Final report on NERC grant GR3/11075

"Cospeciation and relative rates of evolution in lice and birds: a molecular approach"

Roderic D. M. Page, R. Griffiths, and R. W. Furness

This report summarises the main results from different aspects of the project, and discusses future directions for the research. We also point out problems encountered during the study.

1 Louse phylogeny

Our first task was to place seabird lice in a broader phylogenetic context, so that we could establish whether lice hosted by seabirds are monophyletic, and to identify appropriate outgroups for rooting seabird louse phylogenies. Given the absence of a credible phylogeny of lice prior to the start of our project, we chose to survey a number of taxa using the EF1-α gene. This nuclear gene was chosen for an initial survey of the phylogeny of lice for a number of reasons: low copy number, ease of alignment, proven phylogenetic utility in other insect groups. A further consideration was that universal primers already available gave reliable PCR amplification in lice[1].

As our collections grew the scope of this study expanded, and in collaboration with colleagues at the University of Utah we sequenced 129 specimens representing 108 species from all four suborders of lice. Phylogenetic analysis of this data (1) yielded a number of well-supported monophyletic groups, but the relationships among many of these groups could not be resolved. With respect to seabird lice, it is clear that the ischnoceran lice hosted by petrels and albatrosses seabirds are not monophyletic: the genus *Saemundssonia* (also found on gulls, auks and terns) is quite unrelated to the remaining procellarid lice. The group informally recognised as the "*Philoceanus* complex" form a clade restricted to the Procellariformes[2]. The EF1-α data does not resolve the relationships of the two remaining louse genera found on procellariforms: *Docophoriodes* (on albatrosses) and *Trabeculus* (a petrel louse for which we could not amplify EF1-α).

A general feature of the EF1-α phylogeny is the very poor level of resolution at mid- to deep-levels. While it is probable that multiple substitutions at high divergences, coupled with an ancient radiation over a short period of time have contributed to the problem, most of this lack of resolution is probably due to the short length of the sequences (347 bp in length).

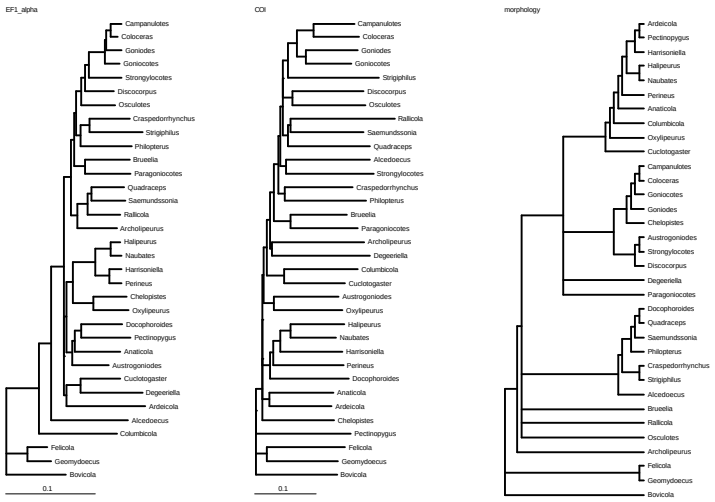
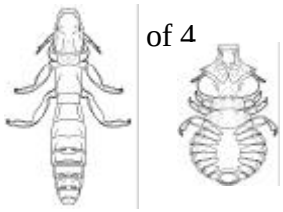


Fig. 1 Phylogenies for a subset of lice for which we have EF-1a, COI and morphological data .

In parallel with this project, a PhD student in my lab undertook a morphological phylogeny of ischnoceran lice (2). There are some striking disagreements between the molecular and morphological results (Fig. 1). In some cases we have independent support for the EF1-α results from other genes (COI and 12S rRNA), suggesting that, in some cases, closely related lice have evolved very divergent morphologies. The apparent morphological division of lice into short, robust "head lice", and long slender "wing lice" does not always reflect phylogeny — it appears that unrelated lice on different host groups have repeatedly evolved similar

[1] Lice have often proved very difficult to sequence, hence once we had identified a gene we could reliably sequence we concentrated on sequencing as many relevant taxa as possible.

[2] With the exception of the louse *Haffneria grandis* found on skuas (Charadriiformes).



wing louse head louse

morphologies (1). The true extent of this convergence remains unclear given our doubts about some aspects of the EF1-α phylogeny.

2 Mitochondrial sequence evolution in lice

2.1 Protein genes

In parallel with our work on EF1-α, together with our colleagues in Utah we sequenced the same taxa for the mitochondrial gene cytochrome oxidase I (COI)[3]. While COI has proved useful in constructing generic-level phylogenies of lice (3), it appears to be of little use at higher levels. The reason for this became apparent when we compared levels of sequence divergence between COI and EF1-α in the same species of lice. Maximum likelihood analyses (4) of louse and other insect sequences show that lice have an astonishingly high rate of substitution in mitochondria. Whereas in insects mtDNA typically evolves 10-20 times faster than nuclear DNA , in lice the difference is 200-300 fold (5). This elevated rate of substitution quickly obliterates phylogenetic signal at all but the tips of the louse tree.

1.2 Ribosomal gene structure and evolution

Our primary gene for investigating seabird louse phylogeny is the mitochondrial small subunit rRNA gene (12S). The choice of this gene was based on Paterson's (6) use of it in his pioneering study of procellariform louse phylogeny[4]. By using the same gene we could incorporate his sequences into a larger louse phylogeny. Paterson et al. (7) noted some unusual features of louse 12S, and in our research it quickly became apparent that louse 12S is very difficult to align. This prompted the PI to explore ways of aligning the sequences using secondary structure. Almost all, automatic, alignment methods use only primary sequence in alignments: however, ignoring secondary structure can produce incorrect alignments (8, 9).

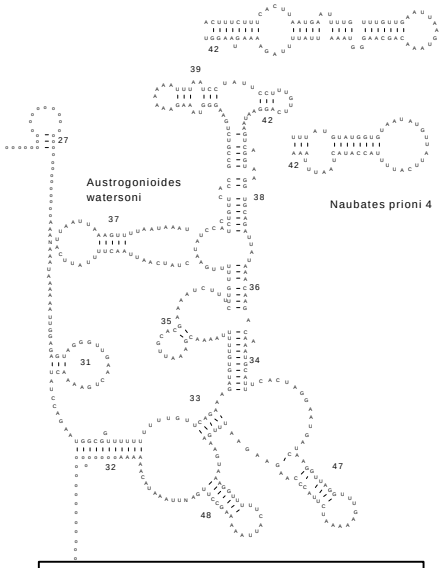


Fig. 2 Novel stems in louse 12S rRNA compared to a typical insect structure

In the absence of a robust model for insect 12S rRNA, the PI developed a computer program to infer secondary structure using comparative methods (10). This method was applied to 225 insect sequences obtained from GenBank to infer a general model of insect 12S secondary structure (11). This model is being used to aligning the louse 12S sequences obtained in this study. Louse 12S rRNA secondary structures show a greater range variation than is found any other group of insects (12). Work is in progress on quantifying this variation using RNA secondary structure metrics (13).

1.3 Summary of mitochondrial evolution

Our research has shown that lice have very elevated levels of substitution in mitochondrial COI (5), and novel 12S rRNA secondary structures (12), suggesting that there is something very distinctive about mtDNA evolution in lice. This conclusion is supported by the first published whole mitochondrial genome of a louse (14), which shows numerous gene order rearrangements, resulting in a genome very different from that of the ancestral insect. At present there is no evidence that louse nuclear genes are evolving at a substantially higher rate than in other insects, hence explanations must be sought that are specific to mitochondrial DNA.

[3] We contributed about a third of the COI sequences used in this study.
[4] This work was completed in 1994, but not published until 2000.

3 Seabird louse phylogeny and cospeciation

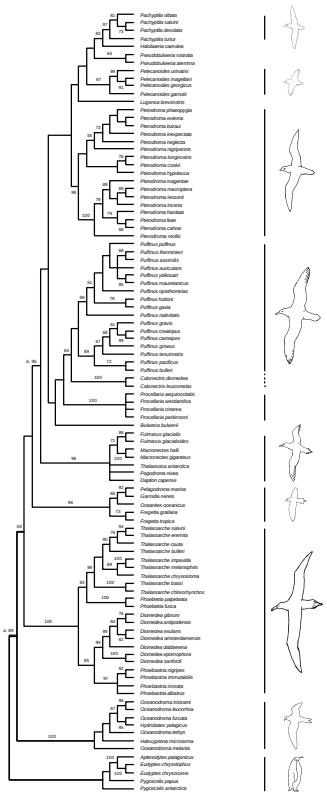


Fig. 3 Phylogeny for seabirds (from Kennedy and Page, submitted)

A main theme of the original proposal was investigating cospeciation and comparative molecular evolution in seabirds and their lice.

3.1 Seabird phylogeny

Studies of cospeciation are complicated by the need for two sets of phylogenies, one for hosts and one for parasites. In order to have a robust phylogeny for seabirds the PI and a NZ FRST funded post doc (Dr Martyn Kennedy) constructed a seabird supertree based on previpously published phylogenies (15). This supertree (Fig. 3) is the single largest phylogeny for these birds, and is being used as the reference point for testing hypotheses of cospeciation between seabirds and their lice.

3.2 Seabird louse phylogeny

The difficulty of identifying outgroups for petrel lice, the rapid rate of mtDNA evolution, and the novel 12S secondary structures have all complicated our efforts to unravel petrel louse phylogeny. We have obtained x 12S and y COI sequences for petrel lice (in addition to z EF1-α). Many of these were obtained in the last weeks of the project, and have not yet been fully analyse. Preliminary results strongly suggest that albatrosses are host to an *in situ* radiation of between 2-4 clades of large lice (*Harrisoniella*, *Paraclisis*, *Perineus*), each clade showing a high degree of cospeciation with their hosts. In contrast, the genera *Halipeurus* and *Naubates* found on petrels show a more complex history, probably involving several host switches.

Part of this difference in degree of cospeciation could be due to differing host biology, but it might also reflect the different shape of the host phylogenies (15). Albatross phylogeny is characterised by long internal branches and shorter terminals, so that the major clades are very distinct. In contrast, the basal branches in the petrel tree are short, suggesting a rapid radiation. It is possible that lice have found it “easier” to track the clearly demarcated lineages of albatrosses compared to the less clearly defined petrel clades. This hypothesis is being investigated quantitatively by comparing relative amounts of cospeciation between petrel and albatross louse clades.

4 Databases

In addition to the results above and the associated publications and manuscripts (either in review or preparation), this project has yielded a substantial collection of lice in alcohol. This collection is a valuable resource. Information on this collection is stored in a database, which is in the being made available online (http://130.209.46.190/cfdocs/exampleapp/lousebase/). Users can query which genera and species we have sequenced, and download those sequences for further study. All sequences reported in published papers are submitted to GenBank, and data matrices are submitted to TreeBASE. Other databases created during this project include a secondary structure database (http://taxonomy.zoology.gla.ac.uk/~rdmp1c/lice/12S/) and an online louse bibliography (http://taxonomy.zoology.gla.ac.uk/~vsmith/bibliography/bib_frame.html). These resources greatly strengthen the infrastructure for future research on these organisms.

5 Future directions

The failure of EF1-α to resolve higher level relationships among lice is disappointing. In retrospect, rather than pursue one gene relentlessly across Phthiraptera, a more sophisticated strategy would be to screen a range of nuclear genes for appropriate levels of variation using a carefully selected set of taxa that

span a range of taxonomic levels. At the time this project started we simply were not in a position to identify an appropriate set of taxa with any confidence — what meagre classifications were available were mutually incompatible, and hence we had grounds for deciding which taxa should be used for a sampling. In the light of our results we are now in a better place to choose appropriate taxa, and some preliminary work was been done on screening other genes (such as wingless and 18S rRNA) for appropriate levels of phylogenetic variation. This is one area we are currently pursuing.

Given that mtDNA sequences have proved to be poor markers for deeper louse phylogenetics, mitochondrial gene order may prove to be a more profitable source of characters for resolving higher level louse phylogeny. Analyses are in progress to evaluate whether there is phylogenetic signal in louse secondary structure.

In light of the difficulty of resolving deep level louse phylogeny, there are several directions future research on this group could take. One is to abandon deep phylogenetics for studies nearer the tips of the tree. This is also where studies of the role of host switching and cospeciation in structuring bird-louse assemblages are most likely to be profitable. Examples of this approach can be found in the forthcoming book edited by the PI (16). Another approach is to frame questions that do not need a fully resolved phylogeny in order to be answered. An example of this is the antiquity of the bird-louse association. Using the methodology of Copper and Penny (17), but substituting birds for the fossil record, we could estimate the age of the bird-louse association by calibrating the rate of evolution in lice with respect to birds and estimating the age of radiation of lice (Fig. 4).

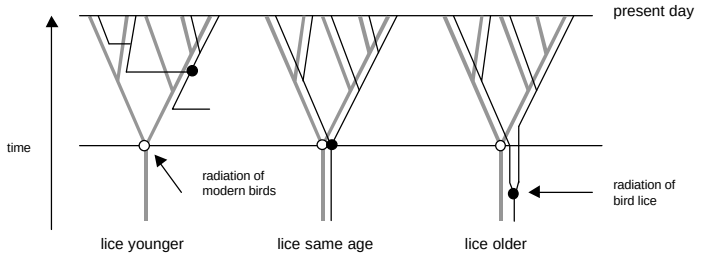


Fig. 4 Different hypotheses of the relative antiquity lice

6 References (publications funded by this grant in boldface)

1. Cruickshank, R. H., Johnson, K. P., Smith, V. S., Adams, R. J., Clayton, D. H. & Page, R. D. M. (2001) *Molecular Phylogenetics and Evolution* **19,** 202-215.
2. Smith, V. S. (2000) (University of Glasgow, Glasgow).
3. Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W. & Nadler, S. A. (1994) *Science* **265,** 1087-90.
4. Yang, Z. (1997) *Computer Applications in the Biological Sciences* **13,** 555-556.
5. Johnson, K. P., Cruickshank, R. H., Adams, R. J., Smith, V. S., Page, R. D. M. & Clayton, D. H. (in review) *Molecular Biology and Evolution* .
6. Paterson, A. M. (1994) in *Deoartment of Zoology* (University of Otago, Otago).
7. Paterson, A. M., Wallis, G. P., Wallis, L. J. & Gray, R. D. (2000) *Systematic Biology* **49,** 383-399.
8. Hickson, R. E., Simon, C., Cooper, A., Spicer, G. S., Sullivan, J. & Penny, D. (1996) *Molecular Biology and Evolution* **13,** 150-169.
9. Hickson, R. E., Simon, C. & Perrey, S. W. (2000) *Molecular Biology and Evolution* **17,** 530-539.
10. Page, R. D. M. (2000) *Bioinformatics* **16,** 1042-1043.
11. Page, R. D. M. (2000) *Nucleic Acids Research* **28,** 3839-3845.
12. Page, R. D. M., Crucikshank, R. & Johnson, K. (in prep) .
13. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M. & Schuster, P. (1994) *Monatshefte für Chemie* **125,** 167-188.
14. Shao, R., Campbell, N. J. H. & Barker, S. C. (2001) *Molecular Biology and Evolution* **18,** 858-865.
15. Kennedy, M. & Page, R. D. M. (submitted) *Auk* .
16. Page, R. D. M. (in press) (University of Chicago Press, Chicago).
17. Cooper, A. & Penny, D. (1997) *Science* **275,** 1109-1113.