Technical Reports in Taxonomy 02-01

taxonomy.zoology.gla.ac.uk/publications/tech-rep/

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

Roderic D. M. Page[1], Robert Cruickshank[1], and Kevin P. Johnson[2]

1 Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK

2 Illinois Natural History Survey, 607 East Peabody Drive, Champaign, IL 61820-6970, USA

Address for correspondence:
Dr Roderic D. M. Page
DEEB, IBLS
Graham Kerr Building
University of Glasgow
Glasgow G12 8QQ, UK
Tel: +44 141 330 4778
Fax: +44 141 330 2972
Email: r.page@bio.gla.ac.uk

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 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

rRNA secondary structure; multiple alignment; 12S rRNA; insects; lice; tree comparison metrics

Introduction

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

The prevalence of cospeciation between lice and their hosts has resulted in these insects playing a key role in recent methodological (Hafner and Page 1995; Huelsenbeck et al. 1997; Page 1996) and empirical (Hafner et al. 1994; Page et al. 1998; Paterson et al. 2000) 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 and 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 secondary structure requires considerable manual effort. Concern about the potential for subjectivity in such alignments has motivated 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 correctly align RNA sequences (Hickson et al. 2000), and methods for automatically aligning sequences using both primary and secondary structure are in their infancy (Corpet and 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) 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 3end 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 (2000) constructed a model based on an alignment of 225 insect sequences. In the present study we use Page's (2000) model (Fig. 1) as a benchmark for determining the degree of variation in 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).

Methods

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). 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 *Austrogoniodes cristati* and *Naubates prioni*) represent those 12S rRNA sequences for which EF1α and COI sequences are also available (Cruickshank et al. 2001; Johnson et al. in review).

Insect secondary structure

We used as our point of reference the 225 insect 12S rRNA sequence alignment employed by Page (2000) 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 2000, 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 landmarks were located on the core conserved helices whose structure is 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 structures, which is the minimum number of operations needed to transform one structure into the other. 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 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 *Gomphocerippus 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 and 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 (2000) 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 (2000) 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 and Wennborg 1996). This technique can be used to suggest

secondary structures, but is not as reliable as the comparative techniques (Konings and Gutell 1995). Secondary structures were drawn using RnaViz 1.0 (de Rijk and de Wachter 1997).

Phylogenetic information

Higher level louse phylogeny is currently uncertain (Cruickshank et al. 2001), so we constructed trees for the 23 of the 25 louse taxa in Table 1 for which we have sequences from the12S rRNA, elongation factor 1α (EF1α), and mitochondrial cytochrome oxidase I (COI) genes. Trees were computed for each gene separately, all three genes combined, and for just the EF1α and COI sequences (to ensure complete independence from the 12S rRNA sequences). LogDet (Lockhart et al. 1994) sequence distances were computed and trees constructed using neighbor-joining in PAUP* (Swofford 2001). A tree of RNA secondary structures was constructed from tree edit distances (see above) using neighbor joining. Trees were compared using the partition metric (Robinson and Foulds 1981) and the symmetric difference metric for triplets (the rooted analogue of the quartet symmetric difference Day 1986). Null distributions for these two measures were computed by computing the distances between 1000 pairs of random trees. All tree comparisons were computed using COMPONENT (Page 1993).

Results

Louse secondary structure

Secondary structures were inferred for each sequence listed in Table 1. The key features of these structures are summarised in Fig. 2. Drawings of structure for each individual sequence are available online

(http://130.209.46.190/cfdocs/exampleapp/lousebase/12s/). The inferred 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 (2000) model, in lice helices 39 and 40 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 of between helices 34' and 33' is not easy because of difficulties in satisfactorily aligning the 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

Fig. 3 shows the variation in sequence length between some key landmarks in 12S rRNA domain III. For the 225 insect sequences used by Page (2000) the range in length for the subsequence starting at helix 33 (segment 1 in Fig. 1) is 213-259 bp, with a mean of 225 (s.d. 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 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 (s.d. 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 3 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, s.d. 1.9), whereas in lice there may be anything from 33 to 78 bases in this region. Most lice fall within the range of 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 (S.D. 15.8) whereas for the 25 lice it is 46.8 (S.D. 21.1). Many of the extreme values in Fig. 4 involve the penguin louse *Austrogoniodes cristati*, which has 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 (S.D. 14.8).

Phylogenetic informativeness

Comparison between neighbor-joining trees for louse sequences and for RNA structures suggests that secondary structure contains limited phylogenetic information. Although 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 expected due to chance, based on the distribution of pairwise distances between 1000 random trees (partition metric $p < 0.001$, triplet symmetric difference $p < 0.01$). Inspecting the trees (Fig. 5) shows that there are some similarities between the sequence and structure trees, such as the grouping of the duck lice *Anaticola* and *Anatoecus*, and the petrel lice *Halipeurus* and *Naubates*. However, some groupings, such as *Austrogonioides* and *Campanulotes* with *Oxylipeurus* receive no support from the sequence data; rather they reflect the presence of helix 37 (Fig. 2), which appears to have evolved independently in each lineage.

Electronic availability of data

The sequences in Table 1 are available in GenBank. The alignments used to infer secondary structures are available from http://taxonomy.zoology.gla.ac.uk/rod/data/louse12S/**.**

Discussion

The difficulties Paterson et al. (2000) experienced when aligning louse 12S rRNA are readily understandable. Within Ischnocera there is 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 (quantitative comparison is hampered because of differences between Hickson et al.'s model and that used here, see Page 2000, 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.

The diversity of 12S rRNA structures found in lice may be symptomatic of 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 in lice these genes evolve 2-3 times more rapidly 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 (Johnson et al. in review). 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 it is specific to particular louse clades. To date we have only found structural variation in the Ischnocera; the single amblyceran and anopluran sequences available display 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 correlated with clades, such as the extra stem in helix 47 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 pimary sequence and secondary structure such as those shown by the louse sequences makes multiple sequence alignment difficult. Even if secondary structure can be inferred with confidence, alignment of loops (and in some cases stems) can be problematic. The problem of difficult to align regions has received considerable attention (e.g., Wheeler 1994), with most efforts directed at automated methods of aligning primary sequence, which are portrayed as more objective than manual alignment (Phillips et al. 2000). More recently, Wheeler (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 optimisation 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 in to another). Although in existing implementations the edit costs are based on primary structure, the approach could be generalised to include measures of secondary structure difference (Hofacker et al. 1994; Moulton et al. 2000). Hence rather than opposing automatic methods such as optimisation alignment

and fixed-state optimisation with secondary structure alignment, these approaches might be usefully merged into a single automated method for aligning variable RNA sequences.

Acknowledgements

This work was supported by NERC grant GR3/11075 to R.D.M.P, Richard Griffiths, and Robert Furness, and by NSF-CAREER award DEB-9703003 to Dale H. Clayton. We thank the Wolfson Foundation, and the University of Utah DNA Sequencing Facility, supported in part by NCI grant #5p30CA42014. D. H. Clayton, D. Drown, R. Furness, A. Patserson, and P. Thompson provided assistance in collecting louse specimens. Florence Corpet provided help with RNAlign. For comments on various versions of the manuscript we thank Martyn Kennedy and Vince Smith.

References

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

Buckley TR, Simon C, Flook PK, 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 LJ, Moulton V, 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, Michot B (1994) RNAlign program: alignment of RNA sequences using both primary and secondary structures. Comput Applic Biosci 10:389-399

Cruickshank RH, Johnson KP, Smith VS, Adams RJ, Clayton DH, Page RDM (2001) Phylogenetic analysis of elongation factor 1α identifies major groups of lice (Insecta: Phthiraptera). Mol Phylog Evol 19:202-215

Day WHE (1986) Analysis of quartet dissimilarity measures between undirected phylogenetic trees. Syst Zool 35:325-33

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

Gutell RR, Power A, Hertz GZ, Putz EJ, Stormo GD (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res 20:5785-5795

Hafner MS, Nadler SA (1990) Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation. Syst Zool 39:192-204

Hafner MS, Page RDM (1995) Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system. Philos Trans R Soc, Lond 349:77-83

Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA (1994) Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265:1087-90

Hickson RE, Simon C, Cooper A, Spicer GS, Sullivan J, 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 RE, Simon C, Perrey SW (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 IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. Monatsh Chem 125:167-188

Houde P, Cooper A, Leslie E, Strand AE, Montaño GA (1997) Phylogeny and evolution of 12S rDNA in Gruiformes (Aves). In: Mindell DP (ed) Avian molecular evolution and systematics. Academic Press, San Diego, p 121-158

Huelsenbeck JP, Rannala B, Yang Z (1997) Statistical tests of host-parasite cospeciation. Evolution 51:410-419

Johnson KP, Cruickshank RH, Adams RJ, Smith VS, Page RDM, Clayton DH (in review) Elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). Mol Biol Evol

Kjer KM (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 DAM, Gutell RR (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, Vingron M (1998) A polyhedral approach to RNA sequence structure alignment. J Comput Biol 5:517-530

Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. Mol Biol Evol 11:605-12

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

Matzura O, Wennborg A (1996) RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Comput Applic Biosci 12:247-249

Mindell DP, Sorenson MD, Huddleston CJ, Miranda HC, Knight A, Sawchuck S, J., Yuri T (1997) Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In: Mindell DP (ed) Avian molecular evolution and systematics. Academic Press, San Diego, p 213-247

Moran NA, van Dohlen CD, Baumann P (1995) Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. J Mol Evol 41:727-31

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

Notredame C, O'Brien EA, Higgins DG (1997) RAGA: RNA sequence alignment by genetic algorithm. Nucleic Acids Res 25:4570-4580

Page RDM (1993) COMPONENT, Tree comparison software for Microsoft® Windows™. The Natural History Museum, London

Page RDM (1996) Temporal congruence revisited: Comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice. Syst Biol 45:151-67

Page RDM (2000) Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. Nucleic Acids Res 28:3839-3845

Page RDM, Hafner MS (1996) Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. In: Harvey PH, Leigh Brown AJ, Maynard Smith J, Nee S (eds) New uses for new phylogenies. Oxford University Press, Oxford, p 255-70

Page RDM, Lee PLM, Becher SA, Griffiths R, Clayton DH (1998) A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. Mol Phylog Evol 9:276-293

Paterson AM, Wallis GP, Wallis LJ, Gray RD (2000) Seabird and louse coevolution: complex histories revealed by 12S rRNA sequences and reconciliation analysis. Syst Biol 49:383-399

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

Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Math Biosci 53:131-147

Shao R, Campbell NJH, Barker SC (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, 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 MS, 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, Dress AWM (1997) DCA: an efficient implementation of the Divide-and-Conquer multiple sequence alignment algorithm. Comput Applic Biosci 13:625-626

Swofford DL (2001) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

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

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

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

Wheeler WC (1994) Sources of ambiguity in nucleic acid sequence alignment. In: Schierwater B, Striet B, Wagner GP, DeSalle R (eds) Molecular ecology and evolution: approaches and applications. Birkhäuser Verlag, Basel, p 323-352

Tables

Table 1. Taxa and sequences used in this study. EF1α sequences from Cruickshank et al (2001), COI sequences from Johnson et al.(in review).

Louse	Host	12SrRNA	COI	EF1a
Anaticola crassicornis	*Anas platyrynchos* (Mallard)	XXXXXXX	XXXXXXX	AF320354
Anatoecus	*Cygnus olor* (Mute Swan)	XXXXXXX	XXXXXXX	AF320356
Ancistrona vagelli	*Puffinus tenuirostris* (Short-tailed Shearwater)	AF189128	XXXXXXX	AF320358
Ardeicola geronticorum	*Geronticus calvus* (Bald Ibis)	XXXXXXX	XXXXXXX	AF320361
Austrogoniodes watersoni	*Eudyptula minor* (Little Penguin	AF189129	XXXXXXX	AF320362
Austrophilopterus subsimilis	*Ramphastos sulfuratus* (Keel-billed Toucan	AF189130	XXXXXXX	AF320365
Campanulotes compar	*Columba livia* (Feral Pigeon)	AF189131	AF278670	AF320377
Columbicola baculoides	*Zenaida macroura* (Mourning Dove)	AF190425	AF278628	AF320384
Columbicola columbae	*Columba livia* (Feral Pigeon)	AF190415	AF278620	AF320385
Discocorpus mexicanus	*Crypturellus cinnamomeus* (Thicket Tinamou)	AF189133	XXXXXXX	AF320392

Docophoroides brevis	*Diomedia epomophora* (Royal Albatross)	XXXXXXX	XXXXXXX	AF320394
Echinophthirius horridus	*Phoca vitulina* (Harbour Seal)	AF189134	XXXXXXX	AF320396
Haffneria grandis	*Catharacta skua* (Great Skua)	AF189135	XXXXXXX	AF320406
Halipeurus pelagicus	*Bulweria bulwerii* (Bulwer's Petrel)	AF189136	XXXXXXX	AF320408
Harrisoniella densa	*Diomedea immutabilis* (Laysan Albatross)	XXXXXXX	XXXXXXX	AF320410
Naubates harrisoni	*Puffinus assimilis* (Little Shearwater)	XXXXXXX	XXXXXXX	AF320432
Oxylipeurus chiniri	*Ortalis vetula* (Plain Chachalaca)	AF189140	XXXXXXX	AF320436
Paraclisis confidens	*Diomedea nigripes* (Black-browed Albatross)	XXXXXXX	XXXXXXX	AF320439
Pectenosoma verrucosa	*Crypturellus cinnamomeus* (Thicket Tinamou)	AF189141	XXXXXXX	AF320440
Pectinopygus brevicornis	*Phalacrocorax aristotelis* (European Shag)	AF189142	XXXXXXX	XXXXXXX
Quadraceps sp. A	*Uria aalge* (Common Murre)	XXXXXXX	XXXXXXX	AF320458
Rallicola colombiana	*Dendrocolaptes certhia* (Barred Woodcreeper)	AF189144	XXXXXXX	AF320459
Saemundssonia stresemanni	*Catharacta skua* (Great Skua)	AF189145	XXXXXXX	AF320466
Austrogoniodes cristati	*Eudyptes pachyrhynchus* (Fiordland Penguin)	Y14909	N/A	N/A
Naubates prioni	*Pachyptila vittata* (Broad-billed Prion)	XXXXXXX	N/A	N/A

Table 2. Pairwise distances between neighbor joining trees for lice constructed from12S rRNA, COI and EF1a sequences (either separately or combined) and from RNA tree edit distances. Lower triangle is the partition metric, upper triangle the triplet symmetric difference.

	12S rRNA	COI	EF1a	EF1a + COI	12S rRNA + EF1a + COI	RNA edit distance
12S rRNA	-	0.37	0.43	0.33	0.29	0.50
COI	34	-	0.33	0.28	0.28	0.51
Ef1a	28	28	-	0.32	0.33	0.52
EF1a + COI	26	26	14	-	0.07	0.51
12S rRNA + EF1a + COI	24	26	20	10	-	0.53
RNA edit distance	36	36	36	36	36	-

Figure captions

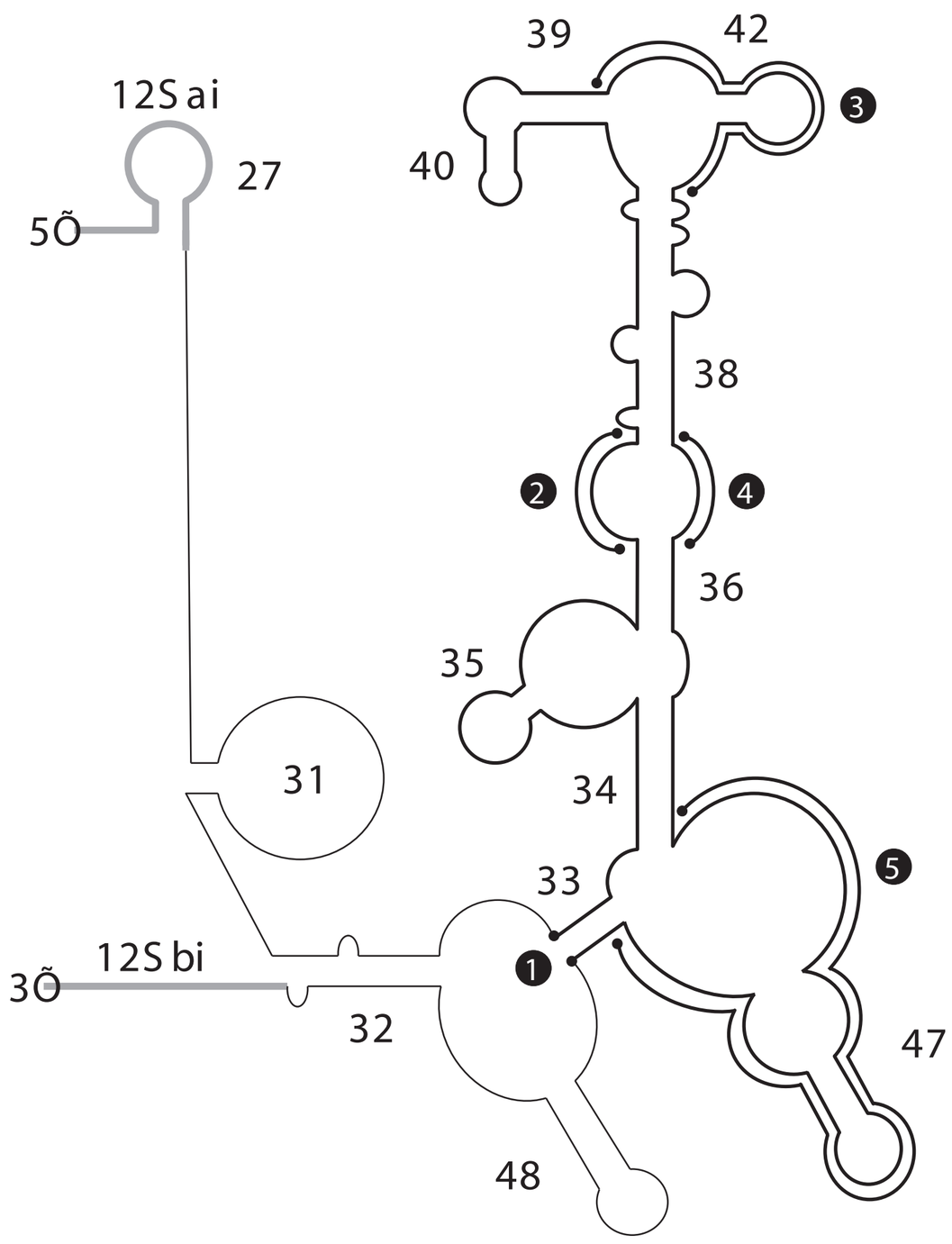
Fig. 1. Generalised secondary structure of insect domain III 12S rRNA (from Page 2000, fig. 6), showing the location of the 12Sai and 12Sbi PCR primers (Simon et al. 1994). The segments used to measure length variation in sequences (Fig. 3) are numbered 1-5. Segment 1 corresponds to the entire structure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

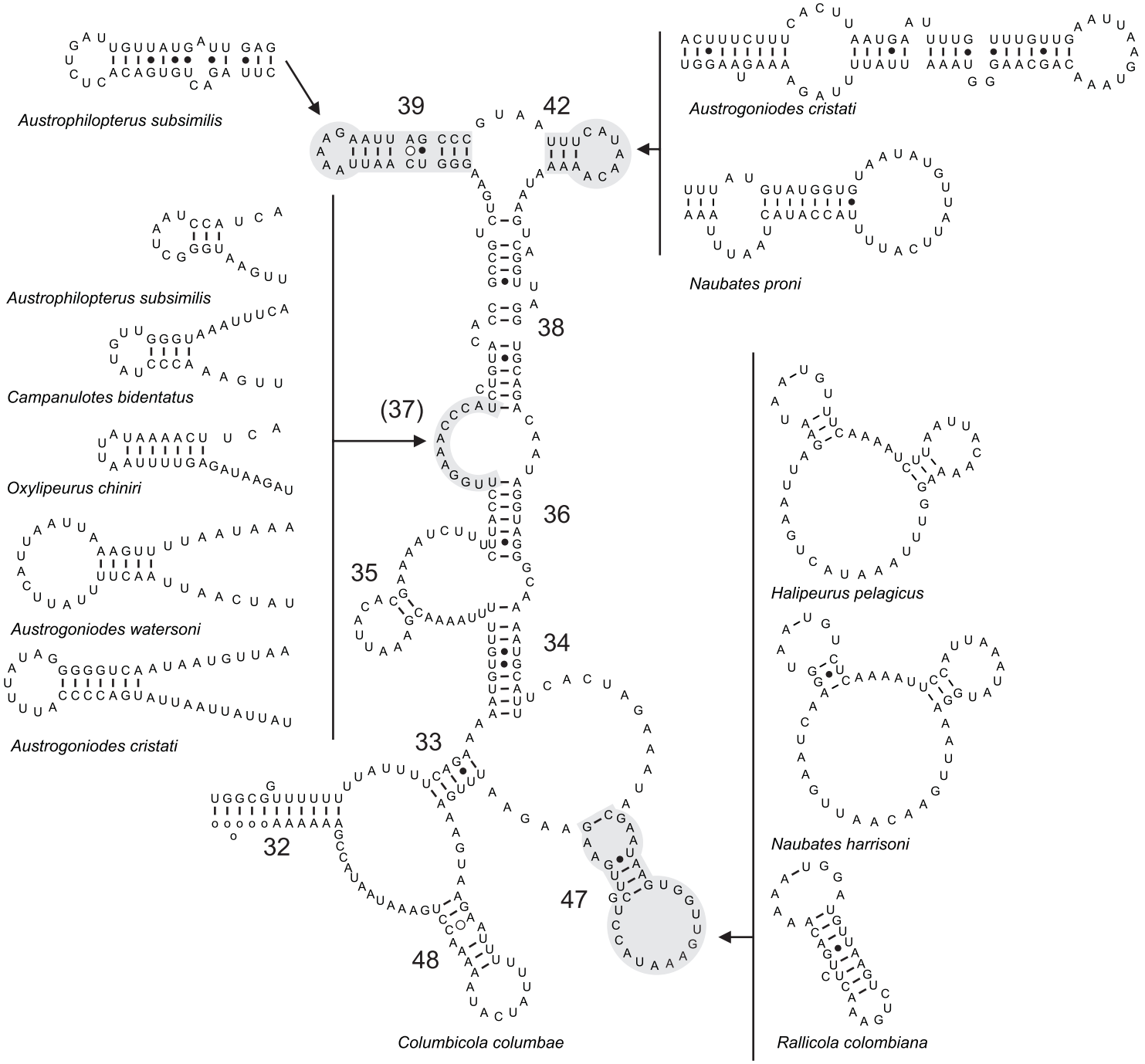
Fig. 2. Secondary structures for domain III for mitochondrial 1S rRNA for selected lice species. The complete structure for *Columbicola columbae* is shown, with regions that vary in other lice highlighted.

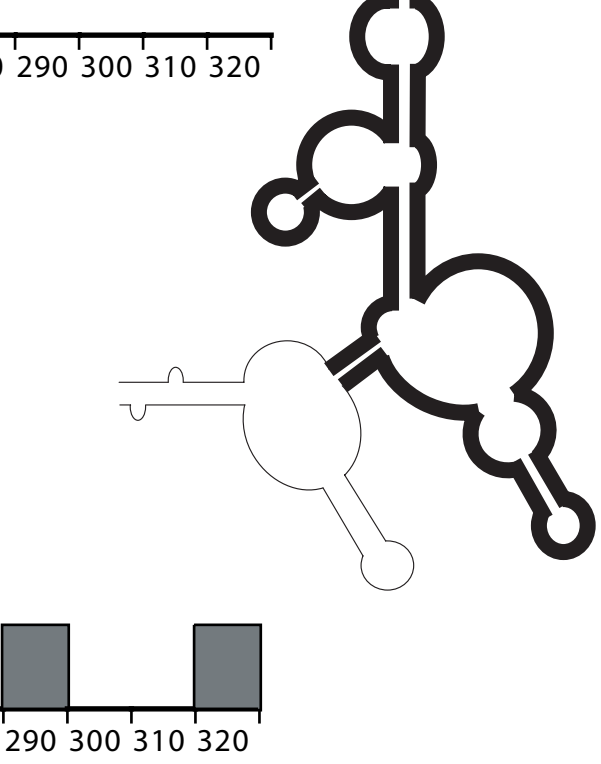
Fig. 3. Distribution of lengths in base pairs for four regions of 12S rRNA (see Fig. 1) for 225 insect and 25 louse sequences. Note the greater variation in lice compared to other insects.

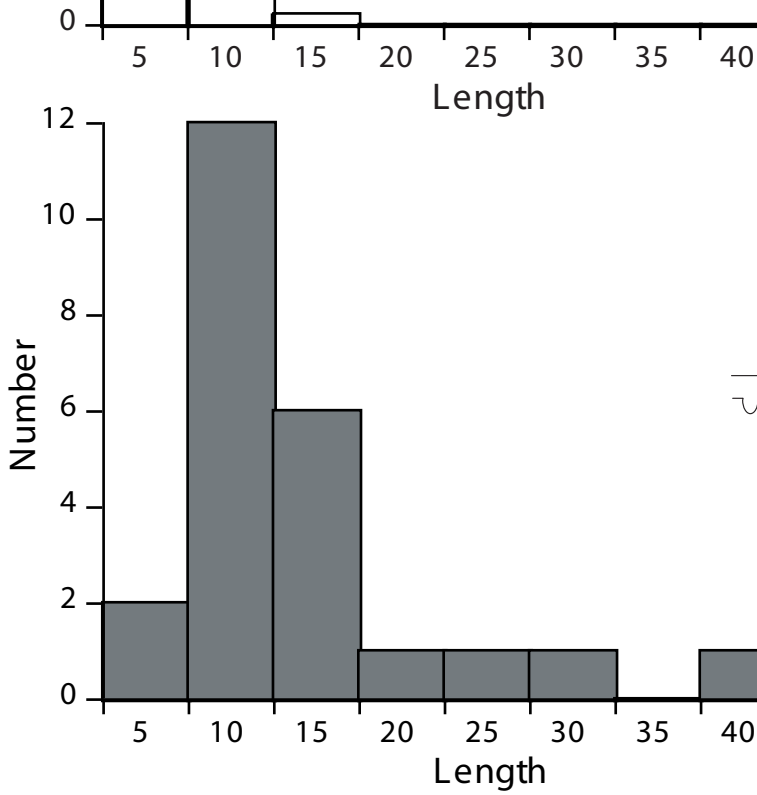
Fig. 4. Distribution of pairwise tree edit distances between 12S rRNA secondary structures for 19 insects and 25 lice (Table 1).

Fig. 5. Neighbor joining trees for the 23 lice in Table 1 constructed from LogDet distances for nucleotide sequences (12S rRNA, and EF1α and COI sequences combined), and from tree edit distance between 12S rRNA secondary structures.

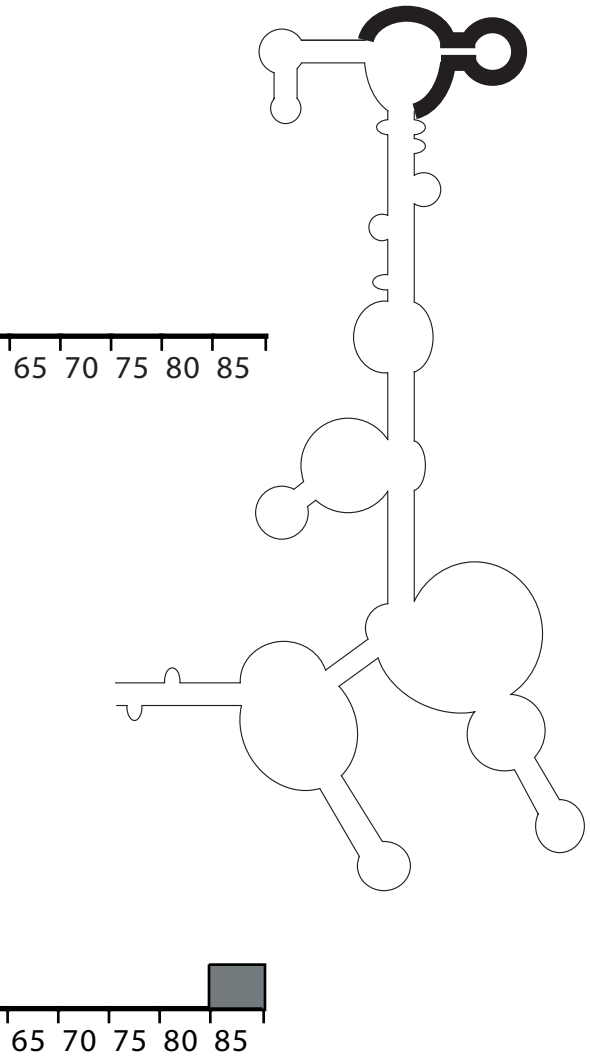

12S ai
5Õ
27
39
42
3
40
38
2
4
36
35
31
34
33
5
12S bi
3Õ
1
32
47
48


Austrophilopterus subsimilis
Austrophilopterus subsimilis
Campanulotes bidentatus
Oxylipeurus chiniri
Austrogoniodes watersoni
Austrogoniodes cristati
Austrogoniodes cristati
Naubates proni
Halipeurus pelagicus
Naubates harrisoni
Rallicola colombiana
Columbicola columbae
39
42
38
(37)
36
35
34
33
32
47
48

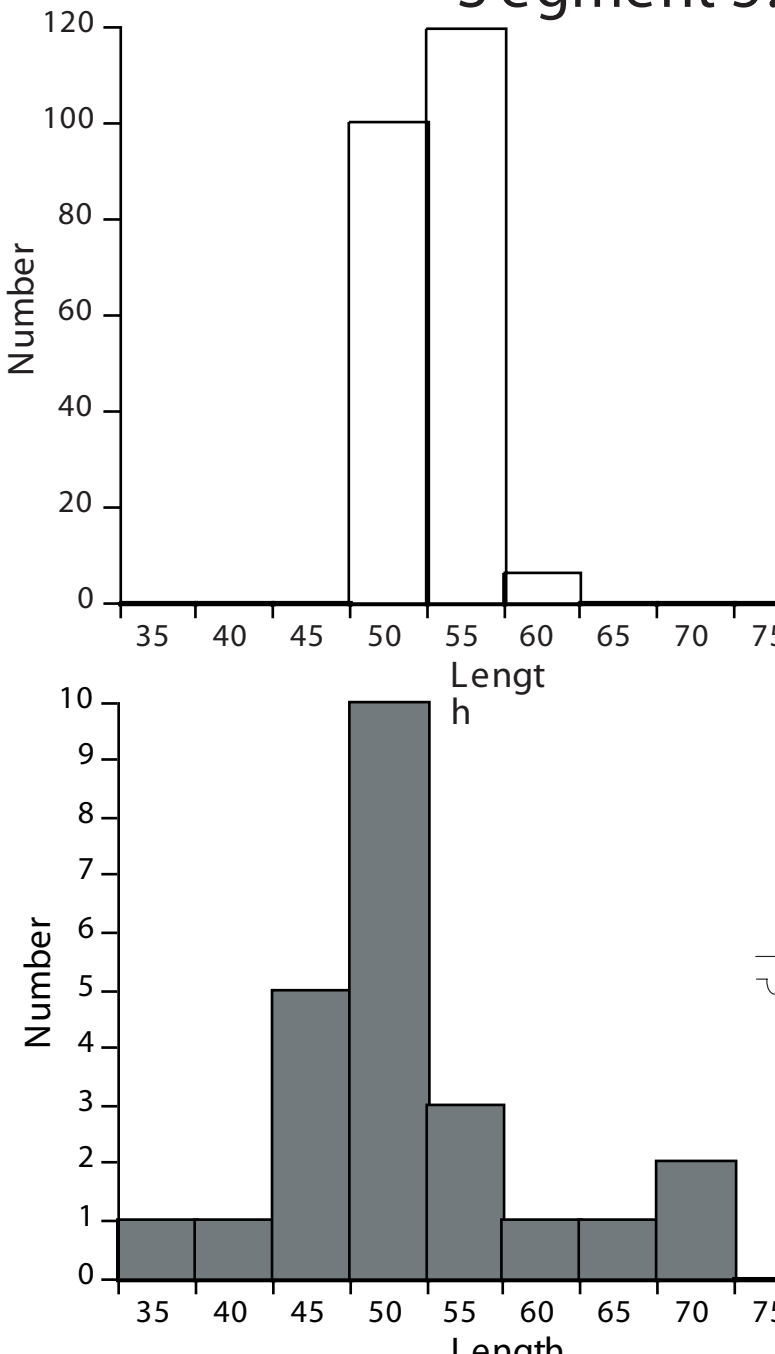


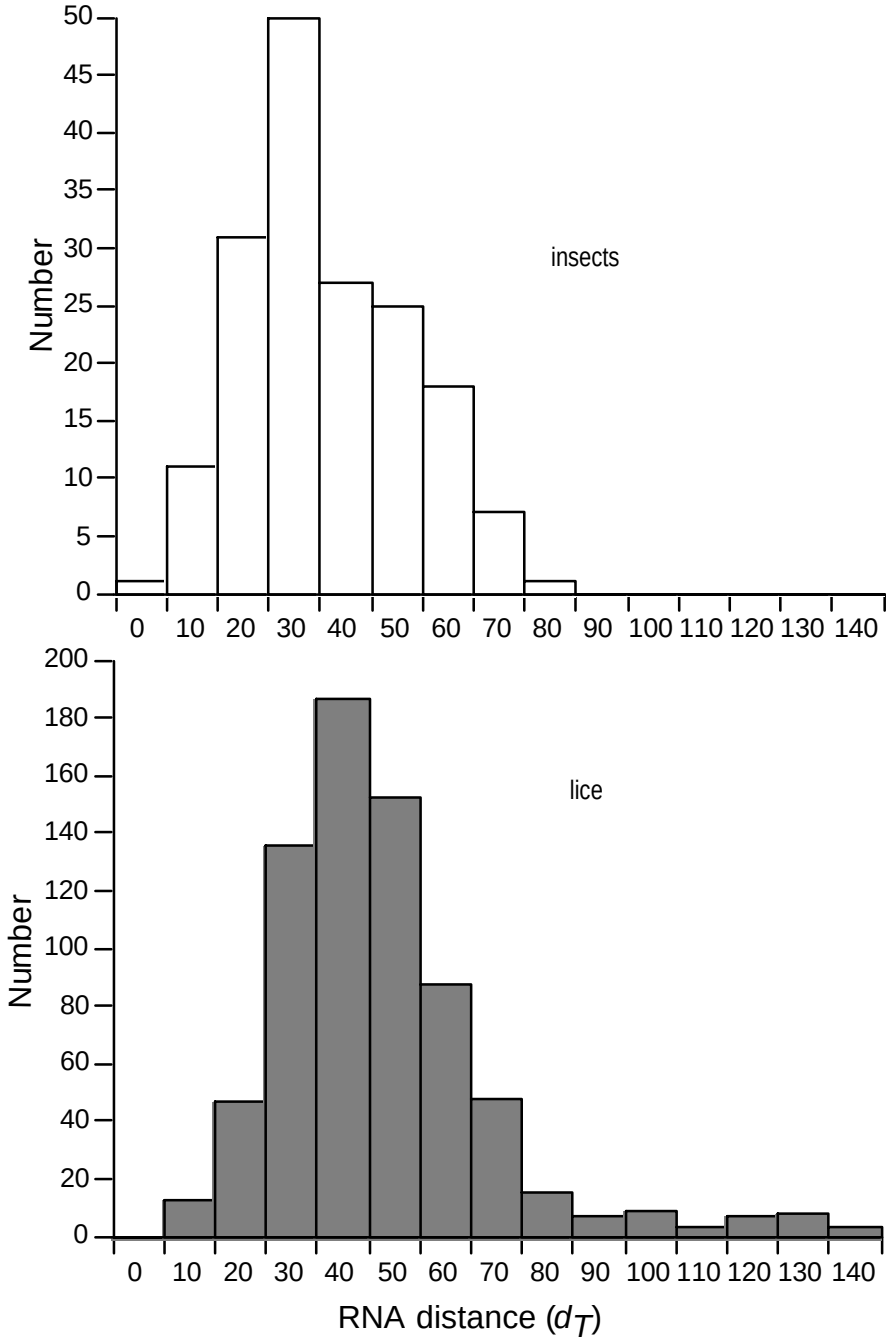


nt 3: 38'-39'

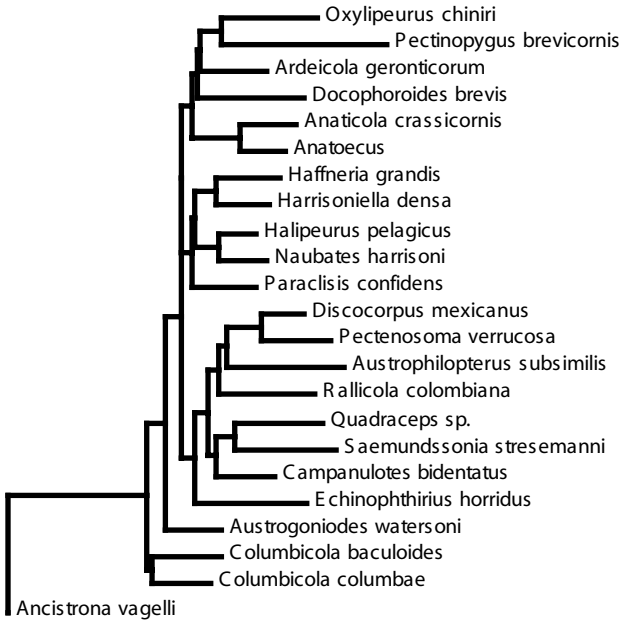


Segment 5:



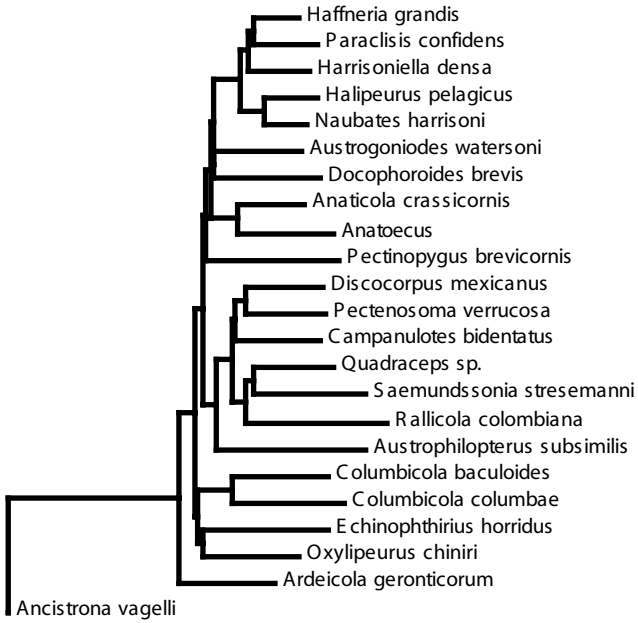

insects
Number
50
45
40
35
30
25
20
15
10
5
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
lice
Number
200
180
160
140
120
100
80
60
40
20
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
RNA distance (d_T)

12S rRNA sequences



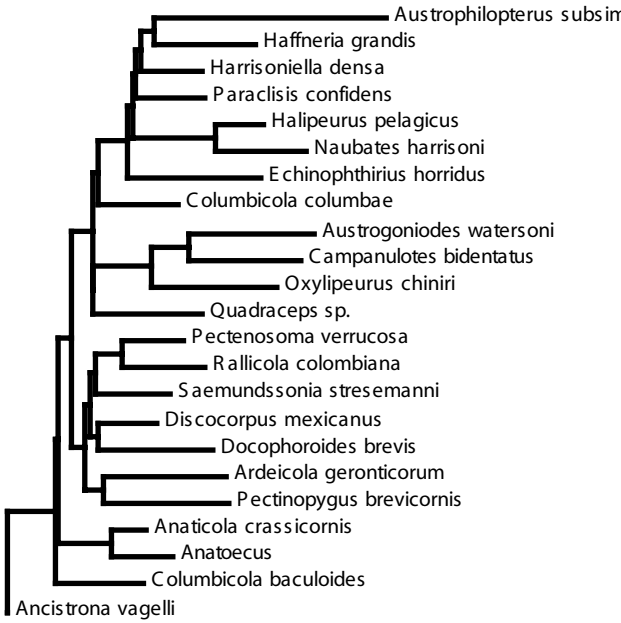
0.1 substitutions per site

EF1a and COI



0.1 substitutions per site

12S rRNA secondary structure



- 1 unit