP. Furthermore, ND5 recovers the known topology only for distance methods without the gamma shape parameter (see figs. 2 and 3). Two regions of this gene have some degree of rate heterogeneity[5], a possible cause of its inconsistent performance. It is also possible that these regions of heterogeneity correspond to unique structural features that play a major role in protein function. Recent work suggests that protein structure may be an important feature to consider in new methods of phylogenetic analysis[27, 28, 29].

Table 1: Topological distances from the highest ranking models for each method.

Method	UPGMA		MP		ME	NJ	ML NC	Bay C	ML C		Sum dT	Gene Len.
Wt. or Model	p^a	JC^b	$\frac{1}{4}^c$	$\frac{1}{8}^c$	p^d	p^e	JC^f	HKY	HKY + I	*nc-gene*	(gene)	(bp)
Genes												
ND4	0	0	0	0	0	0	0	0	0	82	1	1335
ND3	0	0	0	0	0	0	0	0	0	73	16	348
ND2	4	4	0	0	0	0	0	0	0	65	54	1020
lrRNA	2	2	0	0	0	0	0	0	0	64	18	1331
COI	0	0	0	0	2	2	0	0	0	53	72	1533
COIII	2	2	2	4	2	2	4	1	0	25	173	783
srRNA	2	2	2	2	0	0	2	0	2	22	104	795
ND5	0	0	3	3	0	0	2	0	2	20	160	1713
ND4l	4	6	1	1	2	2	0	1	0	20	203	285
ND1	4	4	2	2	2	2	0	2	2	14	228	933
COII	0	0	4	4	4	4	4	3	1	9	285	681
Cytb	2	2	4	4	2	2	4	2	2	0	264	1131
Atp8	6	4	3	3	4	4	3	4	4	0	361	153
ND6	4	4	3	1	4	4	6	3	3	0	381	519
Atp6	2	2	4	4	4	4	6	3	2	0	395	669
Sum dT (method)	32	32	28	28	26	26	31	19	18			
nc	5	5	5	5	6	6	7	7	7			

The maximum value of *nc* is 15; the maximum value of *nc-gene* is 83. Abbreviations: Wt. - Weighting Scheme; Bay. - Bayesian; C - with clock; NC - without clock; Len. - Length. Superscripts: ^{a}UPGMA: identical results with HKY, GTR; ^{b}UPGMA: also with JC+g; ^{c}Ratio of transitions to transversions; ^{d}ME: also JC, K2P, HKY, GTR; ^{e}NJ: also K2P, HKY, GTR; ^{f}ML (NC): also JC+g, JC+I. *nc-gene* is calculated from all 83 methods/models. Genes ranked from most efficient (top) to least efficient (bottom). Methods/models ranked from least accurate (left) to most accurate (right). Note: See methods section for details on ranking.

Cytochrome *b* also performed poorly (Figs. 2 and 3). It is possible that distinct regions of the gene may be subject to high levels of homoplasy. Mutational and evolutionary studies of the cytochrome *b* gene have facilitated the development of a structure/function model and the identification of the sites of electron transfer and inhibitor action [30, 31, 32]. Ballard and Kreitman [33] considered the synonymous and nonsynonymous changes within and between *Drosophila* species among five internal, eight transmembrane and four external regions. They hypothesized that changes should be randomly distributed if there are no selective constraints, or if constraints are the same for all regions. In that study there was no evidence for heterogeneity of nonsynonymous or synonymous changes among regions or species. In our study there is also no evidence for distinct patterns of substitution in the three regions of cytochrome *b* ($G_2^2 = 0.35$, $p = 0.91$; 9 homoplastic changes and 6 non-homoplastic parsimony informative positions in the internal region, 12/6 in the transmembrane region, and 11/7 in the external region). Future work may include a tree-independent analysis on the phylogenetic signal in this gene [34].

Short genes generally are not expected to perform well in phylogenetic studies because they have a limited amount of phylogenetic signal [25]. Surprisingly, the third shortest gene in our dataset (ND3) recovered the correct topology for over 80% of the methods/models. At this moment, we cannot distinguish if this result arises from stochastic effects, from functional constraints, or some other undetermined process.

We also wanted to compare the accuracy of different tree-building algorithms in recovering the mitochondrial genealogy of this group of organisms. The various phylogenetic methods performed very differently in their ability to recover the known genealogy. ML with a clock enforced (HKY+I) was the best overall method/model. However, even this method fails to recover the well-supported topology for more than half of the genes. This result suggests that there is something inherently different about the evolutionary constraints on these *Drosophila* genes that cannot be explained or modeled by phylogenetic methods/models used in this study.

ML models with few parameters (*e.g.,* *p*, JC) performed as well as more complex models (*e.g.,* GTR). It is possible, however, that the simpler models are wrong because, as Yang [37] points out, simple (but wrong) models can recover the topology generated by a more complex model. In some cases, however, more complex models, particularly those using site-specific rates (*i.e.,* g+I+codons), performed poorly in comparison to other variations of the models (*e.g.,* gamma shape parameter,g; proportion of variable sites, I; and both g+I). Ideally, one should attempt to objectively identify the best model for a given data set. We are currently investigating whether the "best" model,

as determined by the log likelihood ratio test [36] and the Akaike Information Criterion (AIC) [38], recovers the known genealogy.

For MP, we found that unequal weighting schemes are necessary to recover the correct topoogy. However, we found only few differences among the different weighting schemes, suggesting that any one of these unequal weighting schemes might be appropriate. In studies with simulated data, Hillis *et al.* [3] and Nei *et al.* [39] found that unequally weighted MP performed much better than equally weighted MP. However, with their vertebrate mitochondrial sequence data, Russo *et al.* [24] found little difference between unequally and equally weighted MP. When this algorithm and these weighting schemes are applied to experimental data, the difference in the ability to recover the known topology does not appear to be large.

We expected comparable performance between ME and NJ [4]; in fact, we found that the best models for ME and NJ resulted in identical topological distances. As in the Russo *et al.* [24] study, these models fail to improve when we include the gamma shape parameter. In particular GTR+g was the worst model for all distance methods, suggesting that overparameterization for these methods will lead to an incorrect topology. We were not surprised to observe that UPGMA performed poorly [2, 3].

4.1 Future Directions

These results form the preliminary stage of this investigation. We intend to use Monte Carlo simulations to further investigate the surprising performances of ND3, ND5, and cytochrome *b*. If the ability of genes to recover the known phylogeny differs by chance, then we may be detecting a gene-by-organism interaction. By using closely related organisms in other groups, we may be able to determine if this interaction is restricted to the *D. melanogaster* subgroup. Alternatively, if the performance of these genes do not differ by chance, then the differences likely arise because of stochastic effects.

As more *Drosophila* mitochondrial genomes become sequenced, we intend to include new taxa so that our data set might include both closely- and distantly-related species. Since some genes have been demonstrated to be appropriate for distantly-related taxa (*e.g.,* ND5 [24]), but not for these closely-related taxa, we would like to see if these trends hold in an expanded data set within *Drosophila*.

While the nodes in the known genealogy are well-supported at the 95% confidence level using equally weighted MP analysis, we have yet to address the robustness of the internal nodes for all methods under consideration. As a first attempt to assess nodal support, we will apply Felsenstein's traditional non-

parametric bootstrap method [40] to all methods/models in our study. Given that we are confident that the known genealogy is correct, it is also appropriate to use the computationally intensive parametric bootstrap approach [41]. We expect that, for many of the methods used here, even the best topologies should show less than 100% bootstrap support on internal nodes.

Acknowledgments

Special thanks to Dave Swofford for his assistance in writing the scripts for our PAUP* analysis and for his phylogenetic insights. We received helpful comments from Shannon Hackett, Paul Goldstein, Jason Bond, and an anonymous reviewer. This work was funded by NSERC (to JES, PDF-220441-1999), a DOE/Sloan Fellowship (to NVS), and NSF (to JWOB, DEB-9702824). Funding for computer equipment came from NSF (DBI-9871372) and from Sun Microsystems (EDUD-7826-990416-US).

References

1. D.L. Swofford, G.J. Olsen, P.J. Waddell, and D.M. Hillis in *Molecular Systematics*, Eds. D.M. Hillis, C. Moritz, and B.K. Mable, pp. 407–514 (Sinauer Associates, Sunderland, MA, 1996)
2. J.P. Huelsenbeck and D.M. Hillis, *Syst. Biol.* **42**: 247–264 (1993)
3. D.M. Hillis, J.P. Huelsenbeck, and C.W. Cunningham, *Science* **264**: 671–677 (1994)
4. A. Rzhetsky and M. Nei, *Mol. Biol. Evol.* **9**: 945–967 (1992)
5. J.W.O. Ballard, *J. Mol. Evol.* **51**: 48–63 (2000)
6. D.O. Clary and D.R. Wolstenholme, *J. Mol. Evol.* **22**: 252–271 (1985)
7. J.W.O. Ballard, *J. Mol. Evol.* **51**: 64–75 (2000)
8. J.W.O. Ballard, *Mol. Biol. Evol.* **17**: 1126–1130 (2000)
9. J. Rosas and R. Rosas, *Comput. Applic. Biosci.* **13**: 307–311 (1997)
10. N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**: 406–425 (1987)
11. T.H. Jukes and C.R. Cantor in *Mammalian Protein Metabolism*, Ed. H.M. Munro, pp. 21–132 (Academic Press, New York, NY, 1969)
12. M. Kimura, *J. Mol. Evol.* **16**: 111–120 (1980)
13. M. Hasegawa, K. Kishino, and T. Yano, *J. Mol. Evol.* **21**: 160–174 (1985)
14. C. Lanave, G. Preparata, C. Saccone, and G. Serio, *J. Mol. Evol.* **20**: 86–93 (1984)
15. S. Tavaré, *Math. Life Sci.* **17**: 57–86 (1986)
16. X. Gu and J.J. Zhang, *Mol. Biol. Evol.* **14**: 1106–1113 (1997)

17. J. Felsenstein, *J. Mol. Evol.* **17**: 368–376 (1981)
18. D.L. Swofford, *PAUP* Phylogenetic analysis using parsimony and other methods* (Sinauer Associates, Sunderland, MA, 1998)
19. D. Simon and B. Larget, *Bayesian analysis in molecular biology and evolution (BAMBE), version 2.02 beta* (Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA, 2000)
20. http://www.mathcs.duq.edu/larget/bambe/Manual/manual.html
21. D.F. Robinson and L.R. Foulds, *Math. Biosci.* **53**: 131–147 (1981)
22. D. Penny and M.D. Hendy, *Syst. Zool.* **34**: 75–82 (1985)
23. W.E. Goddard, E. Kubicka, G. Kubicki, and F.R. McMorris, *Math. Biosci.* **123**: 215–226 (1994)
24. A.M. Russo, N. Takezaki, and M. Nei, *Mol. Biol. Evol.* **13**: 525–536 (1996)
25. R. Zardoya and A. Meyer, *Mol. Biol. Evol.* **13**: 933–942 (1996)
26. M.P. Cummings, S.P. Otto, and J. Wakeley, *Mol. Biol. Evol.* **12**: 814–822 (1995)
27. G.J.P. Naylor and W.M. Brown, *Nature* **388**: 527–528 (1997)
28. N. Goldman, J.L. Thorne, and D.T. Jones, *Genetics* **149**: 445–458 (1998)
29. Z. Yang in *Pacific Symposium on Biocomputing 2000*, Eds. R.B. Altman, A.K. Dunker, L. Hunter, K. Lauderdale, and T.E. Klein, pp. 18–92 (World Scientific Publishing, Singapore, 2000)
30. N. Howell and K. Gilbert, *J. Mol. Biol.* **203**: 607–617 (1988)
31. N. Howell, *J. Mol. Evol.* **29**: 157–169 (1989)
32. J.-P. di Rago, P. Netter, and P.P. Slonimski, *J. Biol. Chem.* **265**: 3332–3339 (1990)
33. J.W.O. Ballard and M. Kreitman, *Genetics* **138**: 757–772 (1994)
34. J. Lyons-Weiler and G.A. Hoelzer, *Mol. Biol. Evol.* **16**: 1400–1405 (1999)
35. W.M. Brown, E.M. Prager, A. Wang, and A.C. Wilson, *J. Mol. Evol.* **18**: 225–239 (1982)
36. J.P. Huelsenbeck and B. Rannala, *Science* **276**: 227–232 (1997)
37. Z. Yang, *J. Mol. Evol.* **42**: 294–307 (1996)
38. H. Akaike, *IEEE Trans. Automat. Contr.* **AC-19**: 716–723 (1974)
39. M. Nei, N. Takezaki, and T. Sitnikova, *Science* **267**: 253–255 (1995)
40. J. Felsenstein, *Evolution* **39**: 783–791 (1985)
41. D.M. Hillis, B.K. Mable, and C. Moritz in *Molecular Systematics*, Eds. D.M. Hillis, C. Moritz, and B.K. Mable, pp. 515–543 (Sinauer Associates, Sunderland, MA, 1996)