Louse (Insecta: Phthiraptera) mitochondrial 12S rRNA secondary structure is highly variable

R. D. M. Page,* R. Cruickshank* and K. P. Johnson†

*Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK; †Illinois Natural History Survey, 607 East Peabody Drive, Champaign, IL 61820-6970, USA

Abstract

Lice are ectoparasitic insects hosted by birds and mammals. Mitochondrial 12S rRNA sequences obtained from lice show considerable length variation and are very difficult to align. We show that the louse 12S rRNA domain III secondary structure displays considerable variation compared to other insects, in both the shape and number of stems and loops. Phylogenetic trees constructed from tree edit distances between louse 12S rRNA structures do not closely resemble trees constructed from sequence data, suggesting that at least some of this structural variation has arisen independently in different louse lineages. Taken together with previous work on mitochondrial gene order and elevated rates of substitution in louse mitochondrial sequences, the structural variation in louse 12S rRNA confirms the highly distinctive nature of molecular evolution in these insects.

Keywords: lice, multiple alignment, rRNA secondary structure, 12S rRNA, tree comparison metrics.

Introduction

Congruent host and parasite phylogenies provide a unique framework for comparing rates of molecular evolution in taxonomically distant organisms (Hafner & Nadler, 1990; Moran *et al.*, 1995; Page & Hafner, 1996). By comparing sequence divergence in homologous genes in co-speciating hosts and their parasites it is possible to determine relative rates of sequence evolution without reference to the fossil record. The prevalence of co-speciation between lice and

their hosts has resulted in these insects playing a key role in recent methodological and empirical (Page, 2002; Paterson & Banks, 2001) studies of relative rates of evolution in host and parasite assemblages.

The large numbers of mitochondrial small subunit (12S) rRNA sequences available for birds (Houde *et al.*, 1997; Mindell *et al.*, 1997) and mammals (Springer & Douzery, 1996) makes this gene an attractive candidate for comparing molecular evolution in lice and their vertebrate hosts. However, the first published louse 12S rRNA sequences were described as 'highly unusual' (Paterson *et al.*, 2000, p. 390), lacking some highly conserved motifs typical of animal 12S rRNA (Hickson *et al.*, 1996), and having large (35–59 base pair) insertions.

Faced with highly variable rRNA sequences that are difficult to align using standard methods (e.g. Stoye *et al.*, 1997; Thompson *et al.*, 1997), one approach is to use secondary structures to guide the alignment (Buckley *et al.*, 2000; Hickson *et al.*, 1996; Kjer, 1995). However, aligning sequences using their secondary structure requires a considerable manual effort. Concern about the potential for subjectivity in such alignments has motivated the development of automatic tools for generating sequence alignments (e.g. Wheeler, 1994). Indeed, some have argued that manual alignments should be avoided altogether (Phillips *et al.*, 2000). However, the main reason manual alignments are needed is that existing algorithms are not up to the task. Automatic methods for aligning primary sequences can fail to align RNA sequences correctly (Hickson *et al.*, 2000), and methods for automatically aligning sequences using both primary and secondary structure are in their infancy (Corpet & Michot, 1994; Lenhof *et al.*, 1998; Notredame *et al.*, 1997).

We have obtained sequences from domain III of 12S rRNA for a wide range of lice, including representatives from three of the four suborders of Phthiraptera. Within small clades of lice (such as a single genus) the alignment is usually relatively straightforward. Automatic alignment methods such as Clustal (Thompson *et al.*, 1997) and DCA (Stoye *et al.*, 1997) generate clean alignments, in agreement with the results of Hickson *et al.* (2000). However, applying programs such as Clustal to all louse sequences together resulted in very poor quality

alignments, especially towards the 3′ end of domain III. We therefore pursued the use of secondary structure to improve the alignment. Our preliminary attempts to use published secondary structures for other animals (including insects) met with mixed success, and suggested that some louse taxa contained large insertions in regions that are highly conserved in other taxa. Furthermore, some helices were not easy to locate in lice, or had different structures.

In order to assess the degree of sequence and structure variation in insect domain III 12S rRNA, Page (2000b) constructed a secondary structure model based on an alignment of 225 insect sequences using maximum weighted matching (Page, 2000a; Tabaska *et al.*, 1998). In the present study we use this model (Fig. 1) as a benchmark for determining degree of variation in the secondary structure in lice. Given that louse 12S rRNA secondary structure is variable, we investigated whether the variation contains information about the evolutionary relationships of lice (rather than being merely a nuisance when constructing alignments).

Results

Louse secondary structure

Secondary structures were inferred for each sequence listed in Table 1. The key features of these structures are summarized in Fig. 2. Drawings of structure for each individual sequence are available online <http://r6-page.zoology.gla.ac.uk/lousebase/2/12s/>. The inferred
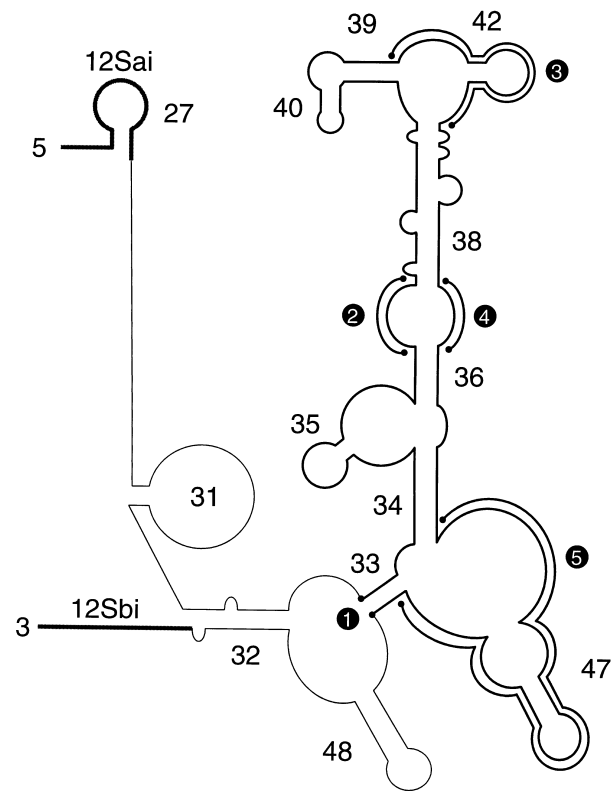

Figure 1. Generalized secondary structure of insect domain III 12S rRNA (from Page, 2000b, fig. 6), showing the location of the 12Sai and 12Sbi PCR primers (Simon *et al.*, 1994). The segments used to measure length variation in sequences (Fig. 3) are numbered 1–5. Segment 1 corresponds to the entire structure bounded by helix 33.

Taxon	12S rRNA	COI	EF1α
Anaticola crassicornis	AF396482	AF396541	AF320354
Anatoecus spp.	AF396485	AF396542	AF320356
Ancistrona vagelli	AF189128	AF497798	AF320358
Ardeicola spp.	AF396486	AF396545	AF320361
Austrogoniodes watersoni	AF189129	AF497799	AF320362
Austrophilopterus subsimilis	AF189130	AF348874	AF320365
Campanulotes compar	AF189131	AF348836	AF320377
Columbicola baculoides	AF190425	AF414764	AF320384
Columbicola columbae	AF190415	AF278620	AF320385
Discocorpus mexicanus	AF189133	AF498002	AF320392
Docophoroides brevis	AF396488	AF396547	AF320394
Echinophthirius horridus	AF189134	AF498001	AF320396
Haffneria grandis	AF189135	AF396553	AF320406
Halipeurus sp.	AF189136	AF396559	AF320408
Harrisoniella densa	AF396501	AF396567	AF320410
Naubates harrisoni	AF396504	AF396571	AF320432
Oxylipeurus chiniri	AF189140	AF348872	AF320436
Paraclisis confidens	AF396511	AF396579	AF502566
Pectenosoma verrucosa	AF189141	AF348862	AF320440
Pectinopygus brevicornis	AF189142	AF497800	AF320442
Quadraceps sp.	AF396527	AF396599	AF320458
Rallicola spp.	AF189144	AF348867	AF320459
Saemundssonia stresemanni	AF189145	AF396612	AF320466
Austrogoniodes cristati	Y14909	–	–
Naubates prioni	AF396509	–	–

Table 1. Taxa and sequences used in this study. For each sequence the GENBANK accession number is given. The designation 'spp.' indicates that the three sequences came from different species, otherwise the same species was used
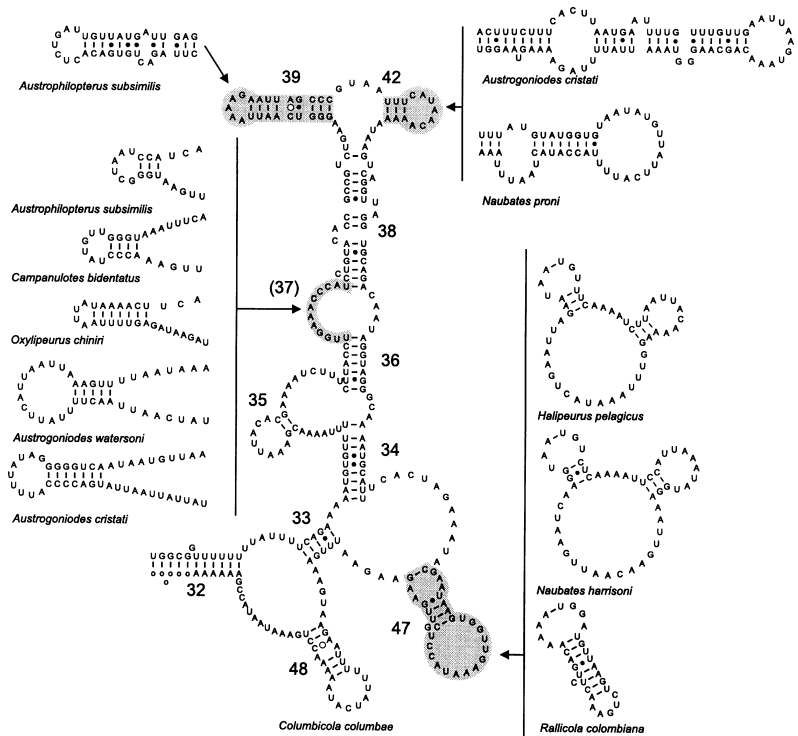
Figure 2. Secondary structures for domain III for mitochondrial 12S rRNA for selected lice species. The complete structure for *Columbicola columbae* is shown, with regions that vary in other lice highlighted.

secondary structure for domain III 12S rRNA for the pigeon louse *Columbicola columbae* (Fig. 2) closely resembles that found in most insects (Fig. 1). Compared to the Page (2000b) model, helices 39 and 40 in lice cannot be distinguished, and so are drawn here as a single helix. Of the 25 lice sequences we studied, most closely resemble the structure for *Columbicola*. However, some taxa show notable differences (Fig. 2). Five species in four different genera show an additional helix between 36 and 38 that we refer to as helix 37 (following van de Peer *et al.*, 1998). Two taxa have a much extended helix 42. Inferring secondary structure in the region between helices 34′ and 33′ is not easy because of difficulties in satisfactorily aligning their sequences. However, *Halipeurus pelagicus* and *Naubates harrisoni* show clear evidence for an additional stem on the side of helix 47. Support for this stem comes from comparisons with sequences for other species of the genus (Paterson *et al.*, 2000), which show compensating changes with respect to the sequences presented here.

Length variation in insects

Figure 3 shows the variation in sequence length between some key landmarks in 12S rRNA domain III. For the 225 insect sequences used by Page (2000b), the range of lengths for the sub-sequence starting at helix 33 (segment 1 in Fig. 1) is 213–259 bp, with a mean of 225 (SD 5.7). Although most insect sequences are within a narrow 23 bp range of sizes (Fig. 3), there are some notable exceptions.

In almost all insects the region between helices 39′ and 38′ comprises a 16–27 bp stretch which includes the short 3–4 bp helix 42. The beetle *Caryedon immaculatum* (GENBANK accession no. AF004121) and the locust *Ruspolia nitidula* (Z97602) both have large (21–27 bp) insertions in this region. The cockroaches *Nyctibora azteca* (U17795) and *N. lutzi* (U17801) have a 15–20 bp insertion between helices 38′ and 36′, which in other insects is a short 3 or 4 bp loop.

Length variation in lice

For the 25 sequences in the louse alignment, the length between the landmarks at the base of helix 33 ranged from 205 to 302 bp, with a mean of 235.1 (SD 25.9). While many sequences are of similar length to other insects (Fig. 3), numerous sequences have large insertions in one or more of three different locations. In lice, the region between helices 36 and 38 comprises anywhere from 4 to 45 bases and appears to form a helix in some taxa (Fig. 2), whereas in other insects this region is a simple bulge of 3–15 bases. The region occupied by helix 42 is variable in lice, and in some *Naubates* and *Austrogonoides* species this helix can be extended from a small three-base helix to a much larger structure (Fig. 2). The final region of variability spans the region between helices 34′ and 33′, which in most insects shows little length variation (49–58 bp, mean 53, SD 1.9), whereas in lice there may be anything from 33 to 78 bases in this region. Most lice fall within the range of
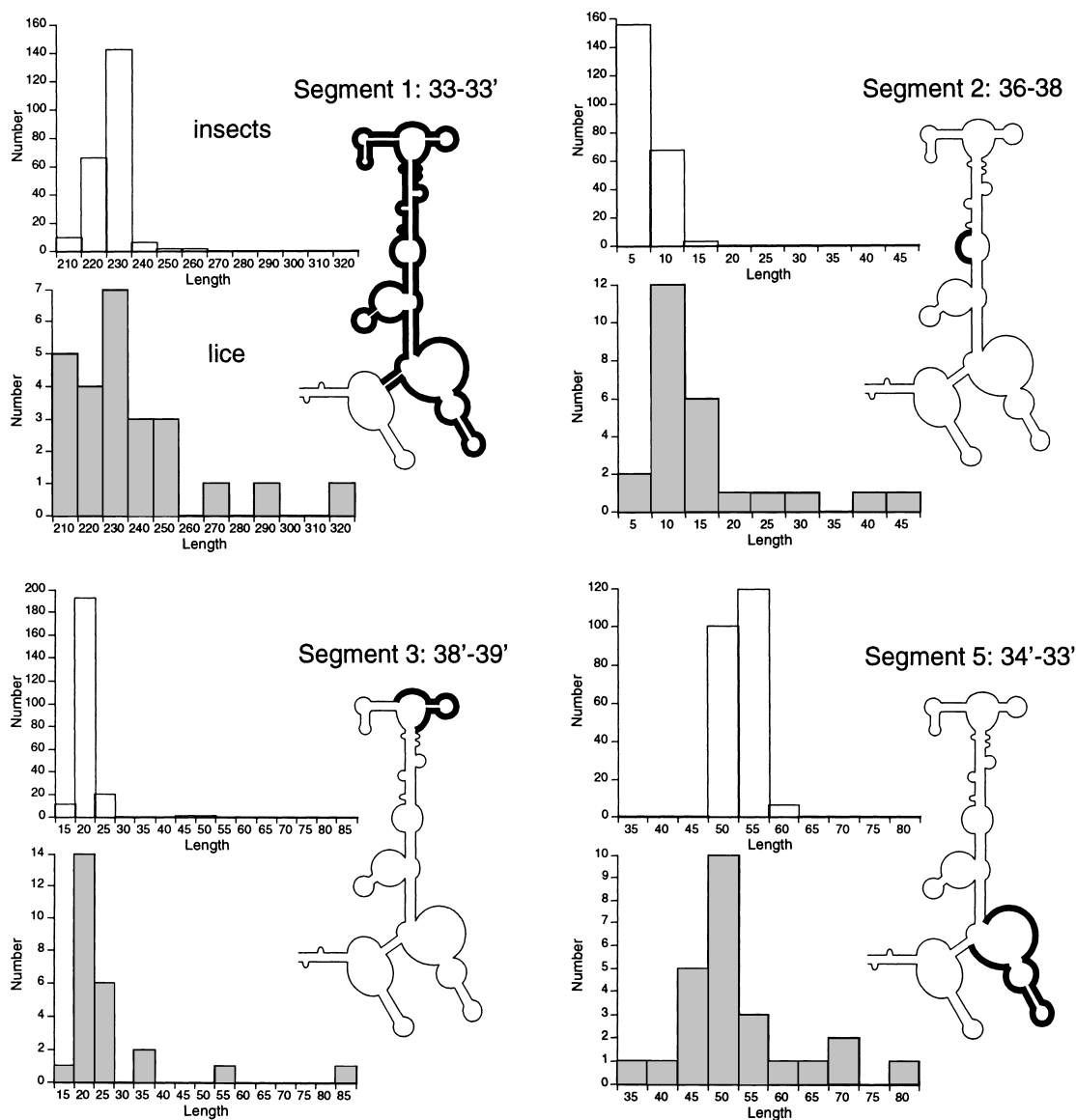
Figure 3. Distribution of lengths in base pairs for four regions of 12S rRNA (numbered as in Fig. 1) for 225 insect and 25 louse sequences. Note the greater variation in lice compared to other insects.

other insects (Fig. 3), but lice from the *Philoceanus* complex (*Halipeurus* and *Naubates*) have a 10–20 bp insertion that can be folded to form an additional helix (Fig. 2).

Structural variation

Both lice and other insects show a wide range of tree edit distances for 12S rRNA secondary structure (Fig. 4). For the 19 exemplar insects, the mean tree edit distance is 35.7 (SD 15.8) whereas for the 25 lice it is 46.8 (SD 21.1). Many of the extreme values in Fig. 4 involve the penguin louse *Austrogoniodes cristati*, which has a helix 37 and a much-elongated helix 42 (Fig. 2). After removing this louse sequence, the mean edit distance for louse structures drops to 41.0 (SD 14.8).

Phylogenetic informativeness

A comparison between Bayesian trees for louse sequences and the neighbour-joining tree for RNA structures suggests that the secondary structure contains limited phylogenetic information (neighbour-joining trees for the sequence data are very similar to the Bayesian trees and are not shown). Although the trees for lice based on different genes are all different (Fig. 5, Table 2), they are more similar to each other than any are to the secondary structure tree. Indeed, whether we compare the RNA structure tree with individual gene trees, non-RNA-based trees, or the tree from all three genes combined, it makes little difference to the comparison. However the RNA structure tree is more similar to the gene trees than
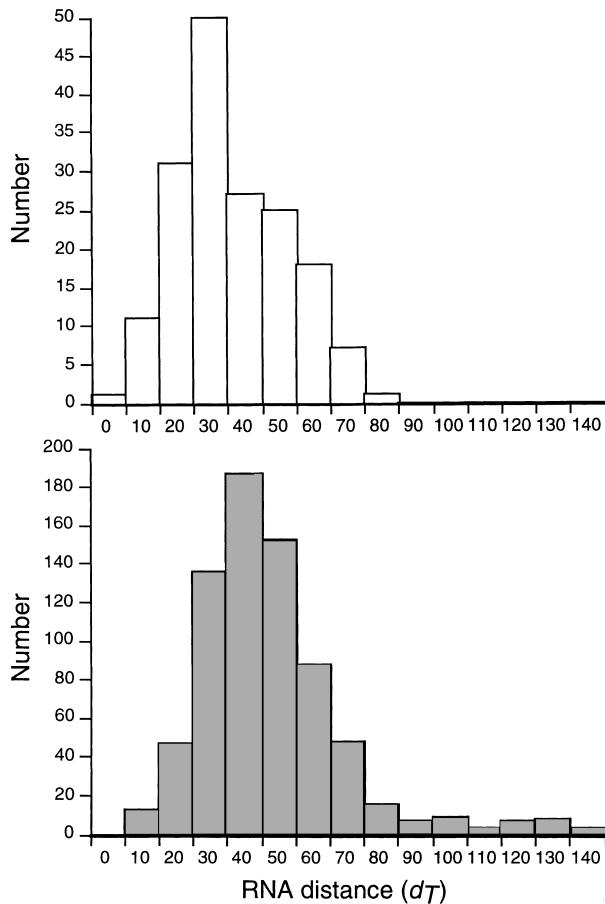
Figure 4. Distribution of pair-wise tree edit distances between 12S rRNA secondary structures for 19 insects and 25 lice (Table 1).

as the grouping of the duck lice *Anaticola* and *Anatoecus*, and the petrel lice *Halipeurus* and *Naubates*, the latter sharing an extra stem in helix 47 (Fig. 2). However, other groupings in the RNA structure tree, such as *Austrogonioides* and *Campanulotes* with *Oxylipeurus* do not occur in any of the sequence trees. This later grouping reflects the presence of helix 37 (Fig. 2), which all three lice share. If there was support for this grouping in the sequence data, then we would expect find this group in at least some of the trees sampled from the Monte Carlo chain in the Bayesian analysis of the sequence data. Given that none of the 5000 trees sampled contained a *Austrogonioides* + *Campanulotes* + *Oxylipeurus* clade, we can be confident that there is no evidence from sequence data for this grouping. Hence, helix 37 has evolved more than once in lice and its presence or absence cannot be used as a reliable phylogenetic marker.

Discussion

The difficulties Paterson *et al.* (2000) experienced when aligning louse 12S rRNA are readily understandable. Within Ischnocera there is a considerable variation in both sequence length and secondary structure. Indeed, the variation shown in louse secondary structures is greater than that depicted by Hickson *et al.* (1996) in their survey of animal 12S rRNA (a quantitative comparison is hampered because of differences between Hickson *et al.*'s model and that used here, see Page, 2000b, for details). The variation in length exceeds that found in other insects. When measured using tree edit distances, there is as much, if not more, variation within lice than across all insect orders.

It would be interesting to see whether the other mitochondrial ribosomal gene (16S rRNA) shows an increased structural variation. To date, the only published 16S rRNA sequence from a louse is for the amblyceran *Heterodoxus macropus* (Shao *et al.*, 2001). This sequence appears
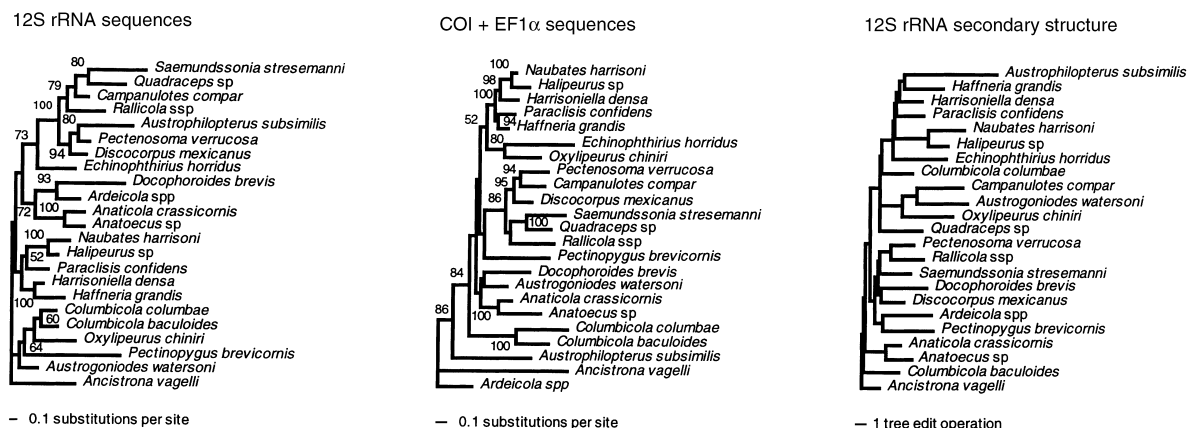
would expected due to chance, based on the distribution of pair-wise distances between 1000 random trees ($P < 0.001$).

Inspecting the trees (Fig. 5) shows that there are some similarities between the sequence and structure trees, such

Figure 5. Trees for the 23 lice in Table 1 constructed by Bayesian analysis of nucleotide sequences (12S rRNA, and combined EF1α and COI sequences), and neighbour-joining analysis of tree edit distance between 12S rRNA secondary structures. Numbers on the internal branches of the sequence trees are the posterior probability values (Huelsenbeck & Ronquist, 2001), where they are greater than 50%.

Table 2. Pair-wise distances between trees for lice constructed from 12S rRNA, COI and EF1α sequences (either separately or combined) and from RNA tree edit distances. Lower triangle is the DC measure, upper triangle is SJA (Day, 1986)

	12S rRNA + EF1α + COI	COI	EF1α	EF1α + COI	12S rRNA	RNA edit distance
12S rRNA + EF1α + COI	–	0.218	0.114	0.077	0.290	0.517
COI	45	–	0.238	0.238	0.460	0.553
Ef1α	53	66	–	0.040	0.392	0.519
EF1α + COI	38	52	21	–	0.361	0.507
12S rRNA	124	118	155	157	–	0.616
RNA edit distance	476	372	453	476	463	–

completely consistent with Buckley *et al.*'s (2000) model for domains IV and V of insect 16S rRNA secondary structure, and shows no length variation beyond that found in other insects.

The diversity of 12S rRNA structures found in lice may be symptomatic of an elevated rate of evolution in louse mitochondrial genomes. Studies of the mitochondrial protein genes cytochrome oxidase I (COI) and cytochrome b (cyt *b*) (Hafner *et al.*, 1994; Page *et al.*, 1998) have shown that these genes evolve 2–3-fold more rapidly in lice than in their vertebrate hosts. Page *et al.* (1998) found that louse cyt *b* showed elevated levels of amino acid replacements with respect to other insects. The only complete mitochondrial genome for a louse sequenced to date (Shao *et al.*, 2001) shows numerous gene rearrangements with respect to other insects, resulting in a gene order unlike any other animal. Our own studies have shown that the relative rate of substitution in mitochondrial and nuclear genes is an order of magnitude greater in lice than in other insects (unpublished observations). Taken together, these studies suggest that lice are indeed highly unusual. It would be very interesting to know whether the novel gene order, rRNA structures and elevated substitution rates in louse mitochondria all evolved in concert with the lice becoming parasitic, or whether this is specific to particular louse clades. To date, we have only found an RNA structural variation in the Ischnocera; the single amblyceran and anopluran sequences available display a typical insect structure.

The poor match between louse phylogenies based on primary sequence and secondary structure suggests that the latter is of limited value in inferring louse phylogenies, in contrast to its utility in other organisms (e.g. Billoud *et al.*, 2000; Collins *et al.*, 2000). Although some structures do seem to be correlated with clades, such as the extra stem in helix 47 which is shared by the closely related *Halipeurus* and *Naubates* (Fig. 2), some striking features such as helix 37 appear to have evolved repeatedly in different clades.

The variation in both primary sequence and secondary structure such as those shown by the louse sequences makes a multiple sequence alignment difficult. Even if secondary structure can be inferred with confidence, the alignment of loops (and in some cases stems) can be problematic. The problem of regions that are difficult to

align has received considerable attention (e.g. Wheeler, 1994), with most efforts directed at automated methods of aligning the primary sequence, which are seen as more objective than manual alignment (Phillips *et al.*, 2000). More recently, Wheeler (1999) and Lutzoni *et al.* (2000) have described an approach which abandons trying to create a complete multiple alignment for all sequences. Fixed-state optimization treats strings of nucleotides as character states. The cost of transformation between these states is an edit cost (the number of operations required to transform one sequence into another). Although in existing implementations the edit costs are based on primary structure, the approach could be generalized to include measures of secondary structure difference (Hofacker *et al.*, 1994; Moulton *et al.*, 2000). Hence, rather than opposing automatic methods such as optimization alignment and fixed-state optimization with secondary structure alignment, these approaches might be usefully merged into a single automated method for aligning variable RNA sequences.

Experimental procedures

Sequences

Total genomic DNA was extracted from single lice using the DNAeasy Tissue Kit (Qiagen). The third domain of the 12S rRNA gene was amplified and sequenced using the insect specific primers 12Sai and 12Sbi (Simon *et al.*, 1994). Amplification products were gel purified using the QIAquick Gel Extraction Kit (Qiagen) or the Qiagen PCR Purification Kit, and sequenced using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase, FS (Perkin Elmer). The sequencing products were ethanol precipitated and run on an ABI 373 or 377 Stretch automated sequencing machine. Previously published 12S rRNA for seabird lice (Paterson *et al.*, 2000) were retrieved from GENBANK. The sequences used in this study are listed in Table 1, and (with the exception of *Austrogonioides cristati* and *Naubates prioni*) represent those 12S rRNA sequences for which EF1α and COI sequences are also available (Cruickshank *et al.*, 2001; Johnson & Clayton, 2002; Johnson *et al.*, 2001a,b, 2002).

Insect 12S rRNA secondary structure

We used as our point of reference the 225 insect 12S rRNA sequence alignment employed by Page (2000b) to infer a general model of insect 12S rRNA secondary structure (this alignment is available from the EMBL WEBALIGN database

<http://www3.ebi.ac.uk/Services/align/listali.html> as alignment DS43718). The original alignment was constructed using CLUSTALX (Thompson *et al.*, 1997) without reference to secondary structure, and contained some regions that – in light of the resulting secondary structure model – are obviously misaligned. In the present study we manually edited regions of poor quality (Page, 2000b, fig. 7) in the 225 sequence alignment to improve the fit to the secondary structure model.

Length variation

As a simple measure of secondary structure conservation, we counted the number of bases (ignoring gaps) between key 'landmarks' in the secondary structure model, such as the start and end of helix 33 (Fig. 1). Variation in length amongst different species will reflect insertion and deletion events, and varying lengths of stems and loops. We measured length variation for the 225 insect sequences, and the louse sequences. For some of the 225 insect sequences there is uncertainty about the secondary structure of the more variable helices (such as 39, 40, 42 and 47), hence these landmarks were located on the core conserved helices whose structure has been firmly established.

Quantifying secondary structure differences

To quantify the difference between secondary structures we used the RNADISTANCE program in the Vienna RNA package (Hofacker *et al.*, 1994). This program computes a tree edit distance, d_T, between two RNA structures. This measure counts the minimum number of insertions and deletions of paired and unpaired bases needed to transform one RNA structure into another (Fontana *et al.*, 1993; Moulton *et al.*, 2000). Although the louse 12S rRNA sequences were all obtained with the same primers, our ability to read the 5′ and 3′ ends of the sequences varied among taxa. Hence, when computing edit distances the sequences were pruned to include only the structure rooted at helix 33 (segment 1 in Fig. 1). To assess the variation among major insects groups, we used a subset of 15 sequences for which we were confident in the secondary structure of both the core helices and the more variable regions. We also added four taxa with the most divergent secondary structures to ensure that the full range of variation in insect structures was represented. The taxa and sequences used are the dipterans *Drosophila yakuba* (X03240), *Diplonevra nitidula* (AF126298), and *Anopheles gambiae* (L20934); the butterfly *Aglais urticae* (AF232882); the honey bee *Apis mellifera* (L06178); the beetles *Tachinus luridus* (AF021047), *Caryedon acaciae* (AF004114), *Caryedon immaculatum* (AF004121), *Molops piceus* (AF190021); the bugs *Panstrongylus megistus* (AF021178) and *Dalbulus charlesi* (AF051276); the cicada *Magicicada cassini* (X97149); the cockroaches *Archimandrita tessellata* (U17762) and *Nyctibora azteca* (U17795); the termite *Hodotermopsis japonica* (AB006580); the grasshoppers *Gomphoceripus rufus* (Z93247) and *Ruspolia nitidula* (Z97602); the damselfly *Ischnura barberi* (AF067703), and the silverfish *Ctenolepisma longicaudata* (L02381).

Louse secondary structure

Initial secondary structures were computed using the program RNALIGN (Corpet & Michot, 1994; ftp://ftp.toulouse.inra.fr/pub/rnalign/). This program takes a sequence and aligns it to a reference alignment and secondary structure model using both primary and secondary structure simultaneously. Page (2000b) constructed a server <http://taxonomy.zoology.gla.ac.uk/cgi-bin/rna.cgi> that aligns a user-supplied sequence to a reference alignment of five insect sequences and the Page (2000b) secondary structure. The server returns a secondary structure for the user's sequence, and an alignment of that sequence to the five reference sequences.

Where necessary, we adjusted the louse alignments and structures found by RNAlign, either manually or using CLUSTALW, taking in to account evidence for compensating mutations (Gutell *et al.*, 1992). In cases where there were large insertions relative to the general insect model, possible secondary structures for these insertions were obtained using thermal folding as implemented in the program RNADRAW (Matzura & Wennborg, 1996). This technique can be used to suggest secondary structures, but is not as reliable as comparative techniques (Konings & Gutell, 1995). Secondary structures were drawn using RNAVIZ 1.0 (de Rijk & de Wachter, 1997).

Phylogenetic tree construction and comparison

Higher level louse phylogeny is currently uncertain (Cruickshank *et al.*, 2001; Johnson & Whiting, 2002), so we constructed trees for the 23 of the 25 louse taxa in Table 1 for which we have sequences from the 12S rRNA, elongation factor 1α (EF1α), and mitochondrial cytochrome oxidase I (COI) genes. Trees were computed separately for each gene, all three genes combined, and for just the EF1α and COI sequences (to ensure complete independence from the 12S rRNA sequences). Bayesian analysis was performed using MRBAYES (Huelsenbeck & Ronquist, 2001) with the following settings. The maximum likelihood model employed six substitution types ('nst = 6'), with base frequencies set to the empirically observed values ('basefreq = empirical'). Rate variation across sites was modelled using a gamma distribution, with a proportion of sites being invariant ('rates = invgamma'). The Markov chain Monte Carlo search was run with four chains for 500 000 generations, with trees being sampled every 100 generations (the first 1000 trees were discarded as 'burnin'). Support for individual groups of taxa was determined by finding the frequency of that group in the set of trees sampled from the Monte Carlo chain. This is the posterior probability or 'credibility' of that group. Neighbour-joining trees were also computed from LOGDET (Lockhart *et al.*, 1994) sequence distances using in PAUP* (Swofford, 2001).

A tree of RNA secondary structures was constructed from tree edit distances (see above) using neighbour joining. The similarity between the sequence and structure trees was computed using triplet tree comparison metrics (Critchlow *et al.*, 1996; Day, 1986). Given a pair of trees, these measures break the two trees up into all possible sets of rooted trees ('triplets'). There is a family of tree distance measures based on counting how many triplets are resolved in each tree, and whether they are resolved the same way. For each pair of trees, two measures were computed: DC, the number of resolved triplets which are different in the two trees; and SJA, the number of strict joint assertions, which is DC normalized by dividing by the number of triplets resolved in the two trees. These measures are the rooted equivalents of the measures for quartets described by Day (1986). The value of SJA ranges from 0 (trees have no triplets in common) to 1 (all triplets are identical). Null distributions for these measures were obtained by computing the distances between 1000 pairs of random trees using COMPONENT (Page, 1993).

References

Billoud, B., Guerrucci, M.-A., Masselot, M. and Deutsch, J.S. (2000) Cirripede phylogeny using a novel approach: molecular morphometrics. *Mol Biol Evol* **17**: 1435–1445.

Buckley, T.R., Simon, C., Flook, P.K. and Misof, B. (2000) Secondary structure and conserved motifs of the frequently sequenced domians IV and V of the insect mitochondrial large subunit rRNA gene. *Insect Mol Biol* **9**: 565–580.

Collins, L.J., Moulton, V. and Penny, D. (2000) Use of RNA secondary structure for studying the evolution of RNase P and RNase MRP. *J Mol Evol* **51**: 194–204.

Corpet, F. and Michot, B. (1994) RNAlign program: alignment of RNA sequences using both primary and secondary structures. *CABIOS* **10**: 389–399.

Critchlow, D.E., Pearl, D.K. and Qian, C. (1996) The triples distance for rooted bifurcating phylogenetic trees. *Syst Biol* **45**: 323–334.

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

Day, W.H.E. (1986) Analysis of quartet dissimilarity measures between undirected phylogenetic trees. *Syst Zool* **35**: 325–333.

Fontana, W., Konings, D.A.M., Stadler, P.F. and Schuster, P. (1993) Statistics of RNA secondary structures. *Biopolymers* **33**: 1389–1404.

Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J. and Stormo, G.D. (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucl Acids Res* **20**: 5785–5795.

Hafner, M.S. and Nadler, S.A. (1990) Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation. *Syst Zool* **39**: 192–204.

Hafner, M.S., Sudman, P.D., Villablanca, F.X., Spradling, T.A., Demastes, J.W. and Nadler, S.A. (1994) Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265**: 1087–1090.

Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J. and Penny, D. (1996) Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol Biol Evol* **13**: 150–169.

Hickson, R.E., Simon, C. and Perrey, S.W. (2000) The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. *Mol Biol Evol* **17**: 530–539.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M. and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh Chem* **125**: 167–188.

Houde, P., Cooper, A., Leslie, E., Strand, A.E. and Montaño, G.A. (1997) Phylogeny and evolution of 12S rDNA in Gruiformes (Aves). In *Avian Molecular Evolution and Systematics* (Mindell, D.P., ed.), pp. 121–158. Academic Press, San Diego.

Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Johnson, K.P., Adams, R.J. and Clayton, D.H. (2001a) Molecular systematics of the Goniodidae (Insecta: Phthiraptera). *J Parasitol* **87**: 862–869.

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

Johnson, K.P., Moyle, R.G., Witt, C.C., Faucett, R.C. and Weckstein, J.D. (2001b) Phylogenetic relationships in the louse genus *Penenirmus* based on nuclear (EF-1α) and mitochondrial (COI) DNA sequences. *Syst Entomol* **26**: 491–497.

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

Johnson, K.P., Williams, B.L., Drown, D.M., Adams, R.J. and Clayton, D.H. (2002) The population genetics of host specificity: genetic differentiation in dove lice. *Mol Ecol* **11**: 25–38.

Kjer, K.M. (1995) Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Mol Phylog Evol* **4**: 314–330.

Konings, D.A.M. and Gutell, R.R. (1995) A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. *RNA* **1**: 559–574.

Lenhof, H.-P., Reinert, K. and Vingron, M. (1998) A polyhedral approach to RNA sequence structure alignment. *J Comput Biol* **5**: 517–530.

Lockhart, P.J., Steel, M.A., Hendy, M.D. and Penny, D. (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol* **11**: 605–612.

Lutzoni, F., Wagner, P., Reeb, V. and Zoller, S. (2000) Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating position homology. *Syst Biol* **49**: 628–651.

Matzura, O. and Wennborg, A. (1996) *Rnadraw: an Integrated Program for RNA Secondary Structure Calculation and Analysis Under 32-Bit Microsoft Windows*. CABIOS. 12, 247–249.

Mindell, D.P., Sorenson, M.D., Huddleston, C.J., Miranda, H.C., Knight, A., Sawchuck, S.J. and Yuri, T. (1997) Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In *Avian Molecular Evolution and Systematics* (Mindell, D.P., ed.), pp. 213–247. Academic Press, San Diego.

Moran, N.A., van Dohlen, C.D. and Baumann, P. (1995) Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *J Mol Evol* **41**: 727–731.

Moulton, V., Zuker, M., Steel, M., Pointon, R. and Penny, D. (2000) Metrics on RNA secondary structures. *J Comput Biol* **7**: 277–292.

Notredame, C., O'Brien, E.A. and Higgins, D.G. (1997) RAGA: RNA sequence alignment by genetic algorithm. *Nucl Acids Res* **25**: 4570–4580.

Page, R.D.M. (1993) *COMPONENT, Tree Comparison Software for Microsoft® Windows™*. The Natural History Museum, London.

Page, R.D.M. (2000a) Circles: automating the comparative analysis of RNA secondary structure. *Bioinformatics* **16**: 1042–1043.

Page, R.D.M. (2000b) Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. *Nucl Acids Res* **28**: 3839–3845.

Page, R.D.M., ed. (2002) *Tangled Trees: Phylogeny, Cospeciation and Coevolution*. University of Chicago Press, Chicago.

Page, R.D.M. and Hafner, M.S. (1996) Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. In: *New Uses for New Phylogenies* (Harvey, P.H., Leigh Brown, A.J., Maynard Smith, J. and Nee, S., eds), pp. 255–270. Oxford University Press, Oxford.

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

Paterson, A.M. and Banks, J. (2001) Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly. *Int J Parasitol* **31**: 1012–1022.

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

van de Peer, Y., Caers, A., de Rijk, P. and de Wachter, R. (1998) Database on the structure of small ribosomal subunit RNA. *Nucl Acids Res* **26**: 179–182.

Phillips, A., Janies, D. and Wheeler, W. (2000) Multiple sequence alignment in phylogenetics. *Mol Phylog Evol* **16**: 317–330.

de Rijk, P. and de Wachter, R. (1997) RnaViz, a program for the visualisation of RNA secondary structure. *Nucl Acids Res* **25**: 4679–4684.

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

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am* **87**: 651–704.

Springer, M.S. and Douzery, E. (1996) Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J Mol Evol* **43**: 357–373.

Stoye, J., Moulton, V. and Dress, A.W.M. (1997) DCA: an efficient implementation of the Divide-and-Conquer multiple sequence alignment algorithm. *CABIOS* **13**: 625–626.

Swofford, D.L. (2001) *Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Tabaska, J.E., Cary, R.E., Gabow, H.N. and Stormo, G.D. (1998) An RNA folding method capable of identifying pseudoknots and base triples. *Bioinformatics* **14**: 691–699.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25**: 4876–4882.

Wheeler, W.C. (1994) Sources of ambiguity in nucleic acid sequence alignment. In: *Molecular Ecology and Evolution: Approaches and Applications* (Schierwater, B., Striet, B., Wagner, G.P. and DeSalle, R., eds), pp. 323–352. Birkhäuser-Verlag, Basel.

Wheeler, W. (1999) Fixed character states and the optimization of molecular sequence data. *Cladistics* **15**: 379–385.