ARTICLE IN PRESS

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution xxx (2003) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera)

Kevin P. Johnson,[a,*] Robert H. Cruickshank,[b] Richard J. Adams,[c] Vincent S. Smith,[b] Roderic D.M. Page,[b] and Dale H. Clayton[c]

[a] *Illinois Natural History Survey, 607 East Peabody Drive, Champaign, IL 61820, USA*
[b] *Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK*
[c] *Department of Biology, University of Utah, Salt Lake City, UT, USA*

Received 1 April 2002; received in revised form 6 August 2002

Abstract

Few estimates of relative substitution rates, and the underlying mutation rates, exist between mitochondrial and nuclear genes in insects. Previous estimates for insects indicate a 2–9 times faster substitution rate in mitochondrial genes relative to nuclear genes. Here we use novel methods for estimating relative rates of substitution, which incorporate multiple substitutions, and apply these methods to a group of insects (lice, Order: Phthiraptera). First, we use a modification of copath analysis (branch length regression) to construct independent comparisons of rates, consisting of each branch in a phylogenetic tree. The branch length comparisons use maximum likelihood models to correct for multiple substitution. In addition, we estimate codon-specific rates under maximum likelihood for the different genes and compare these values. Estimates of the relative synonymous substitution rates between a mitochondrial (COI) and nuclear (EF-1α) gene in lice indicate a relative rate of several 100 to 1. This rapid relative mitochondrial rate (>100 times) is at least an order of magnitude faster than previous estimates for any group of organisms. Comparisons using the same methods for another group of insects (aphids) reveals that this extreme relative rate estimate is not simply attributable to the methods we used, because estimates from aphids are substantially lower. Taxon sampling affects the relative rate estimate, with comparisons involving more closely related taxa resulting in a higher estimate. Relative rate estimates also increase with model complexity, indicating that methods accounting for more multiple substitution estimate higher relative rates.

Keywords: Substitution rates; Phthiraptera; Lice; Insecta; COI; Elongation factor-1 α

1. Introduction

The elevated rate of substitution in mitochondrial DNA relative to nuclear DNA in animals is widely appreciated. Several studies have indicated an elevated rate of substitution, probably owing to underlying differences in mutation rates in vertebrate mitochondrial DNA, compared to comparable nuclear genes (Brown et al., 1979; Johnson and Clayton, 2000; Miyata et al., 1982). The relative rate of nucleotide substitution and mutation in mitochondrial versus nuclear genes has been less extensively examined in insects. However, most studies indicate that mitochondrial genes have a substitution and mutation rate that is 2–9 times faster than nuclear genes (DeSalle et al., 1987; Monteiro and Pierce, 2001; Moriyama and Powell, 1997; Satta et al., 1987; Tamura, 1992), and these estimates are similar to those for vertebrates. Studies comparing rates of mitochondrial substitution for parasitic insects and their vertebrate hosts have concluded that parasites show a faster substitution rate than their hosts (Hafner et al., 1994; Page et al., 1998; Paterson et al., 2000). Because all of these studies have involved lice (Insecta: Phthiraptera), it is unclear whether the elevated mitochondrial substitution rate is characteristic of all insects, or just lice. What is needed is a comparison of the relative mitochondrial to nuclear substitution rate in lice versus other insects.

* Corresponding author. Fax: 1-217-333-4949.
E-mail address: kjohnson@inhs.uiuc.edu (K.P. Johnson).

1055-7903/02/$ - see front matter © 2002 Published by Elsevier Science (USA).
PII: S1055-7903(02)00342-1

Estimating the relative rate of substitution or mutation between nuclear and mitochondrial genes is not as straightforward as it might at first appear. First, variation in selective constraints between different gene regions might lead to differences in the substitution rates. For example, such differences are known to occur between mitochondrial protein coding genes (Jacobs et al., 1988; Mindell and Thacker, 1996) and would affect estimates of relative rates of molecular evolution between mitochondrial and nuclear genes. Ideally, estimates of mutation should be derived from neutral sites because their substitution rates should directly reflect underlying mutation rates (Kimura, 1962).

Another potential concern is that, for comparisons between species, rapidly evolving loci are expected to undergo many multiple substitutions (Li, 1997). Methods that do not account for multiple substitution are likely to underestimate the actual genetic divergence and thus potentially bias the estimate of the relative substitution rates. Methods that rely on simple genetic distances or parsimony reconstruction are likely to miss such multiple substitutions (Johnson and Sorenson, 1998; Sanderson, 1990).

Here we use two methods for estimating the relative rates of mitochondrial and nuclear mutation. The first method is analogous to copath analysis, previously developed for comparison of host-parasite rates (Page, 1996). Copath analysis takes advantage of the many independent comparisons that can be gleaned from a large phylogeny, using branches common to two phylogenies. In the case of relative rates between genes, branch lengths for various gene regions are estimated for each branch in the tree, preferably using a method that accounts for multiple substitution. The lengths of the branches for the two genes are compared in a regression analysis to estimate the relative substitution rates (Page, 1996; Paterson et al., 2000). We assume that the same species phylogeny underlies both genes (Maddison, 1997; Pamilo and Nei, 1988; Slowinski and Page, 1999), such that branches in a species phylogeny are analogous to copaths.

To correct for multiple substitutions, we use a maximum likelihood framework (Felsenstein, 1981; Huelsenbeck and Crandall, 1997). Branch lengths for the two gene regions are estimated under a maximum likelihood model. Likelihood models directly incorporate multiple substitutions when estimating branch lengths, so this procedure has the best chance of minimizing unrecovered multiple substitutions that could compromise the rate comparison. To evaluate the effect of functional constraints on the relative rate estimate, we conduct the branch length regression using only third sites and compare this to the estimate from all sites. Since the majority of third position substitutions are synonymous, analysis of only third positions will minimize the impact of potentially selected amino acid replacement substitutions.

For the second method, we fit likelihood models to the two genes and test whether they show a significantly different rate using likelihood ratio tests. In addition, we test for site specific (codon position) rates using likelihood ratio tests and compare the rate parameters between genes. We also test whether the branch lengths for the two different genes are proportional. This method is fundamentally different than the regression of branch lengths, and provides a comparison to the branch length method.

For the rate comparison presented here, we evaluate the relative rates of substitution for a portion of the nuclear gene elongation factor-1 alpha (EF-1α) and a portion of the mitochondrial gene cytochrome oxidase I (COI) for lice (Phthiraptera). We evaluate the effect of taxonomic sampling by comparing a data set for 150 species, including no more than three species from any one genus, to two groups of dove lice that both include many species from the same genus (Johnson and Clayton, 2002). We compare the estimates for lice with those of other paraneopteran insects for the same two genes.

2. Materials and methods

2.1. Samples and sequencing

We obtained 150 species of bird and mammal lice, including representatives of all four suborders, using ethyl acetate fumigation (Clayton et al., 1992). From individual lice, we extracted DNA by removing the head from the body of the louse and placing both in a digestion buffer from a Qiagen Tissue Extraction kit. Extraction was completed according to manufacturer's protocols. After extraction, the head and body of each louse were mounted together on a microslide as a voucher. Vouchers were deposited in the Price Institute of Phthirapteran Research (University of Utah) or the University of Glasgow.

We amplified, using PCR, portions of the COI and EF-1α genes from each extract. For COI we used the primers L6625 and H7005 (Hafner et al., 1994) and for EF-1α the primers EF1-For3 and Cho10 (Danforth and Ji, 1998). PCR and sequencing were performed as described previously (Cruickshank et al., 2001; Johnson and Clayton, 2000). Sequences for complementary chromatograms were resolved and then aligned across species using the computer program Sequencher 3.0 (GeneCodes) or Sequence Navigator. Many of the EF-1α sequences were published previously (Cruickshank et al., 2001) and new sequences for EF-1α and COI were deposited in GenBank (Accession Numbers AF545666–AF545809).

To test the effects of taxon sampling, we used two previously published data sets of EF-1α and COI sequences in lice: *Columbicola* and Physconelloidinae

(Johnson and Clayton, 2002). These data sets contain 15 and 18 closely related taxa, respectively. As a comparison to other insects, we used a published data set for 14 closely related species of aphids (Hemiptera: *Uroleucon*) for the same genes (Moran et al., 1999). Aphid sequences for EF-1α were unavailable for the same region as the louse EF-1α sequences, but comparisons of rate variation among exons of EF-1α for aphids revealed little rate variation, so we selected an exon of similar size (exon 1), for comparison with the louse data.

2.2. Phylogenetic analysis

To compare rates, we used maximum likelihood (ML) to estimate branch lengths using PAUP* (Swofford, 2001). Maximum likelihood models take into account multiple substitutions and thus are more likely to accurately estimate the number of substitutions since an ancestor along a branch. However, because finding the optimal tree for 150 taxa under maximum likelihood is extremely computationally intensive, we chose to use phylogenetic estimates obtained from neighbor joining of combined COI and EF-1α data (uncorrected distances). We rooted trees for Phthiraptera using species in the suborder Amblycera as indicated by Lyal (1985) and Johnson and Whiting (2002).

Direct comparison of branch lengths, as in copath analysis, relies on the same underlying phylogeny for all the paths. This is the species phylogeny, and we feel that this is best estimated by combined analysis in this case. We used a partition homogeneity test (Farris et al., 1994, 1995; Swofford, 2001) and compared 50% bootstrap trees (Felsenstein, 1985) between COI and EF-1α for the 150 species of Phthiraptera to evaluate whether a similar phylogeny underlies both genes. A partition homogeneity test indicated significant heterogeneity between COI and EF-1α data sets ($P = 0.01$). However, no node conflicted at >60% bootstrap support between trees constructed from parsimony analysis of the two genes independently. Evaluations of the performance of the partition homogeneity test with the same underlying phylogeny, but involving data sets of different rates (Barker and Lutzoni, 2002; Darlu and Lecointre, 2002; Dolphin et al., 2000), indicate that when rates between data sets are dramatically different, significant heterogeneity can be detected even though there is no difference in the phylogenies underlying the data sets (see also Johnson and Whiting, 2002 and Johnson et al., 2002). Such an outcome can be detected by examining bootstrap topologies, with the expectation that in such cases, bootstrap topologies will not be in strong conflict (Barker and Lutzoni, 2002). As we show below, combining genes, even in the presence of heterogeneity, does not appear to affect the estimate of relative substitution rates, because data sets with congruent signal provide similar rate estimates. In addition, the dramatic difference in rates that we estimate below is likely to have caused the detection of significant heterogeneity by the partition homogeneity test.

We estimated the relative rates of substitution at third sites for COI and EF-1α using several different likelihood models. First, to provide the most parameter intensive estimate of branch lengths, we fit the most general maximum likelihood model (general time reversible, rate heterogeneity under a γ distribution with eight rate categories, and invariant sites [GTR + G + I]) to third sites for both COI and EF-1α. We estimated branch lengths using the NJ tree topology for all sites and for third sites only, with ML parameters estimated separately in both cases. We estimated the relative substitution rates for COI compared to EF-1α (third sites and all sites) using reduced major axis regression. Reduced major axis regression should be used when both x and y variables are estimated with error (McArdle, 1988). The slope of the reduced major axis regression is the ratio of standard deviations of estimated branch lengths. Because reduced major axis regression does not pair identical branches per se, rather involves the overall variance of branch lengths, we also estimated relative rates under ordinary least squares regression of COI branch lengths on EF-1α branch lengths. All least squares regressions were forced through the origin, because a branch length of zero for one gene should correspond to a branch length of zero for the other gene. In all cases, regression produced a highly significant correlation in branch lengths (all $P < 0.001$).

Branch length estimates may be sensitive to the maximum likelihood model chosen for analysis. Thus, in addition to using the parameter-rich (GTR + G + I) model, we also estimated branch lengths under the model favored by likelihood ratio tests (Huelsenbeck and Crandall, 1997), using the simplest model that could not be rejected in favor of a more complex model. We used the hierarchical likelihood ratio tests procedure as implemented in the program Modeltest (Posada and Crandall, 1998). Again we estimated these models for each gene independently, for third sites only and for all sites. Finally, to evaluate the role of unequal substitution types, base frequencies, and rate heterogeneity in relative rate estimates, we used the simplest maximum likelihood model (Jukes and Cantor, 1969) to estimate branch lengths (again for both third sites only and all sites).

For previously published data sets (Johnson and Clayton, 2002; Moran et al., 1999), we used the reported phylogenetic trees based on all the data over which to estimate branch lengths for particular genes and codon positions. In the case of *Columbicola* and aphids, this involved additional genes beyond COI and EF-1α. We estimated branch lengths using all the data partitions and maximum likelihood models described above, with

both regression techniques. All rates are reported as COI (mitochondrial), relative to EF-1α (nuclear).

We supplemented the comparison of branch lengths by performing a series of likelihood analyses using program *baseml* from the PAML 3.0a package (Yang, 1997). PAML allows data to be partitioned into different genes or codon positions and a rate parameter estimated for each partition independently. Because of the computationally intensive nature of these analyses we restricted our analysis to the *Columbicola*, Physconelloidinae, and aphid data sets. For each data set we constructed series of progressively less restrictive models and inferred the maximum likelihood branch lengths of the same phylogeny used in the branch length regression analyses. The most restrictive model treated the COI and EF-1α sequences as a single gene evolving at the same overall rate (this is model 0 in Table 2 of Yang (1996) and corresponds to setting Mgene = 0 in the baseml.ctl file in PAML). We then relaxed this constraint by partitioning the data into the two genes, but with the constraint that branch lengths for the two genes were proportional (model 3′, Mgene = 4). The program *baseml* reports the relative rate of evolution in the two partitions with respect to the rate in the first partition. To account for the different substitution rates at the three codon positions, we performed a third analysis with a total of six partitions, three for each gene. Finally, we relaxed the constraint that branch lengths are proportional across the six partitions (model 4, Mgene = 1). Under this model, PAML does not report the relative rate of substitution in the different partitions, in contrast to model 3′ in which the difference in substitution rate is constant across the tree and hence can be represented by a single value. Under model 4, two partitions may evolve at different rates, but those rates may vary across the tree, making it harder to describe by a single value (in the branch length comparison analyses we estimated the overall rate difference using regression).

We compared the models using likelihood ratio tests as described by Yang (1996). In each case we compute the deviance $2\Delta\ell = 2(\ell_1 - \ell_0)$, where ℓ_1 is the log likelihood of the more general model, and ℓ_0 is the log likelihood of the more restrictive model. The significance of the deviance is evaluated as χ^2 with the number of degrees of freedom being the difference in parameters between the two models. To test whether mitochondrial and nuclear genes are evolving at different rates we compared the log-likelihoods for the combined and two-partition analyses (model 0 and model 3′, respectively), which has one degree of freedom. Comparison of the two-gene and two-gene/three-codons models has four degrees of freedom. Model 4 estimates branch lengths for the $2n - 3$ branches in the g unrooted trees for n sequences, where g is the number of partitions. Hence, ignoring the sequence substitution and rate variation among site parameters, model 4 has $p = g(2n - 3)$ parameters. Model 3′ constrains the branch lengths in each partition to be proportional, hence estimates $2n - 3$ branch lengths and $g - 1$ rate ratios, giving $q = 2n + g - 4$ parameters. The degrees of freedom for comparing models 4 and 3′ is $p - q$.

In all analyses the REV model of sequence substitution was used, with among-site variation in rate modeled using a γ distribution with eight rate categories to account for many of the substitution properties of these data with reasonable computational complexity. This model is a subset of the most general model used in the PAUP* analyses (GTR + G + I), which has the additional parameter of proportion of invariant sites (I). The values for the parameters of the REV model and γ distribution (shape parameter α) were estimated separately for each partition (i.e., we allowed the substitution process to be different for each partition of the data).

3. Results

For the 150-taxon phthirapteran data set, uncorrected pairwise divergences for COI ranged from 2.4 to 41.5%, while uncorrected divergences for EF-1α ranged from 0 to 29.4% (Fig. 1). Neighbor joining produced a tree indicating monophyly of all suborders except Ischnocera (Fig. 2). *Haematomyzus elephantis* Piaget, the representative of the monogeneric suborder Rhynchophthirina, appeared within Ischnocera. This tree contained the branches on which maximum likelihood branch lengths were estimated for the 150-taxon data set. The GTR + G + I likelihood model for COI third

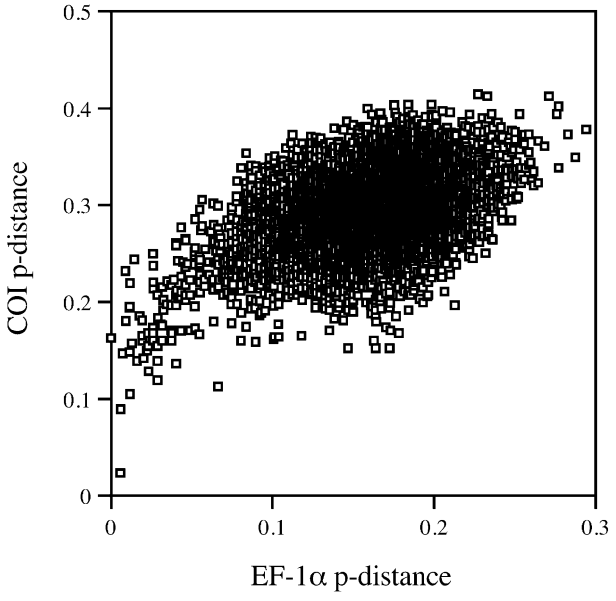

Fig. 1. Plot of uncorrected pairwise sequence divergence for COI against those for EF-1α using the 150 species data set.

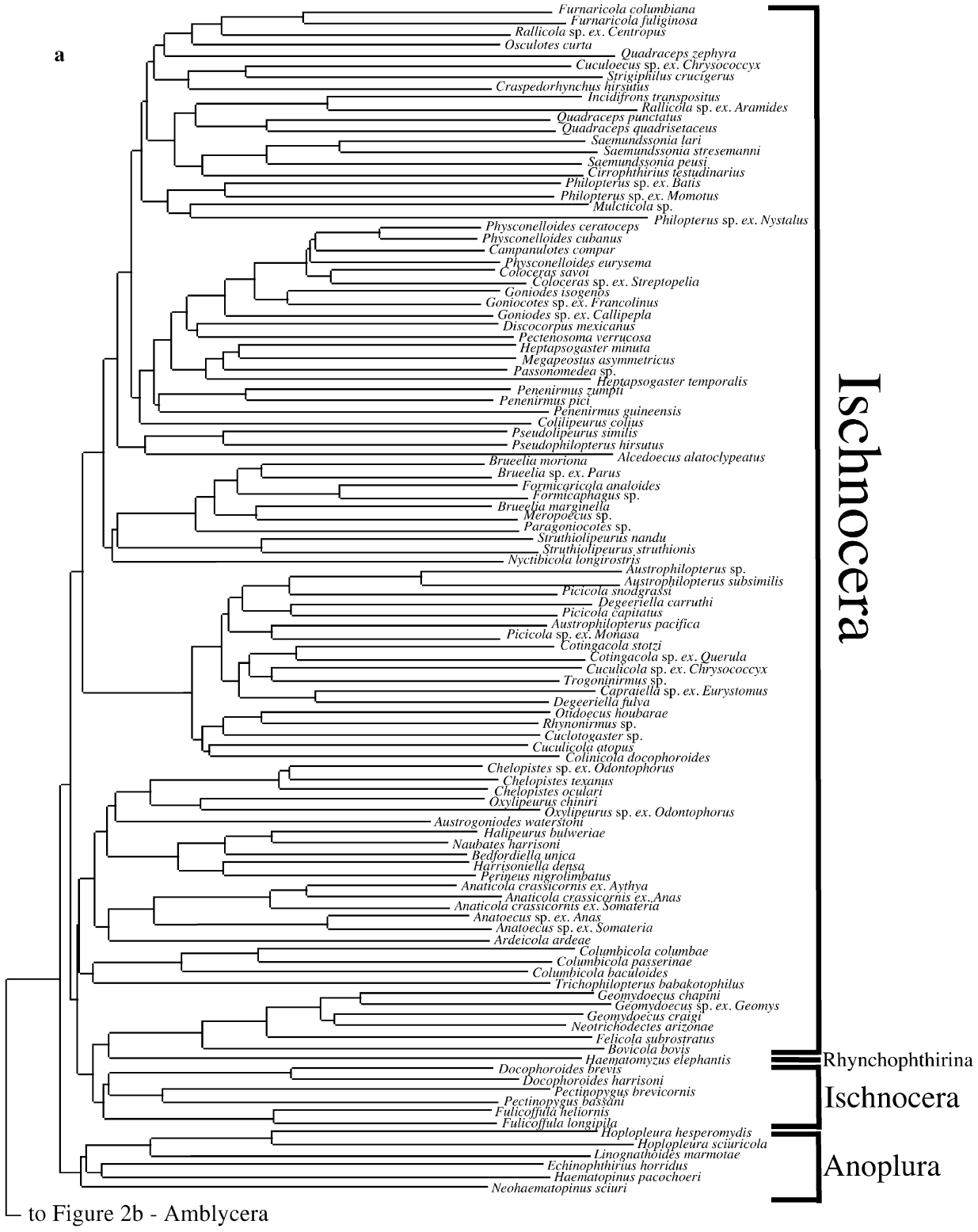

Fig. 2. Neighbor joining tree based on combined COI and EF-1α sequences for the 150 taxon Phthiraptera data set. (a) Ischnocera, Rhynchophthirina, and Anoplura; (b) Amblycera.

sites and EF-1α third sites produced branch lengths that differed dramatically between the two genes (Table 1). The estimated relative substitution rates for third sites was 610:1 using reduced major axis regression of branch lengths and 438:1 using least squares regression ($P < 0.0001$). Considering all sites, relative substitution

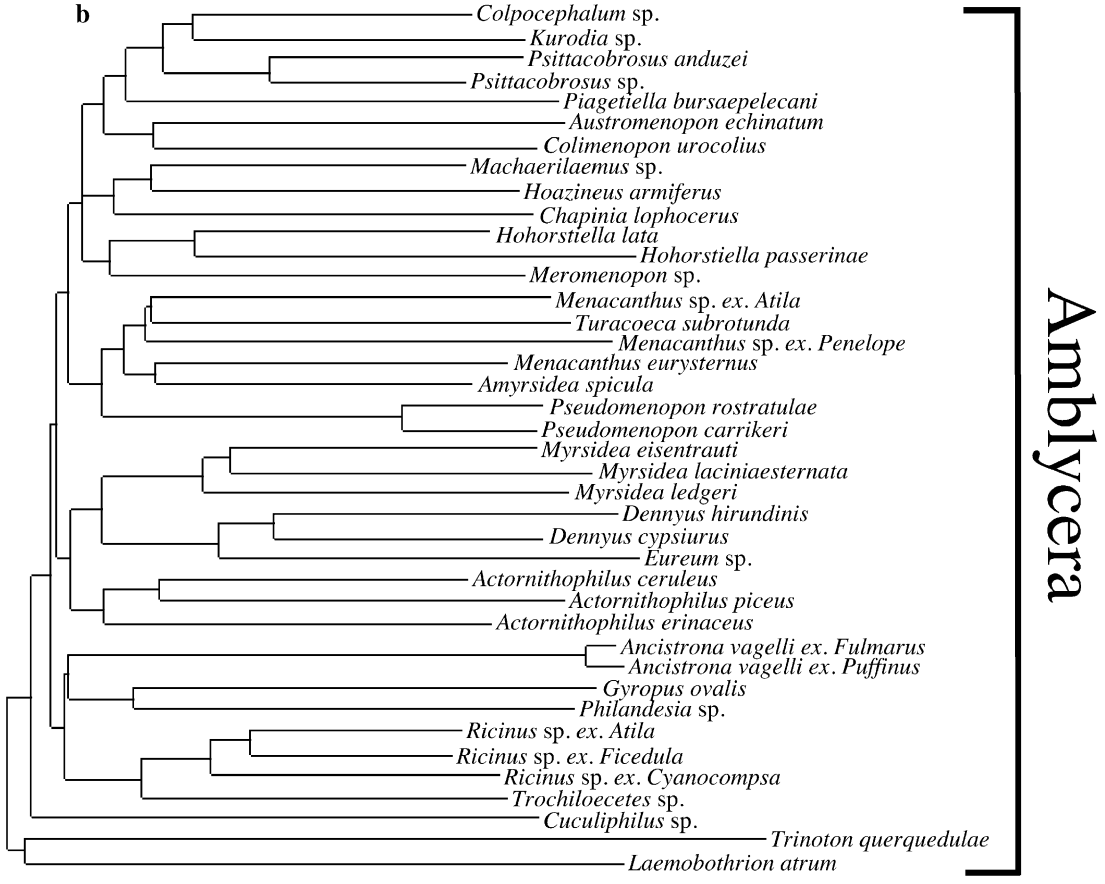

Fig. 2. (*continued*)

Table 1
Relative mitochondrial/nuclear rates (±SE) using all Phthiraptera (150 taxa)

Model	Sites	Major axis regression	Least squares regression
GTR + G + I	Third sites	609.7	437.5 ± 28.0
Modeltest[a]	Third sites	190.5	141.5 ± 8.7
Jukes–Cantor	Third sites	2.3	1.8 ± 0.1
GTR + G + I	All sites	5.5	5.1 ± 0.2
Modeltest[b]	All sites	5.1	4.8 ± 0.2
Jukes–Cantor	All sites	2.2	2.1 ± 0.1

[a] GTR + G (COI) and TIM + G + I (EF-1α).
[b] GTR + G + I (both genes).

rates were around 5:1 under the GTR + G + I model (Table 1). Using the models resulting from maximum likelihood ratio tests from Modeltest, relative rate estimates for third sites were still high, around one third the estimated rates under the GTR + G + I model. Relative substitution rates for all sites under the Modeltest models were slightly lower than those under the GTR + G + I model (Table 1). A Jukes–Cantor (Jukes and Cantor, 1969) model resulted in greatly reduced relative rate estimates, with similar estimates for third sites and all sites (Table 1).

For *Columbicola*, uncorrected pairwise divergences for COI ranged from 3.1 to 29.8%, while uncorrected divergences for EF-1α ranged from 0 to 11.8%. Estimates of relative rates for *Columbicola* from branch length regression (Table 2) were similar in magnitude to those for all Phthiraptera. However, the third site relative rate estimates for *Columbicola* were often higher than those for all Phthiraptera. Fig. 3 shows branch lengths for third positions of the two genes drawn to a similar scale. Relative rate estimates for all sites combined were an order of magnitude lower than those for third sites only (Table 2). Again, relative rates declined as the likelihood model became simpler. A Jukes–Cantor model produced little difference in the estimated relative rates based on third sites only and all sites.

Table 2
Relative mitochondrial/nuclear rates (±SE) using *Columbicola* (15 taxa)

Model	Sites	Major axis regression	Least squares regression
GTR + G + I	Third sites	599.0	412.1 ± 89.3
Modeltest[a]	Third sites	333.8	234.0 ± 50.5
Jukes–Cantor	Third sites	2.7	2.1 ± 0.6
GTR + G + I	All sites	58.5	41.9 ± 7.1
Modeltest[b]	All sites	44.2	33.4 ± 5.2
Jukes–Cantor	All sites	2.9	2.6 ± 0.5

[a] HKY + G (COI) and TVMef + G (EF-1α).
[b] TVM + G (COI) and TVMef + G (EF-1α).

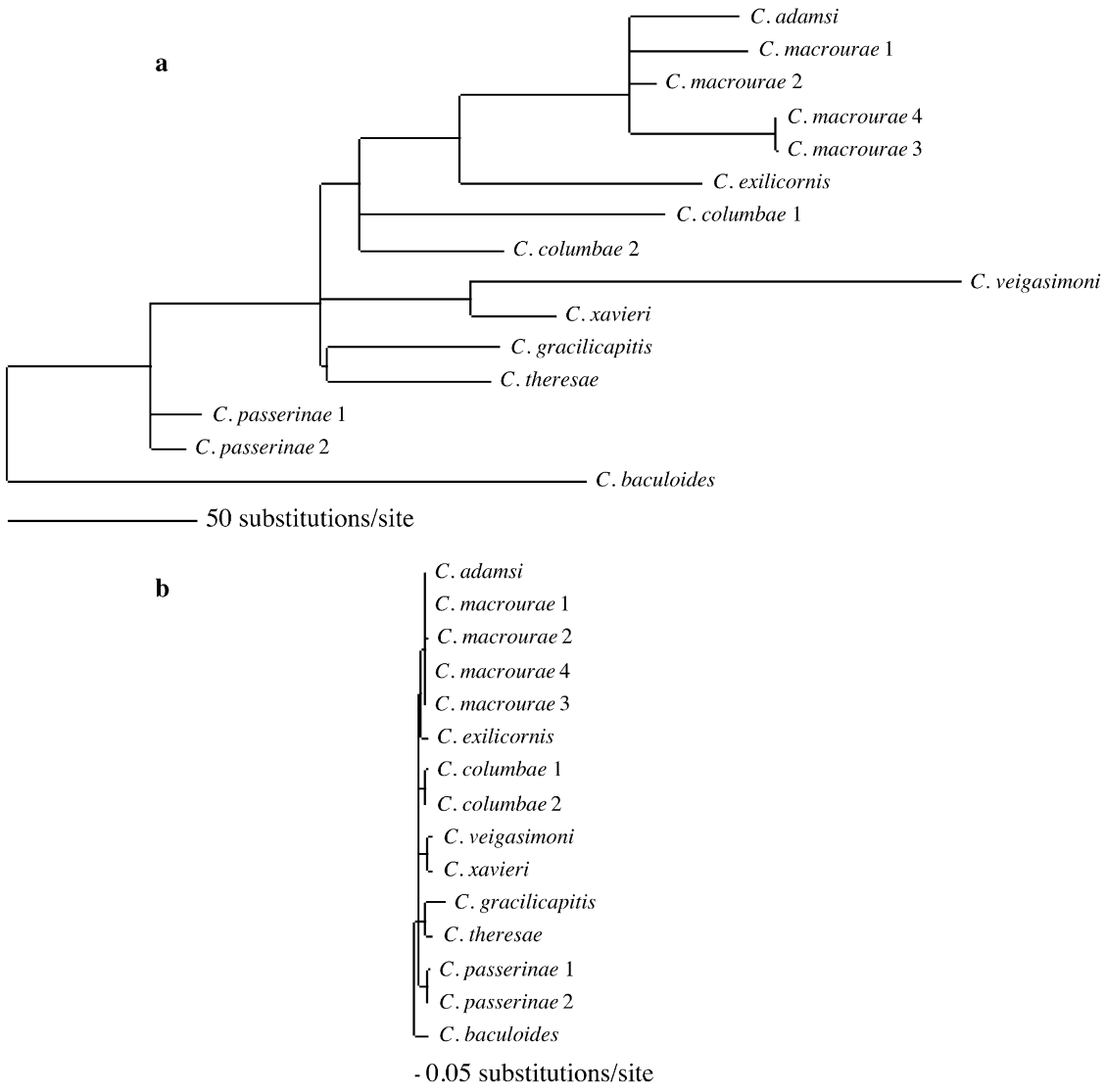

Fig. 3. Reconstruction of ML branch lengths over the *Columbicola* tree for (a) COI 3rd positions and (b) EF-1α 3rd positions. Scale indicated in the figure. Note that the branches of the EF-1α tree should be 30 times shorter to make the trees to the same scale. This was not done because of the difficulty of fitting both trees to scale on the same page. *C.* = *Columbicola*.

Uncorrected pairwise divergences of the two genes for Physconelloidinae ranged from 3.6 to 19.3% for COI and from 0 to 3.1% for EF-1α. Branch length regression estimates of relative rates using the Physconelloidinae data set (Table 3) produced similar results to both the Phthiraptera and *Columbicola* analyses. However,

Table 3
Relative mitochondrial/nuclear rates (±SE) using Physconelloidinae (18 taxa)

Model	Sites	Major axis regression	Least squares regression
GTR + G + I	Third sites	1002.8	515.2 ± 197.5
Modeltest[a]	Third sites	431.3	219.6 ± 84.3
Jukes–Cantor	Third sites	9.4	5.1 ± 2.0
GTR + G + I	All sites	213.2	104.4 ± 37.1
Modeltest[b]	All sites	240.3	117.4 ± 41.1
Jukes–Cantor	All sites	5.1	5.4 ± 1.7

[a] TrN + G (COI) and K81uf + G (EF-1α).
[b] GTR + G + I (COI) and TVMef + G (EF-1α).

relative rate estimates were generally higher for the Physconelloidinae data set than for either the Phthiraptera or *Columbicola* data sets.

For aphids, uncorrected pairwise divergences for COI ranged from 0.8 to 5.8%, while uncorrected divergences for EF-1α ranged from 0 to 3.6%. In contrast to the three louse data sets, estimates of the relative rates using branch length regression of the aphid data set produced markedly lower estimates (Table 4). Relative rate estimates for third sites in aphids using the most complex likelihood models were around an order of magnitude lower (with a maximum around 20:1) than the louse estimates (around 400:1 or greater). For the aphids, more complex models, and analyses using third sites only produced greater relative rate estimates, in a similar manner to the louse comparisons.

Analysis of these data sets using codon specific rate estimates from PAML (Yang, 1997) produced very similar results to the branch length regression analysis. The likelihood ratio tests for the *Columbicola*, Physconelloidinae, and aphid data sets showed the same pattern (Table 5). These tests indicated a significant difference between mitochondrial and nuclear rates and significant variation in rate among codon positions. In all three of these data sets we could not reject the branch lengths proportional model, hence we can use the relative rates reported by the model 3′ analysis (Table 6). For lice, if we combined the three codon positions together, the rate of substitution in the mitochondrial COI gene was 50–60 times higher than the rate in the nuclear EF-1α gene, whereas in aphids the difference was 4 times. At the third codon position the louse data sets showed around 200–300 times higher levels of substitution in their mitochondrial genes, compared to a difference of 18 times in aphids. In both cases the lice showed approximately an order of magnitude higher relative rate of substitution in COI with respect to EF-1α than did aphids.

4. Discussion

Branch length regression analysis provides a means of obtaining many independent points for the calculation of relative substitution rates between genes. Using this method, estimates of relative rates of substitution between mitochondrial (COI) and nuclear genes (EF-1α) at third sites for lice (>100:1) are generally one to two orders of magnitude greater than previous estimates for insects or any other organisms. Similar rates also are estimated when codon specific rate parameters are used to calculate relative rates. Taxon sampling affects the estimate of relative rate. Estimates involving louse taxa with lower overall sequence divergence were generally higher than those involving more distantly related taxa. Such a result would be expected if multiple substitutions are underestimated for the fast gene in more distantly related taxa. Despite the apparent bias in estimated relative rates with overall genetic divergence, relative rate estimates for lice are remarkably high in all data sets and for both methods of analysis. Use of the same methods for an aphid data set resulted in an estimated mitochondrial:nuclear relative rate of around 20:1 or less. This suggests that either the mutation rate of louse mitochondrial DNA is elevated or that the nuclear mutation rate is suppressed, relative to other insects.

Table 4
Relative mitochondrial/nuclear rates (± SE) using aphids (*Uroleucon*) (14 taxa)

Model	Sites	Major axis regression	Least squares regression
GTR + G + I	Third sites	21.5	19.5 ± 2.7
Modeltest[a]	Third sites	12.3	11.2 ± 1.6
Jukes–Cantor	Third sites	1.9	1.7 ± 0.3
GTR + G + I	All sites	3.2	3.1 ± 0.4
Modeltest[b]	All sites	3.0	2.9 ± 0.4
Jukes–Cantor	All sites	1.9	1.9 ± 0.3

[a] HKY+G (COI) and HKY (EF-1α).
[b] GTR + G (COI) and K80 (EF-1α).

Table 5
Likelihood ratio tests of different models of nucleotide substitution in the three insect taxa

Dataset	Comparison	Log-likelihoods						
		$Model_1$	$Model_0$	ℓ_1	ℓ_0	$2\Delta\ell$	df	$p(\chi^2)$
Physconelloidinae	Mitochondrial and nuclear rates	Different	Equal	−3746.467	−3861.166	229.34	1	≫0.001
	Rate at each codon position	Different	Same	−3405.377	−3746.467	682.18	4	≫0.001
	Branch lengths proportional	No	Yes	−3338.039	−3405.377	134.68	180	0.995
Columbicola	Mitochondrial and nuclear rates	Different	Equal	−3813.106	−4179.729	733.25	1	≫0.001
	Rate at each codon position	Different	Same	−4179.729	−4312.841	266.23	4	≫0.001
	Branch lengths proportional	No	Yes	−3735.952	−3813.106	154.31	130	0.072
Aphids	Mitochondrial and nuclear rates	Different	Equal	−1490.312	−1522.976	65.33	1	≫0.001
	Rate at each codon position	Different	Same	−1333.988	−1490.312	312.65	4	≫0.001
	Branch lengths proportional	No	Yes	−1317.614	−1333.988	32.75	50	0.568

Note. Each comparison is between a more general $model_1$, and a more restrictive submodel, $model_0$. In each case the table gives the log-likelihoods for the two models, the deviance ($2\Delta\ell$), the degrees of freedom (df), and the probability of the deviance under the χ^2 distribution.

Phylogenetic analysis of louse EF-1α in relation to other insects indicates that, if anything, branch lengths for EF-1α are *longer* in lice (unpublished analysis). Thus, it appears that mitochondrial substitution rates are elevated in lice, relative to other insects.

Differences in functional constraint at the amino acid level can potentially affect estimates of the mutation rate directly from the substitution rate. This effect can be seen when first and second positions of codons are incorporated into the relative rate estimate. Most first and second position substitutions are nonsynonymous. Estimates of relative mitochondrial:nuclear rates for lice that incorporate all positions generally range from 5:1 to 100:1. These estimates appear to be even more sensitive to taxon sampling than estimates from third positions alone (see Tables 1–3). Selective constraints should only make our relative rate estimate from third sites an underestimate, if anything, because some third sites are nonsynonymous sites.

A relative mutation rate between mitochondrial and nuclear genes for lice of 100:1 or higher is unknown for any group of organisms. Is this estimate reasonable? Detailed examination of pairwise sequence divergences between closely related taxa suggest that this is probably not an artifact of our methods. For example, *Columbicola adamsi* Clayton and Price and *C. macrourae* (Wilson) differ from each other by 20% uncorrected pairwise sequence divergence for the COI gene. Comparisons of third positions only for COI in these same taxa reveal uncorrected divergences of 54%. However, *C. adamsi* and *C. macrourae* show *no* differences in sequences of the EF-1α gene. Certainly correction for multiple substitutions will only increase COI divergence estimates, given that multiple substitutions are likely to be rampant at third positions with such dramatically high divergences. Given these considerations, relative rate estimates should be much greater than 50:1, given multiple substitution and the fact that not one substitution event in EF-1α has yet occurred between these taxa, despite the high COI divergence. As further illustration, we plotted ML corrected pairwise divergences for third sites for COI against those for EF-1α in *Columbicola* (Fig. 4). The initial slope provides an estimate of the relative rate (Sturmbauer and Meyer, 1992), here around 1000:1. A similar pattern emerges in Physconelloidinae, where many highly divergent species at the mitochondrial level show no divergence in EF-1α sequences.

Previous estimates of relative substitution rates in many organisms may be underestimates because of the difficulties of compensating for multiple substitutions. Denver et al. (2000) estimated mitochondrial mutations directly from mutation accumulation lines in *Caenorhabditis elegans* and calculated a mutation rate two orders of magnitude higher than previous indirect estimates. Thus, selection against mutations and the

Table 6
Rate of substitution in COI and EF-1α genes in three insect taxa

Data set	Sites	Rate parameter		Ratio COI:EF-1α
		COI	EF-1α	
Physconelloidinae	All sites	1	0.018	55.6
	pos1	1	0.013	76.9
	pos2	0.12	0.016	7.5
	pos3	129.9	0.41	316.8
Columbicola	All sites	1	0.016	61.5
	pos1	1	0.039	25.6
	pos2	0.15	0.017	8.8
	pos3	169.78	0.80	212.2
Aphids	All sites	1	0.25	4.1
	pos1	1	0.20	5.0
	pos2	0.0001	0.0001	1.0
	pos3	44.79	2.49	18.0

Note. When all sites in the two genes are combined the rate in EF-1α is reported as a proportion of the rate in COI. When the sites are treated by codon position the rates at COI positions 2 and 3, and all sites in EF-1α are with respect to the first position in COI.

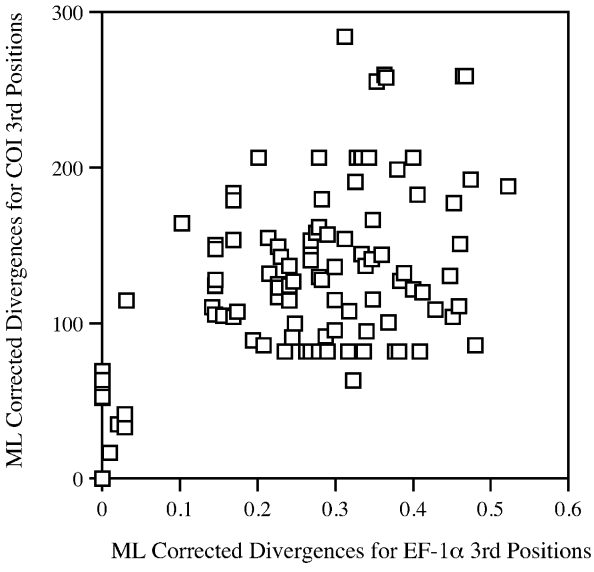

Fig. 4. Pairwise ML (GTR + G + I) corrected distances for COI 3rd sites against ML (GTR + G + I) corrected distances for EF-1α 3rd sites in *Columbicola*. The slope of the initial increase is an approximation of the relative rate (around 1000:1), because multiple substitutions for low divergences are likely to be minimal.

difficulties of identifying multiple substitutions may often lead to underestimates of mutation rates at rapidly evolving loci.

While we are confident that our relative rate estimates using copath and PAML analyses are reasonable, some concerns remain. First, for the 150 taxon data set, the COI and EF-1α data were not homogeneous as determined by the partition homogeneity test (Farris et al., 1994, 1995; Swofford, 2001), and thus whether gene phylogenies reflect the species phylogeny is in doubt. Heterogeneity between gene partitions can arise from at least two processes. First, a different underlying phylogeny between the two partitions may result in significant heterogeneity. Phylogenetic incongruence between genes may occur in cases of lineage sorting incongruent with the species tree (Pamilo and Nei, 1988), hybridization (Johnson and Sorenson, 1999; Mason-Gamer and Kellogg, 1996), or gene duplication (Slowinski and Page, 1999). Alternatively, underlying differences in substitution properties may result in detection of heterogeneity by the partition homogeneity test. When one gene evolves at a much faster rate than another, even given the same underlying phylogeny, the partition homogeneity test can reach significance (Barker and Lutzoni, 2002; Darlu and Lecointre, 2002). In such cases, bootstrap topologies are unlikely to show conflicting nodes between the two gene regions (Barker and Lutzoni, 2002; Johnson and Whiting, 2002; Johnson et al., 2002).

We demonstrate a substantially elevated substitution rate in COI as compared to EF-1α and detect no topological differences that are strongly supported (>60%) by bootstrapping. These observations suggest that the significance of the partition homogeneity test is a result of a difference in rates, rather than a difference in gene phylogenies. Thus, a combined phylogeny is still a reasonable estimate of the species phylogeny, providing homologous branches ("copaths"). The fact that the *Columbicola* and Physconelloidinae data sets did not exhibit heterogeneity ($P > 0.10$), but showed even more dramatic rate differences, suggests that the combination of heterogeneous genes, in the case of the 150 species data set, did not provide an upwardly biased estimate of relative rates.

A second concern is that we did not use the same region of the EF-1α gene in the louse to aphid comparison because such homologous sequences were unavailable. In general, it is likely that variation between nuclear regions is minor and would probably not lead to the order of magnitude difference in substitution rates

between lice and aphids. Sequences for the nuclear gene Wingless in lice show a similar overall divergence to EF-1α sequences (unpub. data). In addition, sequences of other mitochondrial regions in lice (e.g., cyt *b*) show similar, or slightly higher, divergences than sequences for the region of the COI gene presented here (unpub. data). To further test this potential problem, we analyzed sequences for COI and EF-1α regions identical to those for lice for seven species of treehoppers (Insecta: Membracoidea; Cryan et al., 2000; C. Dietrich and J. Cryan, pers. comm.). PAML analysis of these sequences estimated a 38:1 relative rate at third sites, which is similar to the estimate for aphids, but much less than the relative rate for lice. Thus, for insects, it appears that rate variation within the nuclear and mitochondrial genomes (at least for third positions) is relatively minor and unlikely to drastically change our results.

In sum, mitochondrial substitution rates at third codon positions in lice appear to be dramatically elevated (>100 times) over nuclear substitution rates at third codon positions. Both branch length regression and site specific rate parameter estimates have a high potential to uncover multiple substitutions, which would otherwise be missed when rate differences are extreme. Such high relative rates are unknown in other organisms. Differences in mutation rates may arise through differences in mitochondrial mutation pressure (e.g., oxidative stress) or the DNA repair machinery. The louse mitochondrial genome has several gene rearrangements relative to other insects (Shao et al., 2001). These rearrangements may alter mutation rates either through differences in replication rate or in DNA repair. Aphids, treehoppers, and lice are all paraneopteran insects (Yoshizawa and Saigusa, 2001) and thus it appears that the rate elevation in mtDNA is unique to the lineage leading to lice (see also Simmons and Weller, 2001). To uncover the origin of elevated mutation rates, further comparisons of relative rates and mitochondrial genome organization in even more closely related groups of insects (e.g., Psocoptera) are needed to more completely document and understand the origin of such dramatic differences in substitution rates in lice.

Acknowledgments

The following individuals provided assistance in obtaining and identifying louse specimens for this study: B. Appleby, S. Aquino, C. Baduini, P. Battley, S. Bertelli, British Antarctic Survey, R. Cuamatzi, B. Dawson (International Foundation for Conservation and Development of Wildlife), J. de Cruz, J. Demastes, L. Durden, R. T. Engstrom, G. Escalona-Segura, R. Faucett, C. Floyd, B. Furness, J. Pedro Granadeiro, R. Griffiths, M. Hafner, M. Hastriter, H. Hira, N. Hoffman, J. K. Jensen, S. de Kort, C. Lyal, B. Marks, S. Meagher, D. Mellor, E. Mey, R. Moyle, A. Navarro, C. Newton, E. Osnas, R. Palma, J. Parga, A. Paterson, M. Penck, A. T. Peterson, R. Price, H. Proctor, M. Robbins, O. Rojas, D. Reed, S. Sharp, J. Taylor, P. Thompson, J. Weckstein, M. Whiting, C. Witt, and B. Zonfrillo. We thank Z. Yang for advice on using PAML. M. Kennedy, J. Koontz, P. Pamilo, and D. Reed provided helpful comments on the manuscript. We thank the DNA sequencing facilities of the Universities of Glasgow and Utah, the latter supported in part by NCI Grant 5p30CA23014. C. Dietrich and J. Cryan kindly provided us with sequences of Membracoidea for analysis. This work was supported by NERC Grant GR3/11075 (to R.D.M.P.), NSF-CAREER award DEB-9703003 (to D.H.C.), and NSF DEB-0107891 (to K.P.J.).

References

Barker, F.K., Lutzoni, F., 2002. The utility of the incongruence length difference test. Syst. Biol. 51, 625–637.

Brown, W.M., George Jr., M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. PNAS 76, 1967–1971.

Clayton, D.H., Gregory, R.D., Price, R.D., 1992. Comparative ecology of neotropical bird lice. J. Anim. Ecol. 61, 781–795.

Cruickshank, R.H., Johnson, K.P., Smith, V.S., Adams, R.J., Clayton, D.H., Page, R.D.M., 2001. Phylogenetic analysis of partial sequences of elongation factor 1α identifies major groups of lice (Insecta: Phthiraptera). Mol. Phylogenet. Evol. 19, 202–215.

Cryan, J.R., Wiegmann, B.N., Deitz, L.L., Dietrich, C.H., 2000. Phylogeny of treehoppers (Insecta: Hemiptera: Membracidae): evidence from two nuclear genes. Mol. Phylogenet. Evol. 17, 317–334.

Danforth, B.N., Ji, S., 1998. Elongation factor-1α occurs as two copies in bees: implications for phylogenetic analysis of EF-1α sequences in insects. Mol. Biol. Evol. 15, 225–235.

Darlu, P., Lecointre, G., 2002. When does the incongruence length difference test fail? Mol. Biol. Evol. 19, 432–437.

Denver, D.R., Morris, K., Lynch, M., Vassilieva, L.L., Thomas, W.K., 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science 289, 2342–2344.

DeSalle, R., Freedman, T., Prager, E.M., Wilson, A.C., 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. 26, 157–164.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17, 401–406.

Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1994. Testing significance of congruence. Cladistics 10, 315–320.

Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1995. Constructing a significance test for incongruence. Syst. Biol. 44, 570–572.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Hafner, M.S., Sudman, P.D., Villablance, F.X., Spradling, T.A., Demastes, J.W., Nadler, S.A., 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 365, 1087–1090.

Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28, 437–466.

Jacobs, H., Elliott, D., Math, V., Farquharson, A., 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. 202, 185–217.

Johnson, K.P., Clayton, D.H., 2002. Coevolutionary history of ecological replicates: comparing phylogenies of wing and body lice to Columbiform hosts. In: Page, R.D.M. (Ed.), Tangled Trees: Phylogeny, Cospeciation, and Coevolution. University of Chicago Press, Chicago.

Johnson, K.P., Clayton, D.H., 2000. Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). Mol. Phylogenet. Evol. 14, 141–151.

Johnson, K.P., Sorenson, M.D., 1998. Comparing molecular evolution in two mitochondrial protein coding genes (cytochrome *b* and ND2) in the dabbling ducks (Tribe: Anatini). Mol. Phylogenet. Evol. 10, 82–94.

Johnson, K.P., Sorenson, M.D., 1999. Phylogeny and biogeography of dabbling ducks (Genus: *Anas*): a comparison of molecular and morphological evidence. Auk 116, 792–805.

Johnson, K.P., Weckstein, J.D., Witt, C.C., Faucett, R.C., Moyle, R.G., 2002. The perils of using host relationships in parasite taxonomy: phylogeny of the *Degeeriella* complex. Mol. Phylogenet. Evol. 23, 150–157.

Johnson, K.P., Whiting, M.F., 2002. Multiple genes and the monophyly of Ischnocera (Insecta: Phthiraptera). Mol. Phylogenet. Evol. 22, 101–110.

Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 21–123.

Kimura, M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47, 713–719.

Li, W-H., 1997. Molecular Evolution. Sinauer Associates, Inc., Sunderland, MA.

Lyal, C.H.C., 1985. Phylogeny and classification of Psocodea, with particular reference to lice (Psocodea: Phthiraptera). Syst. Entomol. 10, 145–165.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Mason-Gamer, R.J., Kellogg, E.A., 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Syst. Biol. 45, 524–545.

McArdle, B.H., 1988. The structural relationship: regression in biology. Can. J. Zool. 66, 2239–2339.

Mindell, D.P., Thacker, C.E., 1996. Rates of molecular evolution: phylogenetic issues and applications. Annu. Rev. Ecol. Syst. 27, 279–303.

Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M., Koike, K., 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. 19, 28–35.

Monteiro, A., Pierce, N.E., 2001. Phylogeny of Bicyclus (Lepidoptera: Nymphalidae) inferred from COI, COII, and EF-1α gene sequences. Mol. Phylogenet. Evol. 18, 264–281.

Moran, N.A., Kaplan, M.E., Gelsey, M.J., Murphy, T.G., Scholes, E.A., 1999. Phylogenetics and evolution of the aphid genus *Uroleucon* based on mitochondrial and nuclear DNA sequences. Syst. Entomol. 24, 85–93.

Moriyama, E.N., Powell, J.R., 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. J. Mol. Evol. 45, 378–391.

Page, R.D.M., 1996. Temporal congruence revisited: comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice. Syst. Biol. 45, 151–167.

Page, R.D.M., Lee, P.L.M., Becher, A., Griffiths, R., Clayton, D.H., 1998. A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. Mol. Phylogenet. Evol. 9, 276–293.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Paterson, A.M., Wallis, G.P., Wallis, L.J., Gray, R.D., 2000. Seabird and louse coevolution: complex histories revealed by 12S rRNA sequences and reconciliation analysis. Syst. Biol. 49, 383–399.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Sanderson, M.J., 1990. Estimating rates of speciation and evolution: a bias due to homoplasy. Cladistics 6, 387–391.

Satta, Y., Ishiwa, H., Chigusa, S.I., 1987. Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Mol. Biol. Evol. 4, 638–650.

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

Simmons, R.B., Weller, S.J., 2001. Utility and evolution of cytochrome *b* in insects. Mol. Phylogenet. Evol. 20, 196–210.

Slowinski, J.B., Page, R.D.M., 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48, 814–825.

Sturmbauer, C., Meyer, A., 1992. Genetic divergence, speciation, and morphological stasis in a lineage of African cichlid fishes. Nature 358, 578–581.

Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony, ver. 4.0, β. Sinauer, Sunderland, MA.

Tamura, K., 1992. The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. Mol. Biol. Evol. 9, 814–825.

Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biol. Sci. 13, 555–556.

Yoshizawa, K., Saigusa, T., 2001. Phylogenetic analysis of paraneopteran orders (Insecta: Neoptera) based on forewing base structure, with comments on monophyly of Auchenorrhyncha (Hemiptera). Syst. Entomol. 26, 1–13.