Syst. Biol. 56(2):232–251, 2007

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701311370

Multiple Cophylogenetic Analyses Reveal Frequent Cospeciation between Pelecaniform Birds and *Pectinopygus* Lice

JOSEPH HUGHES,[1] MARTYN KENNEDY,[2] KEVIN P. JOHNSON,[3] RICARDO L. PALMA,[4] AND RODERIC D. M. PAGE[1]

[1]*Department of Environmental and Evolutionary Biology, University of Glasgow, UK; E-mail: j.hughes@bio.gla.ac.uk (J.H.)*
[2]*Allan Wilson Centre for Molecular Ecology and Evolution, Department of Zoology, University of Otago, P.O. Box 56, Dunedin, New Zealand*
[3]*Illinois Natural History Survey, Champaign, IL 61820, USA*
[4]*Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand*

Abstract.—Lice in the genus *Pectinopygus* parasitize a single order of birds (Pelecaniformes). To examine the degree of congruence between the phylogenies of 17 *Pectinopygus* species and their pelecaniform hosts, sequences from mitochondrial 12S rRNA, 16S rRNA, COI, and nuclear wingless and EF1-α genes (2290 nucleotides) and from mitochondrial 12S rRNA, COI, and ATPases 8 and 6 genes (1755 nucleotides) were obtained for the lice and the birds, respectively. Louse data partitions were analyzed for evidence of incongruence and evidence of long-branch attraction prior to cophylogenetic analyses. Host-parasite coevolution was studied by different methods: TreeFitter, TreeMap, ParaFit, likelihood-ratio test, data-based parsimony method, and correlation of coalescence times. All methods agree that there has been extensive cospeciation in this host-parasite system, but the results are sensitive to the selection of different phylogenetic hypotheses and analytical methods for evaluating cospeciation. Perfect congruence between phylogenies is not found in this association, probably as a result of occasional host switching by the lice. Errors due to phylogenetic reconstruction methods, incorrect or incomplete taxon sampling, or to different loci undergoing different evolutionary histories cannot be rejected, thus emphasizing the need for improved cophylogenetic methodologies. [Coalescence; coevolution; cospeciation; ILD; lice; ParaFit; Pelecaniformes; Phthiraptera; TreeFitter; TreeMap.]

When two interacting lineages have been in intimate association during much or all of their diversification, as in the case of some host-parasite interactions, it is probable that speciation in one group is paralleled by speciation in the other. This mode of diversification may have resulted in a pattern of shared evolutionary history between two lineages, known as cospeciation (Hafner and Nadler, 1988; Hafner et al., 1994; Huelsenbeck et al., 2000; Moran and Baumann, 1994; Page, 1994a, 1994b; Page and Charleston, 1998). Other coevolutionary phenomena such as host switching, failure of the parasite to speciate, sorting (e.g., parasite extinction), and duplication (speciation of the parasite on the host) also can influence the structure of host-parasite assemblages (Johnson and Clayton, 2003; Johnson et al., 2003).

Determining the extent of shared evolutionary history between two members of an association requires the determination of the phylogenies of both groups of organisms with sufficient confidence to make robust inferences. This can be a tall order, given that estimating a phylogeny for just one group of organisms can be difficult enough. Indeed, most methods for testing cospeciation use tree topologies (Page, 2003) and hence are potentially highly sensitive to the selection of genes and uncertainties in tree reconstruction for hosts and parasites. The expectation that host-parasite cospeciation will necessarily produce congruent phylogenies is naïve because species trees are estimated from gene trees and thus convey the history of a particular region of DNA and not necessarily that of the species (Rannala and Michalakis, 2003). Rannala and Michalakis (2003) also suggest that under certain conditions (e.g., large host and/or parasite population size and short time between speciation events), the probability that host and parasite gene trees are concordant is low, thus emphasizing the need to consider demographic factors when trying to understand differences between host and parasite gene trees.

Methods that can explicitly incorporate uncertainty in trees (e.g., Huelsenbeck et al., 2000, 2003) are in their infancy and rely on demonstrably inadequate models of host-parasite associations. In the absence of methods that have adequate models, we can apply existing tree comparison methods (TreeMap with Jungles: Charleston, 1998a; TreeFitter: Ronquist, 1995) to a range of trees representing alternative hypotheses of host and parasite relationships and identify those elements of the reconstructions that are shared across combinations of trees. Another approach is to use methods that do not compare trees but compare either data (ILD: Johnson et al., 2001) or some derivative of the data, such as distance matrices (ParaFit: Legendre et al., 2002). The ParaFit method (Legendre et al., 2002) uses matrices of patristic distances (summed branch lengths along a phylogenetic tree) or phylogenetic distances calculated directly from sequence data in both hosts and parasites, rather than the fully resolved topologies necessary in TreeMap and TreeFitter. Thus, ParaFit is not sensitive to the selection of different phylogenetic hypotheses, is not affected by polytomies in the tree, and can be used for any number of hosts per parasite or parasites per host.

If there are the same number of host and parasite taxa, and each parasite has a single, unique host, then we can treat the parasite data as just another source of data on host relationships (or vice versa). Hence, we could ask whether we can reject the hypothesis that the sequence data from lice and from their hosts support the same tree using, for example, a likelihood-ratio test (Huelsenbeck et al., 1997). This method is potentially powerful, although it is limited to the case of a

one-to-one host-parasite association and can lead to the hypothesis of host and parasite data underlying identical topologies to be incorrectly rejected. Indeed, in a study of cospeciation between bacterial endosymbionts and aphids (Clark et al., 2000), the patterns of base substitution in nonrecombining, maternally transmitted mitochondrial genomes were erratic enough that, even when model parameters were optimized for a given data set, the assumptions of the maximum likelihood models of evolution were violated. As a result, the assumptions of the LRT were invalid, leading to the false rejection of perfectly congruent phylogenies and cospeciation (Clark et al., 2000).

Lice are very attractive subjects for coevolutionary study because they represent the most significant radiation of insect ectoparasites (Marshall, 1981). Unlike fleas, lice have no free-living stage, and they are the most completely committed to parasitism of all the insects (Askew, 1971). Lice have low mobility and individuals that leave the host die within a few hours or days, such that lice are totally reliant upon the host for survival (Tompkins and Clayton, 1999). Host specificity tends to be high, with louse species usually restricted to either one or a group of closely related host species (e.g., Hellenthal and Price, 1991; Clayton et al., 1992; but see Price, 1975). Louse transfer between hosts usually requires direct contact and often occurs vertically (i.e., between host and their offspring), although phoresy, which occurs when lice attach to other animals such as hippoboscid flies for transportation to other hosts, has been observed (Marshall, 1981).

In this study we explore the effects of phylogenetic uncertainty on the inference of cospeciation between the louse *Pectinopygus* and its pelecaniform hosts. In general, lice have proved to be challenging subjects (Banks et al., 2006; Cruickshank et al., 2001) because sequencing of mitochondrial genes has been hampered by the massive rearrangements undergone by louse mtDNA (Shao et al., 2001). Classic mtDNA markers such as COI evolve very rapidly in lice (Johnson et al., 2003), becoming subject to multiple substitutions at all but shallow evolutionary depths. Other classical markers such as 12S rRNA have proven to be challenging to align due to exceptional secondary structure variation (Page et al., 2002). Few nuclear markers have been successfully amplified in lice, and those that have (such as EF1-α) are highly conserved and of limited value (Cruickshank et al., 2001).

Phylogenetic inferences for the hosts also are plagued with difficulties. Avian phylogeny remains somewhat uncertain (e.g., Garcia-Moreno et al., 2003; van Tuinen et al., 2000), and the Pelecaniformes are no exception. Ornithologists typically recognize 6 families in the order Pelecaniformes with a total of 57 species: Pelecanidae with 7 species, Sulidae with 9 species, Phaethontidae with 3 species, Phalacrocoracidae with 29 species, Fregatidae with 5 species, and Anhingidae with 4 species (Brooke and Birkhead, 1991). However, there is dispute over the membership of the group (e.g., whether the tropic bird belongs here; Kennedy and Spencer, 2004), and even the relationships within the "core" Pelecaniformes (gannets, boobies, darters, and cormorants) have proven resistant to resolution. For example, placing the darters can be hampered by long-branch attraction (Kennedy et al., 2005). The relationship of the pelicans to the core Pelecaniformes is also contentious and recent research suggests that pelicans are not sister to the core Pelecaniformes (Cracraft et al., 2004; Ericson et al., 2006), whereas other data suggest that the frigatebirds do group with the core Pelecaniformes (Ericson et al., 2006) and the monophyly of the core Pelecaniformes is strongly supported in these studies.

The genus *Pectinopygus* (Phthiraptera: Ischnocera) seems an obvious candidate for cospeciation studies as it is host specific and is of manageable size (39 species). *Pectinopygus* lice are found on all members of the Pelecaniformes except tropic birds (*Phaethon* spp.), which have a typical gull louse fauna (comprising the genera *Saemundssonia* and *Austromenopon*). The Pelecaniformes also are host to the amblyceran louse *Piagetiella*, which is found in the throat pouch of pelicans and cormorants. Both *Pectinopygus* and *Piagetiella* are unique to the Pelecaniformes, and the presence of this distinctive louse fauna is consistent with these birds being a monophyletic group (to the exclusion of the tropic birds). Of the 39 species in the genus *Pectinopygus*, 21 are found on cormorants (*Phalacrocorax*), 6 on pelicans (*Pelecanus*), 4 on boobies (*Sula*), 3 on frigatebirds (*Fregata*), 3 on darters (*Anhinga*), 1 on gannets (*Morus*), and 1 on the flightless cormorant (*Nannopterum*; Price et al., 2003). We compare the phylogeny of 18 host birds to the phylogenies estimated for 17 of their chewing lice with particular emphasis on the louse phylogenetic reconstruction. We explore the possibility that errors due to phylogenetic reconstruction methods, different loci having different evolutionary histories, or incompletely or incorrectly sampled taxa could cause incongruence between the host phylogeny and the *Pectinopygus* phylogeny, as well as the possibility that incongruence is a result of historical events such as host switching, extinction, duplication, and/or, lineage sorting.

Materials and Methods

Pectinopygus *Lice and Their Hosts*

Although we could not sequence the lice from the same host individuals that we sampled, we included 17 louse species from the genus *Pectinopygus* (Appendix 1) known to parasitize 18 different host species (listed in Appendix 2). *Pectinopygus bassani* was sampled from both *Morus bassanus* and *Morus serrator*, bringing the total of louse specimens sequenced to 18. The hosts in this study are found worldwide, and most have their breeding sites in coastal or insular regions with broad geographical ranges: (1) *Phalacrocorax pygmaeus*, *P. aristotelis*, and *Morus bassanus* have a European distribution and *P. carbo* occurs in North America, Eurasia, Africa, and Australasia; (2) *Anhinga novaehollandiae*, *P. sulcirostris*, *P. varius*, *P. punctatus*, and *M. serrator* are found in Australia and New Zealand; (3) *Fregata minor*, *Sula sula*, *S. dactylatra*, and *S. nebouxii* share a distribution from the Pacific

coast of the Americas to the Indian Ocean; (4) *Phalacrocorax auritus*, *Pelecanus erythrorhynchus*, and *Pelecanus occidentalis* are found mainly in North America; (5) *Sula leucogaster* is known from the Atlantic, Pacific, and Indian Oceans; and (6) *F. magnificens* is mainly distributed along the Atlantic and Pacific coasts of the Western Hemisphere. *Pectinopygus* lice collected from these bird species were preserved in ethanol until DNA extraction.

Amplification and Sequencing

Methods for polymerase chain reaction amplification for COI, EF1-α, 16S, and 12S in lice follow Cruickshank et al. (2001), Page et al. (2004), and Yoshizawa and Johnson (2003). The method for amplification of wingless was identical to that of EF1-α (Cruickshank et al., 2001) using the primers Lep wg1a (GAR TGY AAR TGY CAY GGY ATG TCT GG) and Lep wg2a (ACT ICG CAR CAC CAR TGG AAT GTR CA) designed by Danforth et al. (2004). For this study, all five genes were sequenced for both DNA strands. All sequences have been deposited in GenBank (Appendix 3) and LouseBase (http://darwin.zoology.gla.ac.uk/~rpage/LouseBase/2/), and the data matrix for lice is available in TreeBase (accession SN2851). The host data set contained ATPases 8 and 6, COI, and 12S, with a number of the sequences new to this study, added to preexisting sequences from Kennedy et al. (2000), Kennedy and Spencer (2004), and Kennedy et al. (2005; Appendix 4). Amplification and sequencing of these genes followed the methodology in the latter studies. Accession numbers are available in Appendix 4 and the data matrix for birds is available in TreeBase (accession SN2851).

Sequence Alignment

Nucleotide sequences of birds were aligned in ClustalX. The alignment of louse COI, wingless, and EF1-α also was performed in ClustalX. The sequences of 12S and 16S for the lice were aligned using the profile alignment mode in ClustalX followed by manual realignment based on secondary structure: the 12S rDNA alignment model for lice was provided by Page et al. (2002), and the 16S rDNA alignment for insects was provided by Buckley et al. (2000). These secondary structure models were used as alignment profiles in ClustalX.

Using the default settings of the Profile Mode of ClustalX, most of the stem regions of 12S and 16S rDNA were well aligned and correspond to secondary structure models. However, some of the more variable stems and loops of 12S and 16S, including long insertion/deletions, were present. Page et al. (2002) showed that some stem loop regions of louse 12S rDNA are highly variable in structure and that these stems and loops are phylogenetically less informative than others. Thus, we divided 12S and 16S into well-aligned regions (12SnoVar and 16SnoVar) and highly variable regions having long indels (12Svar and 16Svar). 12S and 16S are used to refer to the complete gene hereon.

Tests of Phylogenetic Congruence and Signal

The statistical significance of the incongruence length difference (ILD; Farris et al., 1994, 1995) between data partitions (12S, 16S, COI, EF1-α, wingless) was assessed in PAUP v4.0b10 (Swofford, 1998) by executing 200 replicates with only the taxa common to both partitions included in the analysis (Appendices 3 and 4). The signal of each data set for leaf stability was compared by positional congruence of the NJ trees in RadCon (Thorley and Page, 2000) to identify the most unstable taxa and focus upon their impact. When presented with a number of different trees for the same taxa, RadCon computes the degree to which taxa move around in the tree. Phylogenetic signal of the data sets was determined using the g1 statistic compared to the values in Hillis and Huelsenbeck (1992) and the permutation tail probability (PTP) test (Faith, 1991; Faith and Cranston, 1991) in PAUP v4.0b10 (Swofford, 1998).

A chi-square test of homogeneity of base frequencies across taxa was used to estimate the frequency distribution of the observed number of substitutional changes per character for each gene. The spectral analysis (Hendy, 1993) using the program Spectrum 2.3 (Charleston, 1998b) was used to evaluate the phylogenetic information in sequence data independently of a tree topology thus avoiding the difficulty of choosing the "best" method for tree reconstruction and the issue of whether the data are tree-like. Spectral analysis determines support for a split depending on the number of character columns in the alignment whose patterns correspond to that split. The conflict for a split is the sum of the support for the splits that are incompatible with it. As a split may be incompatible with many other splits, its conflict may be much larger than its support.

Phylogeny Reconstruction

To assess the relative stability of trees to methods of analysis, we used three different tree construction methods: parsimony, maximum likelihood, and Bayesian. *Gavia stellata* and *Anaticola crassicornis* were used as outgroups for the hosts and the lice, respectively. Because of questions about the relationship of the genus *Pelecanus* to the core pelecaniforms (darters, boobies, gannets, and cormorants) and the limited sequence data and knowledge of suitable outgroups for the *Pectinopygus* lice, the phylogenetic analyses also were conducted with *Pelecanus occidentalis* and *Pelecanus erythrorhynchos* as additional outgroups for the birds and *Pectinopygus occidentalis* and *Pectinopygus tordoffi* for the lice to determine whether the selection of the outgroup affects the topology of the core pelecaniforms and their lice.

Phylogenies were estimated for each gene as well as the combined data set of lice. The pelecaniform data were analyzed thoroughly by Kennedy et al. (2000, 2005) and Kennedy and Spencer (2004), so all phylogenetic reconstructions for the birds were applied only to the combined bird genes. Maximum parsimony (MP) phylogenies were estimated by heuristic searching with all

sites equally weighted, 1000 random addition replicates with tree bisection-reconnection (TBR) branch swapping in PAUP v4.0b10 (Swofford, 1998). Under the maximum likelihood (ML) criteria, base frequencies and portion of invariant sites were estimated from the empirical levels, and the model of substitution was selected using ModelTest (Posada and Crandall, 1998). Bootstrapping (1000 heuristic replicates) was used to determine the strength of support for individual nodes. MrBayes v3.0 (Ronquist and Huelsenbeck, 2003) was used for calculation of Markov chain Monte Carlo Bayesian posterior probabilities for all genes as well as the combined bird genes and combined *Pectinopygus* genes with the following settings: the model employed 6 substitution types ("nst = 6") and rate variation across sites was modeled using a gamma distribution, with a proportion of sites being invariant ("rates = invgamma"). For the combined bird genes and combined *Pectinopygus* genes, a partitioned Bayesian analysis was performed with unlinked model parameters across the partitions of the data. The Markov chain Monte Carlo search was run with 4 chains for 2,000,000 generations, with trees being sampled every 100 generations (the first 1000 trees [100,000 generations] were discarded as burn-in). A plot of generation versus the log probability was used to check for stationarity, and the partition probabilities were compared in different runs to ensure convergence.

Topology-Based Methods of Cospeciation

TreeMap.—To assess whether species of *Pectinopygus* and seabirds have undergone parallel diversification, we used reconciliation analysis with Jungles analysis as implemented in TreeMap *2.0ß* (Charleston, 1998a). TreeMap (Page, 1994b) uses reconciled trees to compute the fit between the host and parasite phylogenies, allowing the incorporation of host-switching events and considering all potentially optimal solutions. TreeMap allows a graphic display of the results and therefore the identification of coevolutionary events. TreeMap also includes a testing procedure by generating random trees and comparing the random number of cospeciation events in the association to the observed number to assess whether it is significantly higher than chance alone.

TreeFitter.—TreeFitter 1.0 uses a method based on generalized parsimony to assess the fit between the host and parasite phylogenies, incorporating a differential cost to the four types of potential events occurring in a host-parasite association (Ronquist, 1995, 2003): cospeciation (C), duplication or intrahost speciation of the parasite (D), sorting or extinction of the parasite lineage (S), and host switching (H). The optimal reconstruction is the one that minimizes the global cost. TreeFitter uses a permutational procedure to test the overall cost and occurrence of each type of event.

In both TreeMap and TreeFitter, the null hypothesis that the two phylogenies are randomly related was tested by comparing the scores of optimal reconstructions (number of cospeciation events for TreeMap and global cost for TreeFitter) with those of 999 randomly generated phylogenies. Because these programs require fully resolved trees, we tested all combinations of the obtained topologies to account for phylogenetic uncertainties. TreeFitter also allows assignment of different costs to the four types of events, so we varied these costs to assess its effect on the test results.

Data- and Distance-Based Approaches for Testing Cospeciation

ILD test of cospeciation.—The data-based parsimony method of cophylogenetic analysis identifies lineages that are responsible for significant differences between phylogenies (Johnson et al., 2001). The ILD test was run in PAUP v4.0b10 (Swofford, 1998) on 100 replicates with 10 random addition replicates (maxtrees = 100) and tree bisection-reconnection branch swapping. First, significant incongruence was tested between the host partition (combined data set) and the parasite partition (combined data set). Then, a host-parasite association (a pair of host-parasite associations, a triplet of host-parasite associations, etc.) was removed in turn until a nonsignificant result was achieved (Johnson et al., 2001). Next, the combined evidence topology for the congruent host-parasite associations was reconstructed following the previously defined maximum parsimony and bootstrapping methods. This topology depicted a perfectly congruent host-parasite phylogeny to which the deleted hosts and parasites were added. The combined evidence topology for the host was constrained (backbone constraint), the deleted host taxa were added back to the host data set, and a parsimony search was conducted. This yielded a complete host phylogeny. The procedure was then carried out for the parasite. The final step was to compare the host and parasite phylogeny using reconciliation methods in TreeMap (Page, 1994b).

ParaFit.—ParaFit uses matrices of patristic distances (summed branch lengths along a phylogenetic tree) or distances calculated directly from sequence data, rather than tree topologies. In this test, distance matrices of the two groups are transformed using principal coordinates analysis, and a larger matrix is constructed based on these two distance matrices and a matrix that represents the host-parasite associations. The test statistic is based on the trace of this matrix, and its significance is determined using a permutation test (Legendre et al., 2002). The null hypothesis is that the two groups are randomly associated. This method also can test whether each host-parasite association contributes significantly to the global fit. In this study we used both uncorrected distances calculated from the combined sequence data for parasites and for hosts, and patristic distances calculated from the ML and MP trees. Principal coordinates for these distance matrices were calculated using the DistPCoA software (Legendre and Anderson, 1998), and the significance of the test statistic was evaluated by performing 999 permutations using the program ParaFit (Legendre et al., 2002). The formatted data files for the TreeMap, TreeFitter, and

ParaFit analyses are available for download at http://darwin.zoology.gla.ac.uk/~jhughes/Pectinopygus/.

Likelihood-ratio test (LRT).—If *Pectinopygus* and pelecaniform phylogenies show significantly higher congruence than expected by chance, the observed topological incongruence may simply be explained by error (e.g., limited number of informative characters or inadequate taxon sampling) rather than actual biological processes such as host shifting. Therefore, we tested the hypothesis that the same tree topology underlies the *Pectinopygus* and pelecaniform data sets using the likelihood-based Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999; Goldman et al., 2000) using PAUP v4.0b10 with RELL (resampling estimated log-likelihood) optimization and 10,000 bootstrap replicates. In this test, the likelihood score of the best topology for a given data set is compared to the scores of alternative topologies obtained from other data sets. Only the nine species (Appendix 3) for which all the genes were present in both the parasite and the host were used in this test. All unique topologies obtained in the ML searches for the separate *Pectinopygus* loci (12SnoVar, 16SnoVar, COI excluding 2nd and 3rd codon positions, EF-1α, and wingless) and the ML analysis of combined bird loci were used in this test. In this way, the level of conflict among the parasite loci could be determined. We also calculated the likelihood-ratio test statistic ($\delta = 2[\ln L_1 - \ln L_0]$), which measures the difference in the likelihood when each data set is allowed to have a different topology, L_1, from the likelihood obtained when all data sets are constrained to the same topology, L_0. Under the null hypothesis of a common topology, the tree assumed to underlie all data sets is that with the highest summed likelihood across data sets. We also examined the contribution of individual parasite loci to the observed heterogeneity in the data set by excluding individual genes from the calculation of δ (Huelsenbeck and Bull, 1996). The significance of δ was calculated using parametric bootstrapping. The test statistic was compared to a distribution of likelihood scores generated under the null hypothesis of identical topologies given the host and parasite data sets (Huelsenbeck et al. 1997). The null distribution of likelihood scores was constructed by optimizing likelihood parameters for each data set given the constrained tree. The program SeqGen 1.3.2 (Rambaut and Grassly, 1997) using the graphical interface SG Runner 2.0 (T.P. Wilcox; http://homepage.mac.com/tpwilcox/SGRUNNER/FileSharing8.html) was used to generate 100 data sets (Monte Carlo simulation) using the optimized parameters and the constrained topology of each parasite locus and the combined host data. The likelihood-ratio test statistic for the constrained and best topologies for each of these simulated data sets was calculated, and a null distribution of test statistics was constructed. The test statistic derived from the empirical data was then compared to the null distribution to determine if phylogenetic conflict existed between data sets.

Correlation of coalescence times.—Hypotheses of cospeciation imply that speciation events in hosts and parasites should be approximately contemporaneous. If sequence divergence in hosts and parasites is correlated and both taxa have molecular clocks, then fitting a line to a plot of parasite divergence against host divergence allows us to describe two aspects of host-parasite divergence (Hafner and Nadler, 1990; Hafner and Page, 1995). The slope of the line is the relative rate of host and parasite evolution, and the intercept measures the divergence of parasites when their hosts speciate. An intercept of zero implies synchronous cospeciation in the hosts and parasites.

The presence of a molecular clock for each gene was tested in PAUP v4.0b10 (Swofford, 1998) with the appropriate likelihood model selected by ModelTest (Posada and Crandall, 1998). The difference between likelihood score with a clock enforced and without a clock was used as the test statistic for a chi-square test and its significance was calculated using the number of taxa minus 2 degrees of freedom. Given the absence of a constant molecular clock for most of the loci, we employed the relaxed Bayesian methods for multilocus data to obtain the ultrametric trees (Thorne and Kishino, 2002; Thorne et al., 1998). PAUP 4.0b10 was used to infer the parameters for the F84+Γ model used by ESTBRANCHES to estimate the variance-covariance matrices for all 7 DNA fragments independently (wingless was excluded due to the small number of sequences). The output files from ESTBRANCHES were employed in MULTIDIVTIME to estimate the divergence times. The Markov chain Monte Carlo analyses used the default settings of 100,000 cycles in which the Markov chain was sampled 10,000 times every 100th cycle following burn-in. Although there are fossil pelecaniforms that could be use to calibrate the host tree (Dyke and Van Tuinen, 2004; Mayr, 2005), there are no fossils for *Pectinopygus* that would enable us to infer its age. Hence, we are restricted to comparing relative ages. The ingroup nodes in the bird and louse trees were assigned prior ages of 1.0 (SD = 1). If the birds and lice have indeed cospeciated, then we expect the speciation depth of nodes in the bird and louse trees to be highly correlated. The value of the rate of evolution at the root was varied and it was found that quite large changes to the root rate had little influence on the estimates. The correlation between copaths (i.e., homologous paths in the host and parasite phylogenies) were computed in TreeMap 1.0 using reduced major axis regression.

RESULTS

Data Evaluation

The ILD found no significant difference in phylogenetic signal between the data partitions in the parasites (12S, 16S, EF1-α, and wingless) except between COI and wingless ($P = 0.02$) and COI and 16S ($P = 0.01$). We also compared the positional congruence (Estabrook, 1992) between NJ trees resulting from the different data sets, and the comparison indicated topological congruence above 67% for all tree comparisons except from the trees estimated from 12Svar and 16Svar, which showed less congruence with the other trees derived from COI,

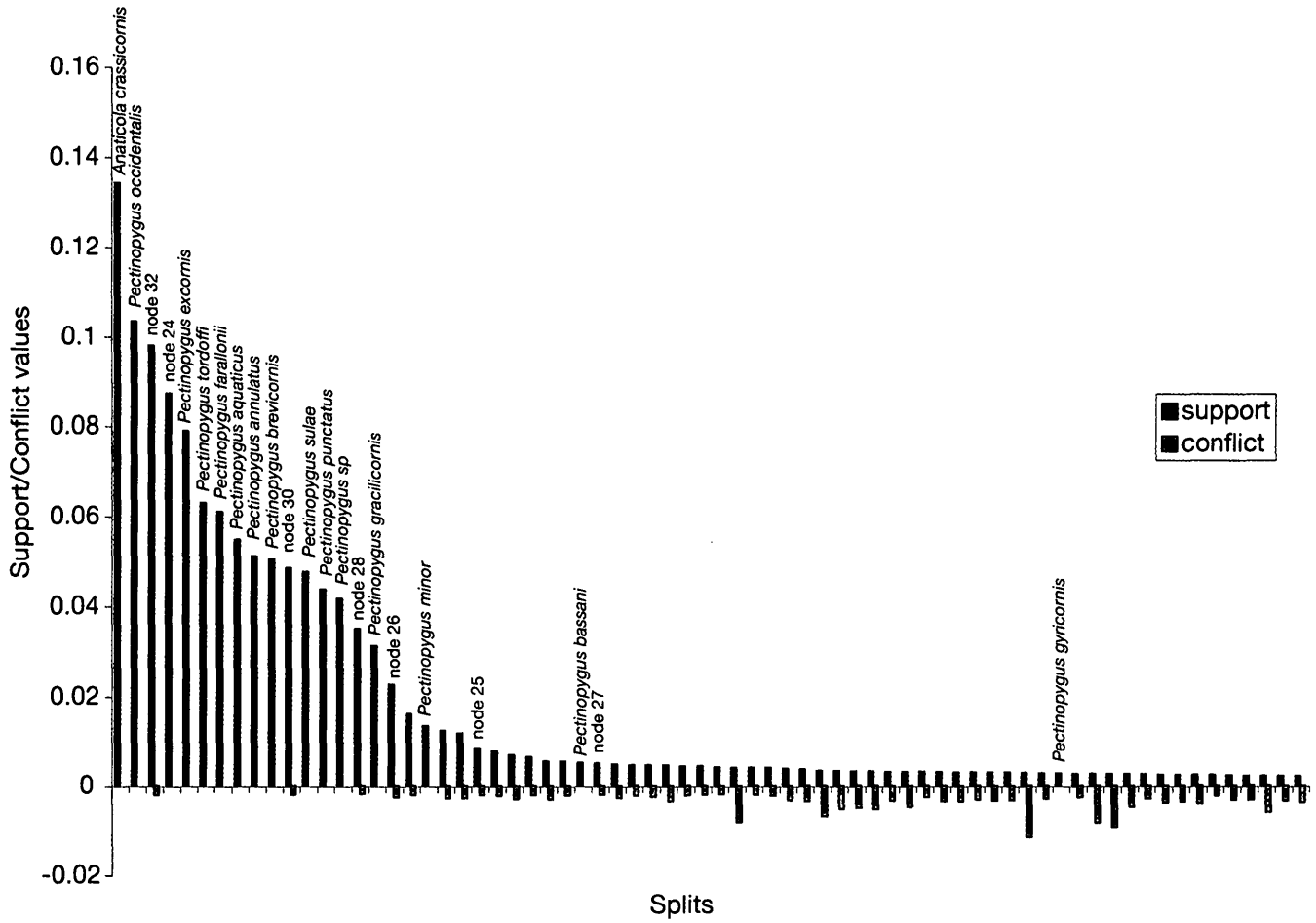

FIGURE 1. The support/conflict spectrum with the Hamming distances option. The node numbers are the labels for the internal branches shown in Figure 2 (right). Only the splits above the threshold of 0.02 are shown. The (negative) conflict values are normalized following Lento et al. (1995).

12SnoVar, 16SnoVar, EF1-α, and wingless (between 49% and 65%).

The chi-square test of homogeneity of base frequencies across louse taxa resulted in significant P-values for 16S and 12S only ($\chi^2 = 88.81$, $df = 54$, $P < 0.01$; $\chi^2 = 169.57$, $df = 54$, $P < 0.01$, respectively). However, after removal of the variable regions of 16S, the base frequencies resulted in no significant P-value ($\chi^2 = 35.85$, $df = 54$, $P < 0.97$), whereas the removal of 12S variable regions did not change the significance of the test of homogeneity. Although the variable regions of 12S and 16S do not show high incongruence with the other data partitions, we preferred to exclude the variable regions from further analyses as the alignment that corresponds to the secondary structure models for these regions is less reliable.

The combined well-aligned data set of lice includes 1903 bp, and out of the 857 variable sites, 712 of the characters were parsimony informative. The data set contained significant phylogenetic signal above background noise as indicated by the g1 statistic (g1 = −0.834584 from 10,000 random trees, $P < 0.01$; Hillis and Huelsenbeck, 1992) and the PTP test (1000 replicates, $P < 0.01$). Figure 1 is a graphic display of all phylogenetic information ("signals") contained in the combined data set of 12SnoVar, 16SnoVar, COI (excluding 2nd and 3rd codon positions), EF1-α, and wingless. The bars in the top half of the spectrum show the frequency of support for all splits for which there is evidence in the data. With 18 taxa, the number of possible splits is $2^{(18-1)} = 131{,}072$. Of these, 70 are supported by some evidence above the threshold of 0.02 and these are the bars shown on the top half of the spectrum. Fifteen splits represent singletons; 5 splits support monophyly of the gannet lice, frigate lice, booby lice, pelican lice, and grouping of the darter louse with the cormorant lice; for the remaining 131,002 splits, there is little evidence in the data. Bars below the x-axis in Figure 1 show the normalized sum of conflicting evidence for the corresponding split above the x-axis. The labeled bars indicate those splits that have been included in the maximum-parsimony tree (Fig. 2, right). The spectral analysis showed that removal of 2nd and 3rd positions of COI (1651 characters remaining, 567 informative) increased the total support and decreased the level of conflict.

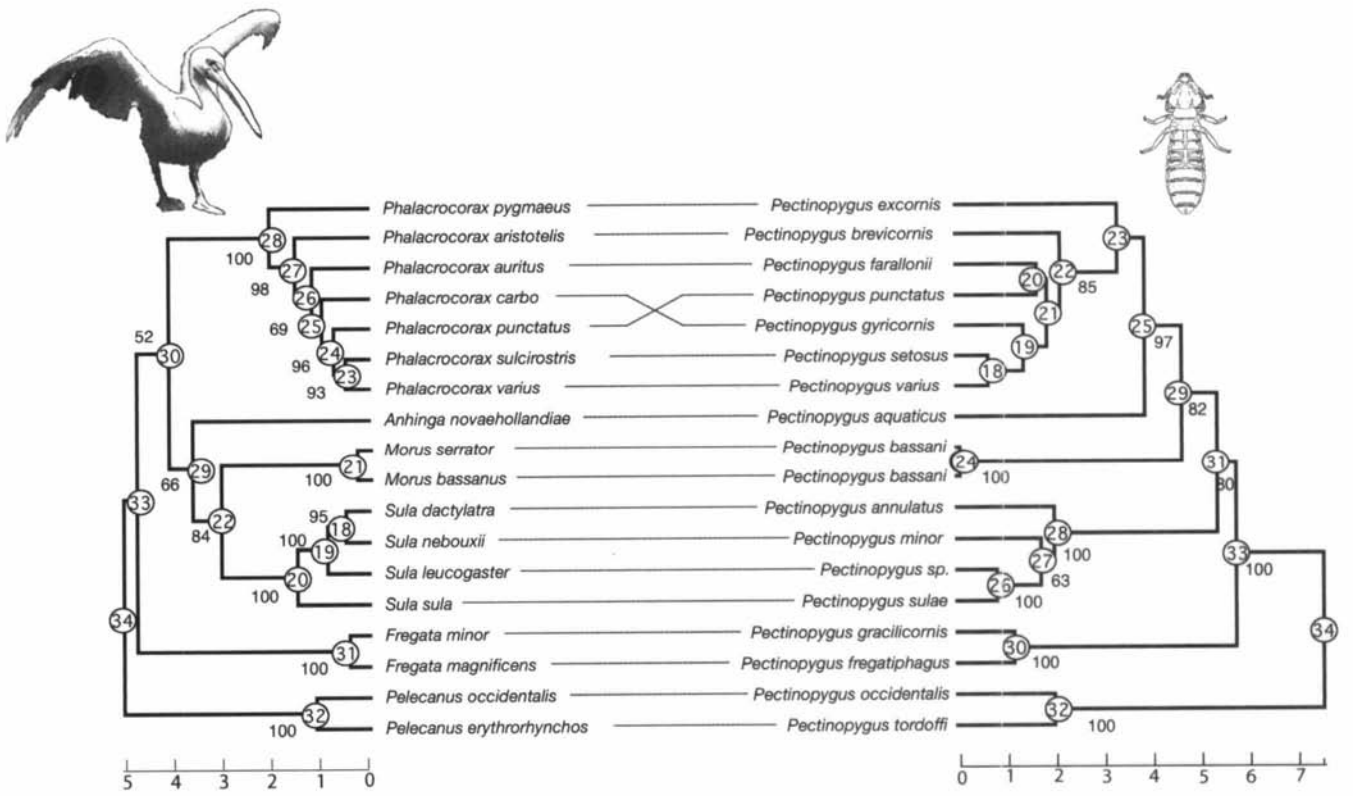

FIGURE 2. Phylogenetic trees for pelecaniform birds (left) and *Pectinopygus* lice (right) with lines connecting coexisting hosts and parasites. The pelecaniform tree is the maximum parsimony topology inferred from the combined sequences of COI, ATPase, and 12S rRNA with MP bootstrap beside the circled node label (length 1730, CI = 0.540, RI = 0.635); the *Pectinopygus* tree is the maximum parsimony based on the combined COI, 12S rRNA, 16S rRNA , EF1-α, and wingless sequences (length 2078, CI = 0.568, RI = 0.571). The bootstrap support is shown below the branch. The circled nodes correspond to the node numbers shown in subsequent figures. *Pectinopygus* image from http://sid.zoology.gla.ac.uk/.

Phylogenetic Analysis

Removal of 2nd and 3rd positions of COI did not significantly alter the topology of the maximum parsimony trees (MPTs) for the lice. Signal above background noise was still significant in the 1651 characters (567 parsimony informative) comprised of 12SnoVar, 16SnoVar, COI excluding 2nd and 3rd positions, EF1-α, and wingless as shown by the PTP test (1000 replicates, $P < 0.01$) and the g1 statistic (g1 = -0.892918 from 10,000 random trees, $P < 0.01$). Thus, these sites were removed from all additional analyses.

The MP analysis of the louse data led to a single MPT (tree length = 3912, CI = 0.504; Fig. 2). The relationships among the core pelecaniform lice were not affected when the pelican lice were also selected as outgroups. MP analyses of the individual partitions, however, resulted in different topologies with multiple MPTs for some partitions (Fig. 3). Gannet lice (*Pectinopygus bassani*) have a particularly unstable position in the latter topologies: sister to *Phalacrocorax* lice (*P. excornis* and relatives) for 12SnoVar, 16SnoVar, and wingless (Fig. 3a, b, and e); sister to *Pelecanus* lice (*P. tordoffi*, *P. occidentalis*) for COI (excluding 2nd and 3rd codon positions, Fig. 3d); or sister to the *Fregata* and *Sula* lice (*P. fegatiphagus*, *P. minor*, and relatives) for EF1-α (Fig. 3c).

The GTR model, which takes into account a proportion of invariable sites and a gamma distribution for substitution rate heterogeneity (GTR+G+I), was selected by ModelTest (Posada and Crandall, 1998) and was used to run the ML and Bayesian analyses. The ML and Bayesian trees from the combined *Pectinopygus* data recovered the same topology (Fig. 4). ML, Bayesian, and MP analyses produced similar topologies that were not significantly different using the likelihood-based Shimodaira-Hasegawa test (SH test; Shimodaira and Hasegawa, 1999; Goldman et al., 2000). However, in the ML and Bayesian trees, gannet lice are basal to the frigate and booby lice, in contrast to the MP topology where they are basal to the cormorant and darter lice. We decided to use both these trees in our subsequent analyses of cospeciation.

The phylogenetic analysis of the combined data sets of birds produced similar topologies for the MP (Fig. 2, left), ML, and Bayesian (Fig. 4, left) analyses that were not significantly different based on the Shimodaira-Hasegawa test. The only observed difference involved relationships within the cormorants (*Phalacrocorax*), wherein *P. carbo* forms a clade with *P. aristotelis* and *P. auritus* in the ML and Bayesian trees. Although these differences were small, both topologies were used in our further analyses. The use of the pelicans as outgroups did not alter relationships within the core Pelecaniformes. All analyses

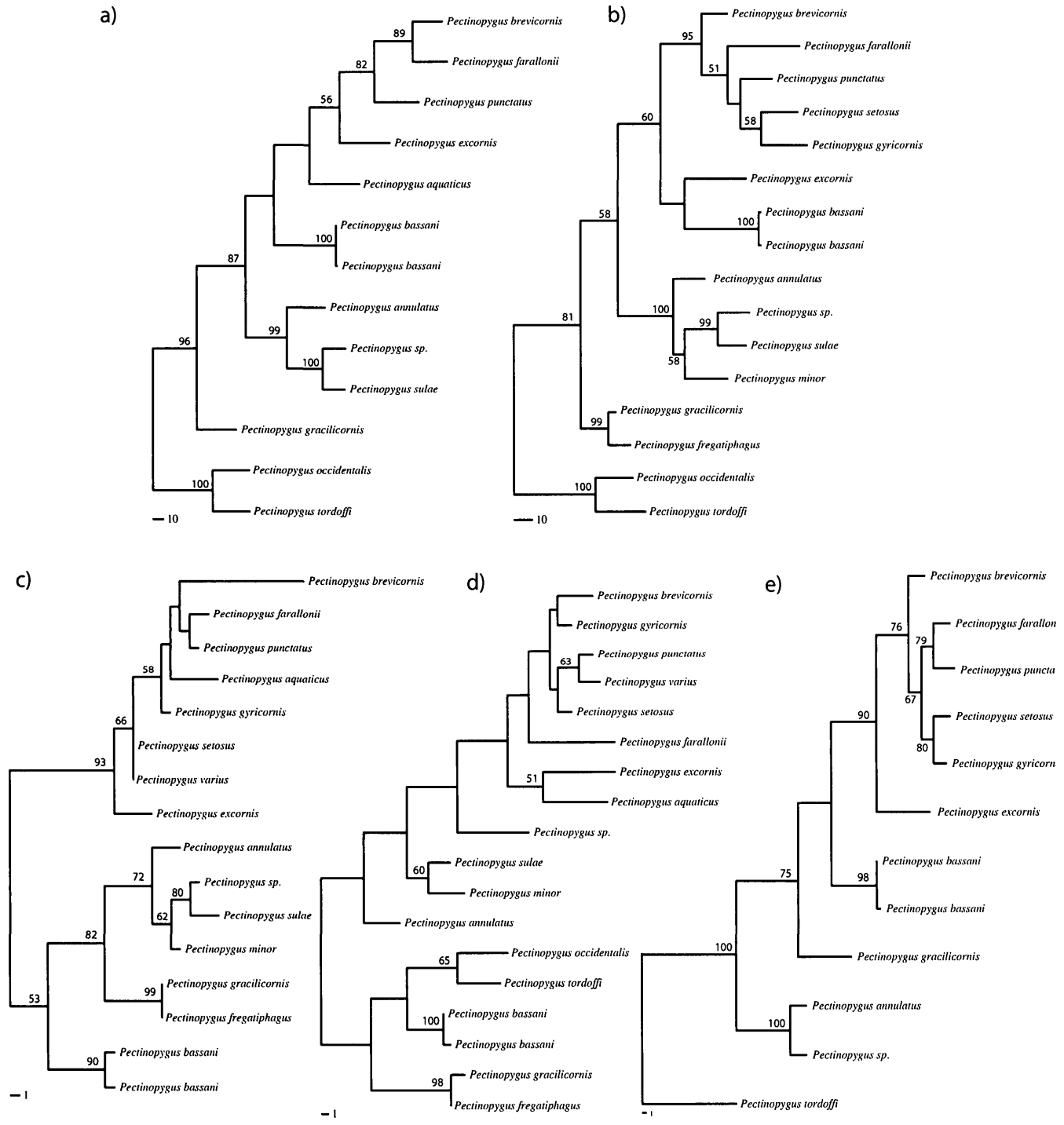

FIGURE 3. Maximum parsimony trees for each gene lice partition. (a) MP analysis of 12S rRNA excluding variable regions (361 characters, 213 parsimony-informative) resulted in 1 MPT (length of 880, CI = 0.574, RI = 0.520). (b) MP analysis of 16S rRNA excluding variable regions (382 characters, 184 parsimony-informative) resulted in 1 MPT (length of 647, CI = 0.564, RI = 0.609). (c) MP analysis of EF1-α (347 characters, 44 parsimony-informative) resulted 39 MPTs (length of 115, CI = 0.730, RI = 0.817). (d) MP analysis of COI excluding 2nd and 3rd position (125 characters, 45 parsimony-informative) resulted in 13 MPTs (length of 173, CI = 0.428, RI = 0.544). (e) MP analysis of wingless (434 characters, 81 parsimony-informative) resulted in 2 MPTs (length of 222, CI = 0.685, RI = 0.662). Bootstrap support is shown beside the branch.

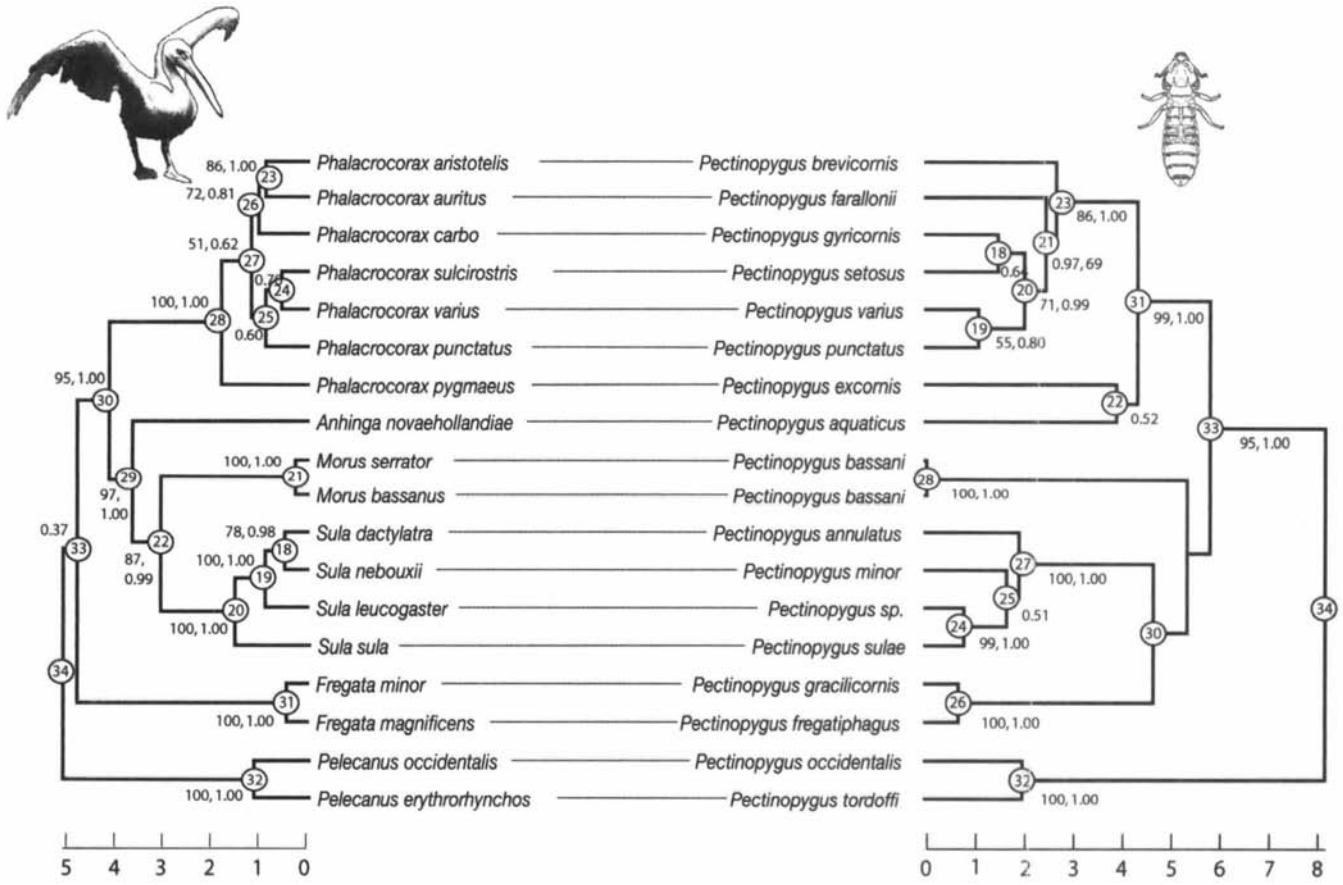

FIGURE 4. Maximum-likelihood phylogeny trees for pelecaniform birds (left) and *Pectinopygus* lice (right) with lines connecting coexisting hosts and parasites. Bird maximum-likelihood tree estimated using GTR+G+I substitution model (−ln likelihood = 9593.36; empirical base frequencies with rate heterogeneity; gamma shape parameter = 1.494; proportion invariable sites = 0.554), ML bootstrap, and posterior probability are beside the circled node label. The *Pectinopygus* maximum-likelihood tree was estimated from all sequences excluding variable regions and 2nd and 3rd position of COI (−ln Likelihood = 10585.38; empirical base frequencies with rate heterogeneity; gamma shape parameter = 0.457; proportion invariable sites = 0.322).

hereafter were performed on the topologies derived from analysis of the combined data of *Pectinopygus* and the combined data of the hosts.

Topology-Based Test of Cospeciation

TreeMap 2.0ß.—TreeMap was performed using all combinations of the ML and MP trees of the combined *Pectinopygus* data with the ML and MP topologies derived from the combined data set of the host. For simplicity we only show the tanglegrams of the MP tree of birds versus the MP tree of *Pectinopygus* (Fig. 2) and the ML tree of the host versus the ML tree of *Pectinopygus* (Fig. 4). The node uniting the two gannets and *P. bassani* are counted as sorting events instead of cospeciation events as there is little divergence in the parasites compared to the hosts. This could not be taken into account directly in TreeMap or TreeFitter as the software cannot deal with one parasite found on several hosts.

Without invoking any host switching, TreeMap had to introduce 11 cospeciation events, 5 duplications, and 19 sorting events (total event cost of 24) to reconcile the MP host tree and the MP parasite tree, and the ML parasite tree required 10 cospeciation events, 6 duplications, and 24 sorting events (total event cost of 30) to be reconciled with the MP host tree. A non-timed analysis (using only the topology obtained through MP) with host switching enabled resulted in 34 optimal reconstructions. The optimal solutions postulated 10 to 11 cospeciations, 5 to 6 duplications, 3 to 19 losses, and 0 to 6 switches for the MP topology. The switch of *Pectinopygus* lice from *Anhinga novaehollandiae* to the *Phalacrocorax* clade or vice versa is consistent with 18 out of 34 reconstructions produced in TreeMap using the MP topology. Twelve other reconstructions suggest a switch to *Anhinga novaehollandiae* from *Phalacrocorax pygmaeus*.

The number of optimal solutions increases to 66 when the pelican associations are removed, with cospeciation events ranging from 7 to 9, 5 to 7 duplications, 2 to 19 losses, and 0 to 7 switches with all reconstructions remaining significant in the randomization test. When using the ML topology of the host, the number of cospeciation events did not change for the MP louse topology and decreased slightly to 8 to 9 cospeciations for the ML lice tree. The randomization test on the complete data set suggested that the global fit between the ML host tree and the ML parasite tree was significant

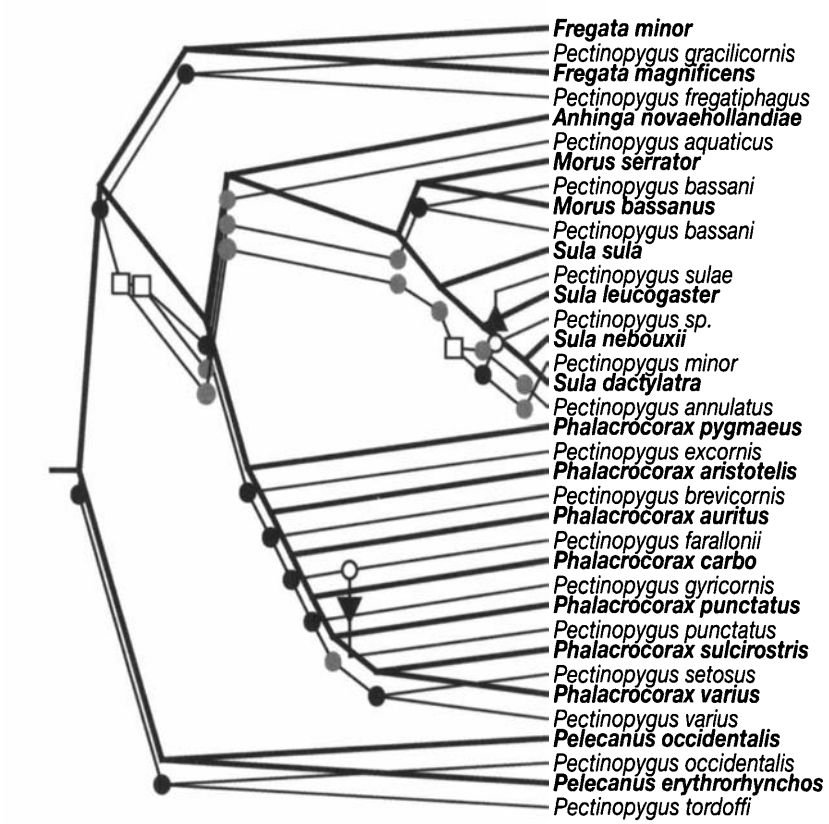

FIGURE 5. One of 34 potentially optimal reconciled trees from TreeMap between the MP bird tree and the MP *Pectinopygus* tree (12 codivergences, 2 switches, 5 duplications, 12 losses, total cost = 24). Thick and thin lines represent the host and associate trees, respectively. Four evolutionary events are denoted: codivergence (black circles), duplications (empty square), loss (grey circle), and switching (arrows).

for all reconstructions (Fig. 5). Analyses with the ML host and parasite trees restricted to the core Pelecaniformes increased the number of optimal solutions to 42, with the number of cospeciation events between 6 and 7, 7 to 8 duplications, 11 to 31 losses, and 0 to 5 switches. The reconciliation of the host MP tree and *Pectinopygus* MP tree involves the largest number of inferred cospeciation events, and reconciliation of the two ML trees involves the smallest number of codivergence events. These two sets of trees were used in all further analyses.

TreeFitter.—The analysis performed with TreeFitter 1.0, using the default settings (C = 0, D = 0, S = 1, H = 2) on the MP bird tree and MP associate tree, suggests that there is phylogenetic structure in this association. The fit between the host and parasite phylogenies, tested by permutation, shows that the overall cost is significantly lower than expected by chance alone (cost = 14, P = 0.0001, permutation = 1000). In all reconstructions, there were 10 codivergences, 0 duplications, 3 sorting events, and 6 switches. Thus, the main factor contributing to this fit is a relatively small number of switching events (P < 0.001) and large number of cospeciations (P < 0.001). Identical results were found when using the same cost settings as in TreeMap (C = 0, D = 1, S = 1, H = 1). Applying maximum codivergence (C = −1, D = 0, S = 0, and H = 0), the global fit between the two trees was still significant (P = 0.0001). If cospeciation and sorting are made almost impossible via a very high cost (Fitch optimization, C = 10, D = 0, S = 10, and H = 1), the significant values disappear, confirming the signal of cospeciation in the present host-parasite system. When sorting and switching costs were varied among 0.5, 1.0, and 2.0 (maximum of fourfold difference), overall fit was still significant, suggesting that these results are not sensitive to cost settings (Table 1). When reconciling the ML

TABLE 1. The results of the TreeFitter analysis under various cost settings. The test was performed between the MP tree of the host and the MP tree of *Pectinopygus* and between the ML tree of the host and the ML tree of *Pectinopygus*. The combined data were used to construct the host MP and ML topologies (12S, CytB and ATPase) and the parasite MP and ML topologies (12SnoVar, 16SnoVar, COI excluding 2nd and 3rd codon positions, EF1-α, wingless). Probabilities are based on 999 permutations. The asterisks represent significance at 5%.

Cost settings				Probability	
Cospeciation	Duplication	Sorting	Host shift	ML topology	MP topology
0	0	1	2	<0.001*	<0.001*
0	0	0.5	2	<0.001*	<0.001*
0	0	1	1	0.004*	<0.001*
0	0	2	0.5	0.0031*	0.004*
0	0	2	1	0.0037*	<0.001*
100	0	1	2	1.00	1.00
0	100	1	2	<0.001*	<0.001*
0	0	100	2	0.0034*	0.0034*
0	0	1	100	<0.001*	<0.001*

bird tree with the ML parasite tree, the cost increased to 19 (cost = 19, $P = 0.0001$, permutation = 1000). In all the reconstructions, there were 8 codivergences, 0 duplications, 4 sorting events, and 8 switches. When applying maximum codivergence, the global fit between the two ML trees was no longer significant ($P = 0.1419$). In all analyses, the correlations became non-significant only when cospeciation was prevented by assigning it a very high cost (100). This was not observed when the three other events were prevented suggesting that the significant correlation is solely a result of the number of cospeciation events (Table 1).

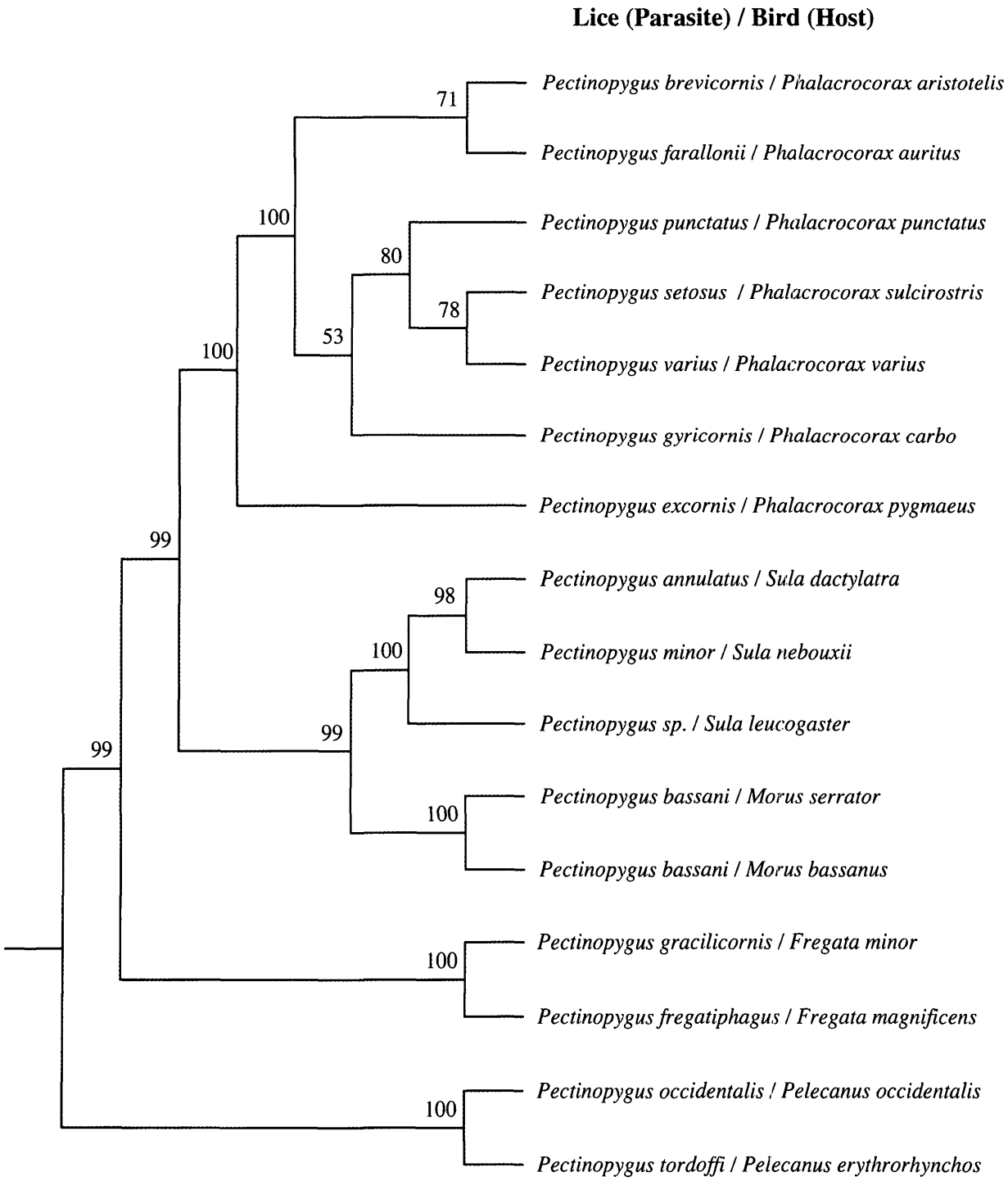

FIGURE 6. Combined evidence topology derived by combining bird and louse data and excluding host-parasite associations that show significant incongruence between the bird and louse data sets as measured by the ILD test (length of 3206, CI = 0.587, RI = 0.618). Numbers above branches are bootstrap support values from the combined analysis. The support for these nodes has increased from Figure 2 as it is based on both the *Pectinopygus* data and the bird data.

Data- and Distance-Based Analyses of Cospeciation

ILD test of cospeciation.—Congruence between the host and parasite gene partitions was determined using the data based parsimony test (ILD) following the method of Johnson et al. (2001). The bird and louse data sets were incongruent ($P < 0.01$) but became congruent ($P = 0.14$) after the removal of two host-parasite associations (*Pectinopygus aquaticus–Anhinga novaehollandiae*, *Pectinopygus sulae–Sula sula*). The removal of the pelicans and their associated lice from this analysis did not affect the results, although the significance of the congruence was decreased ($P = 0.07$). Figure 6 shows the combined evidence topology derived from bird and *Pectinopygus* data excluding the host-parasite associations causing the incongruence. After adding back these associations to the constrained combined evidence topology, one MPT was found for *Pectinopygus* and one MPT was recovered for the birds (Fig. 7).

ParaFit.—Because TreeFitter and TreeMap require fully resolved trees and thus are sensitive to the selection of different optimal trees, we also performed the ParaFit analysis using uncorrected distances calculated from the combined data sets (Table 2). The global test using ParaFit corroborated results obtained by TreeFitter and TreeMap that there is a significant correlation between *Pectinopygus* and their host birds (ParaFitGlobal = 0.01956, $P = 0.0001$, and in the core Pelecaniformes, ParaFitGlobal = 0.01242, $P = 0.0001$). However, the test of individual links (Table 2) indicated that not all louse-pelecaniform associations contribute to the global fit between the two data sets. These results show that some associations, such as those between *Pectinopygus excornis*, *Pectinopygus aquaticus*, *Pectinopygus setosus*, *Pectinopygus varius*, and their respective hosts, do not contribute to the overall fit between birds and lice. When using the phylogenetic trees expressed as distance matrices by calculating patristic distances among the species, the global test of cospeciation revealed a global association between hosts and parasites (ParaFitGlobal = 2.2, $P = 0.0001$ for the MP trees and ParaFitGlobal = 1.8, $P = 0.0001$ for the ML trees). Considering the individual host-parasite links, only 4 of 18 were not significant for the uncorrected distances and the patristic distances from the ML trees and only 2 of 18 for patristic distances from the MP trees. The links *Pectinopygus varius–Phalacrocorax varius* and *Pectinopygus aquaticus–Anhinga novaehollandiae* were not significant in any of these analyses. Nonetheless, these analyses suggest that an important amount of cospeciation between *Pectinopygus* and seabirds is present irrespective of the phylogenetic reconstruction.

LRT.—Because the results of TreeMap, TreeFitter, and ParaFit indicated a significantly better fit between lice and seabirds than expected by chance, we tested the hypothesis that the same topology underlies lice and bird phylogenies. The results of the SH test indicated that there is significant disagreement between the lice gene

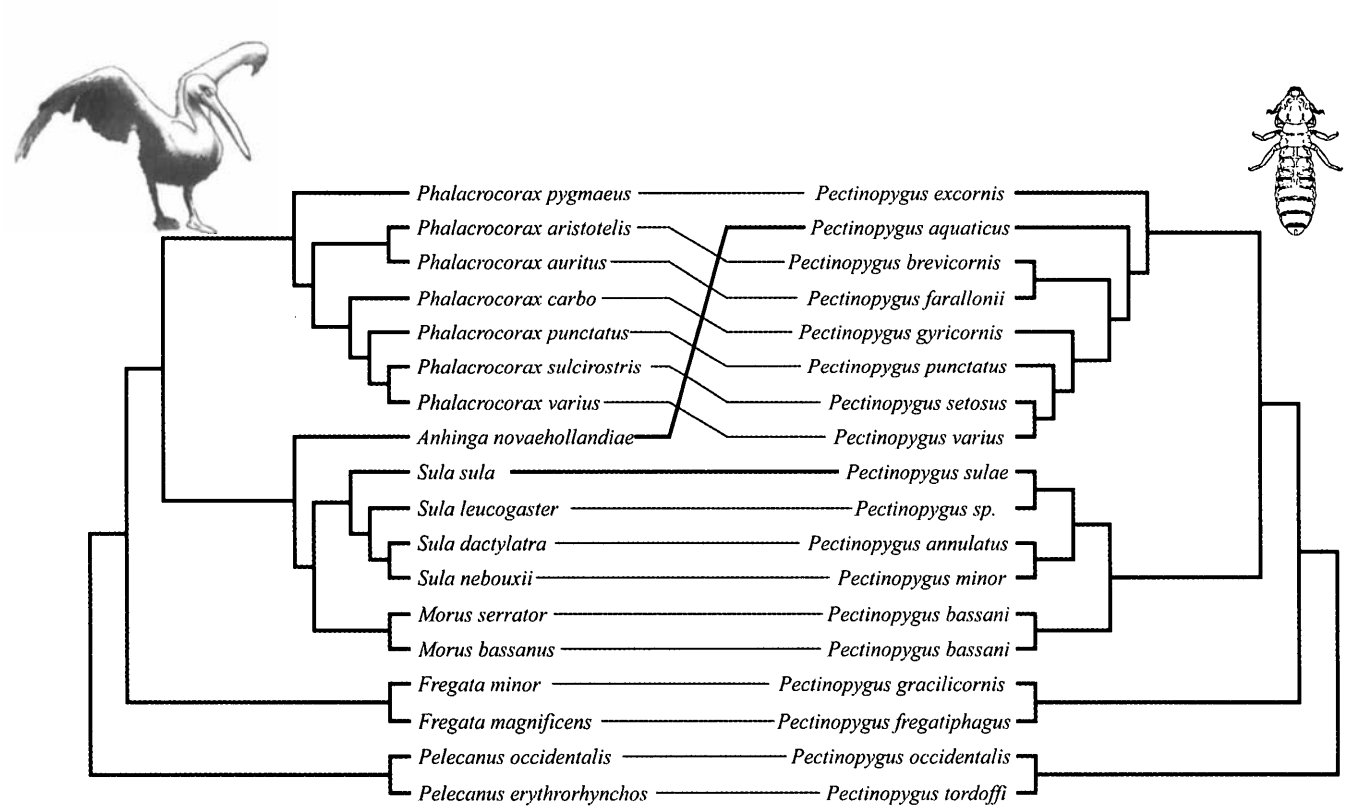

FIGURE 7. Phylogenies for birds (length of 1731, CI = 0.540, RI = 0.634) and lice (length of 1855, CI = 0.554, RI = 0.562) after reintroducing the taxa showing conflict to the constrained combined evidence topology. Bold lines indicate the host-parasite associations that are significant sources of incongruence between the bird and louse data sets.

TABLE 2. The result of ParaFit analyses conducted using raw distances, patristic distances of the MP trees, and patristic distances of the ML trees. Topologies were built from the combined data sets as in Table 1. Probabilities are based on 999 permutations. The global fit (in bold) indicates significant associations between hosts and parasites. Significant P-values (in bold) suggest that the link under evaluation has a significant contribution to the global fit (Legendre et al., 2002). Probabilities in bold are significant at a level of 5%.

		Probability		
Pectinopygus species	Host species	Under uncorrected distances	Under MP patristic distances	Under ML patristic distances
Pectinopygus brevicornis	*Phalacrocorax aristotelis*	**0.009**	**0.0186**	**0.0014**
Pectinopygus farallonii	*Phalacrocorax auritus*	**0.0225**	**0.0133**	**0.0031**
Pectinopygus punctatus	*Phalacrocorax punctatus*	**0.0091**	**0.0019**	**0.0081**
Pectinopygus excornis	*Phalacrocorax pygmaeus*	0.0912	0.127	0.0911
Pectinopygus annulatus	*Sula dactylatra*	**0.0038**	**0.0005**	**0.0001**
Pectinopygus	*Sula leucogaster*	**0.0091**	**0.0006**	**0.0004**
Pectinopygus sulae	*Sula sula*	**0.0184**	**0.0034**	**0.0029**
Pectinopygus aquaticus	*Anhinga novaehollandiae*	0.3663	0.8229	0.8816
Pectinopygus occidentalis	*Pelecanus occidentalis*	**0.0028**	**0.0038**	**0.0016**
Pectinopygus tordoffi	*Pelecanus erythrorhynchos*	**0.003**	**0.0032**	**0.0011**
Pectinopygus gracilicornis	*Fregata minor*	**0.0148**	**0.0359**	0.37
Pectinopygus bassani	*Morus serrator*	**0.0326**	**0.0336**	**0.0277**
Pectinopygus bassani	*Morus bassanus*	**0.027**	**0.0346**	**0.0273**
Pectinopygus fregatiphagus	*Fregata magnificens*	**0.0062**	**0.035**	0.3659
Pectinopygus minor	*Sula nebouxii*	**0.0012**	**0.0002**	**0.0001**
Pectinopygus setosus	*Phalacrocorax sulcirostris*	0.0647	**0.0016**	**0.0078**
Pectinopygus gyricornis	*Phalacrocorax carbo*	**0.0385**	**0.0048**	**0.0129**
Pectinopygus varius	*Phalacrocorax varius*	0.3653	**0.0011**	**0.0072**
Global test		**0.0001**	**0.0001**	**0.0001**

trees (12SnoVar, 16SnoVar, COI excluding 2nd and 3rd positions, EF1-α, and wingless) and the bird combined data set (Table 3). Moreover, because none of the simulated values of δ exceeded the observed value, the null hypothesis of identical topologies given the host and parasite data can be rejected. This level of data set heterogeneity is significant, as evidenced by a δ-value (42.02) that is much higher than any value in the null distribution generated by Monte Carlo simulation of data sets (null distribution ranges from 0.09 to 27.04). If we specifically consider the bird data set, we can ask whether the bird topology is less consistent with louse genes than louse genes are with one another. The exclusion of individual genes from the calculation of δ shows that the combined bird data contributes most to the phylogenetic conflict. Thus, the louse genes are more consistent with one another ($\delta = 13.95$) than they are with the combined bird data. The rejection of the null hypothesis of identical trees underlying host and parasites indicates that there is more than systematic error occurring and, thus, differences between host and parasite phylogenies are probably the result of other historical events such as host switching or extinction.

Correlation of coalescence times.—Figure 8a is a plot of the correlation between coalescence times for the MP trees measured using MULTIDIVTIME. The intercept of the slope is positive but not significantly different from zero and the coalescence times are significantly correlated (Fig. 8a, $r = 0.94$ for the MP tree, $P < 0.01$; and, Figure 8b, $r = 0.92$ for the ML tree, $P < 0.01$), suggesting

TABLE 3. ln Likelihoods of *Pectinopygus* and bird trees under alternative data sets for nine species (listed in Appendix 3) for which all the genes are present calculated for the likelihood-ratio test. Significance levels of differences between likelihoods were tested using the Shimodaira-Hasegawa test. Significant results (asterisks) indicate that the score of the best tree for a given data set is significantly higher than the scores of the optimal trees based on alternative data sets. The results obtained using likelihood topologies are presented. L_{sum}, summed likelihoods across all six data sets constrained to a given topology; $L_0 = -13{,}350.86$, the summed ln likelihoods across data sets for the bird topology ; L_1, the sum of the maximum-likelihood scores for each data set (along diagonal) = −13,329.85905; the likelihood-ratio test statistic (δ) for the inclusion of all data sets = 2[(−13,329.85) − (−13,350.86)] = 42.02. The contribution of individual loci to the observed data heterogeneity was calculated by excluding individual genes.

	Tree						
	Pectinopygus					Host	δ (Excluding
Data sets	12S	16S	COI	EF1-α	wingless	Combined	individual genes)
Pectinopygus 12S	**−1955.62**	−1958.01*	−1963.71*	−1957.49*	−1958.01*	−1957.26*	36.4
Pectinopygus 16S	−1754.92*	**−1748.59**	−1778.18	−1752.31*	−1748.59*	−1755.56*	28.07
Pectinopygus COI	−2045.2*	−2042.74*	**−2038.79**	−2046.9	−2042.74*	−2045.07*	29.46
Pectinopygus EF1-α	−831.32*	−829.27*	−835.67	**−828.64**	−829.27*	−830.78*	37.73
Pectinopygus wingless	−1093.26*	−1088.89*	−1107.73	−1091.85*	**−1088.89**	−1092.87*	34.04
Combined bird	−5680.86	−5682.92	−5682.91	−5680.86	−5682.92	**−5669.3**	13.95
L_{sum}	−13,361.2	−13,350.45	−13,407.02	−13,358.07	−13,350.45	13,350.86	42.02

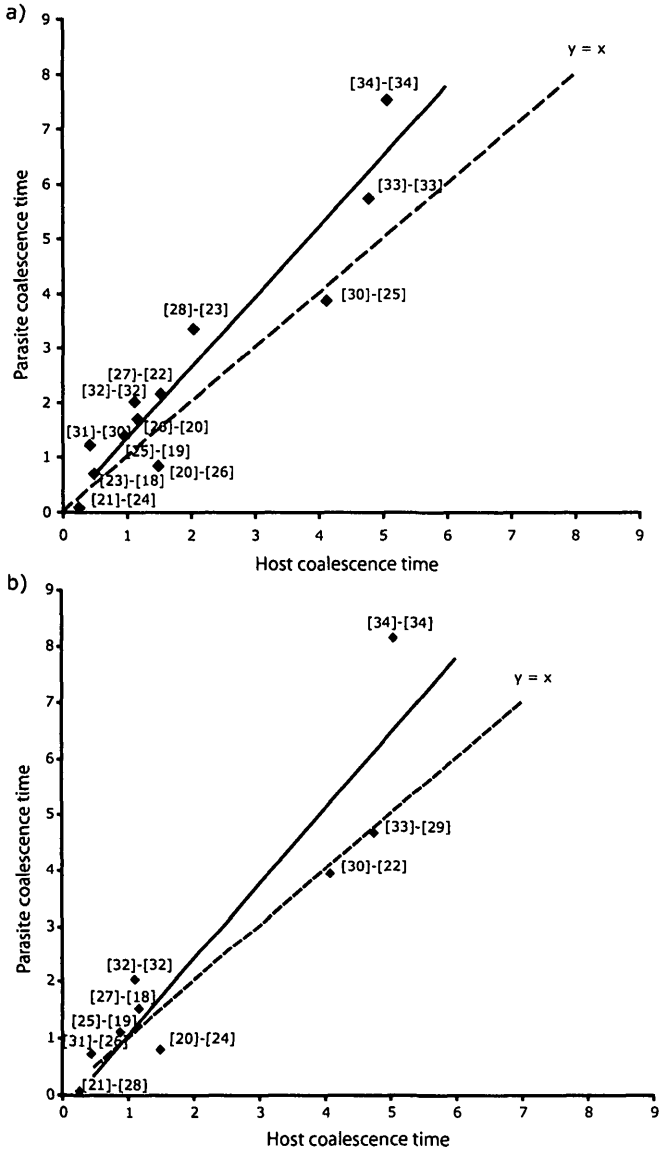

FIGURE 8. (a) Plot of estimated coalescence times from MULTIDIVTIME along copaths in the lice and bird MP phylogenies as shown in Figure 2. Data labels refer to node numbers in Figure 2. The regression line found using the reduced major axis method has a slope of 1.29 ($r^2 = 0.895$). (b) Coalescence times along copaths in the lice and bird ML phylogenies as shown in Figure 4. Data labels refer to node numbers in Figure 4. The regression line found using the reduced major axis method has a slope of 1.35 ($r^2 = 0.852$).

that there has been synchronous cospeciation. The plot does show some scatter and the slope of the regression is not equal to 1. However, removal of the basal node [34]-[34] from the reduced major axis analyses, which suggests that the coalescence time is greater for the basal lice than the birds, results in a slope that approaches 1 with fewer outliers.

DISCUSSION

A phylogenetic approach has proved to be invaluable for coevolutionary studies (e.g., Hafner et al., 1994; Page et al., 1998; Paterson and Gray, 1996). However, most coevolutionary methods presuppose that the "true" host and parasite trees are known. For small numbers of taxa, we can enumerate all possible cladograms (either explicitly or implicitly using a branch-and-bound search); but for larger trees, as in this study, we must rely on heuristics. Thus, the usefulness of most coevolutionary studies is limited by the accuracy of the phylogenies on which they are based. Finding reliable estimates of species phylogenies is a long-studied and challenging problem (Felsenstein, 2003; Semple and Steel, 2003) integral to the effective study of coevolutionary patterns. The problems of establishing accurate molecular phylogenies for insect orders are well known and for lice these problems appear to be even greater (Banks et al., 2006; Cruickshank et al., 2001; Johnson et al., 2003). Previous attempts to reconstruct louse phylogenies using various nuclear and mitochondrial markers have produced conflicting trees with poor phylogenetic resolution and support particularly at a basal level (Cruickshank et al., 2001; Johnson et al., 2003).

Here, we have used a data set of 2 nuclear, 2 rRNA, and 1 mitochondrial markers to elucidate relationships within the genus *Pectinopygus*. The addition of 16S rDNA and wingless to the molecular tool kit and the favorable characteristic of a combined molecular data set promises improved understanding of louse phylogeny for the future and consequently a more accurate appreciation of cospeciation and the origin of a parasitic lifestyle in lice. Several reasons might explain the differences in topology of COI (excluding 2nd and 3rd positions), wingless, EF1-α, 12SnoVar, and 16SnoVar with regards to the unstable position of the gannet lice, such as sampling error with respect to the slightly different set of taxa used in each data set, different stochastic processes acting on the characters, or possibly true differences in branching histories of different genes. Although we chose to combine the data sets, the issue of combining data for phylogenetic analysis is the subject of some debate without a current consensus (Bull et al., 1993; De Queiroz et al., 1995; Huelsenbeck et al., 1996; Miyamoto and Fitch, 1995).

Taxonomic sampling is another important consideration, both for accurate estimation of amounts of molecular evolution (Fitch and Bruschi, 1987) and for accurate identification of cospeciation events. Indeed, it has been shown that increased taxon sampling is usually more advantageous to accurate tree estimation than adding more characters (Pollock et al., 2002; Zwickl and Hillis, 2002). As a result, our phylogenies of hosts and parasites might change with greater taxon sampling and consequently the number of cospeciation events could change. For example, basal relationships in the louse phylogeny are only weakly supported and result in the inference of a host switch from boobies (*Sula*) to cormorants (*Phalacrocorax*) in the case of the ML lice tree. Further sampling might change these basal relationships and provide further evidence for cospeciation.

The contentious relationship of the pelicans relative to the core Pelecaniformes suggests that the presence of *Pectinopygus* on the core Pelecaniformes may be due to

host switching rather than by decent. Removing the pelicans from the TreeMap analyses reduces the number of cospeciations, but the global fit of the core Pelecaniformes remains significant. However, the significance of the fit is decreased (e.g., in the ILD test and ParaFit) and the number of optimal solutions increases (e.g., TreeMap analyses) when the sample size is decreased; i.e., when only analyzing the core Pelecaniformes, although the presence of significant cospeciation remains unambiguous. Thus, taxon sample size can affect the significance of the fit between the hosts and parasites emphasizing the need for further sampling in this group.

The *Pectinopygus* specimen (RF-15) collected on *Sula leucogaster* is a nymph and could not be identified. Thus, cospeciation events might falsely be reconstructed among the *Sula* lice as three different species can be found on *Sula leucogaster* (*P. annulatus*, *P. garbei*, *P. sulae*) and *Sula sula* and *Sula leucogaster* can host the same louse species (Clay, 1964). Here, the *Pectinopygus* species (RF-15) has diverged sufficiently from *P. sulae* and *P. annulatus* to suggest that it is most probably *P. garbei*. Nonetheless, the presence of three different lice on both *Sula leucogaster* and *Sula sula* would suggest host switching and/or little cospeciation for these three louse species. Without sampling all three louse species from both hosts, the question of cospeciation among *Sula* lice will remain unresolved. This further underlines the need to determine clear unambiguous monophyletic clades with extensively sampled taxa for both hosts and parasites when conducting cospeciation analyses.

To date phylogenetic congruence is imperfect or absent in most of the louse–bird interactions that have been studied (Johnson et al., 2002; Weckstein, 2004; Banks et al., 2006). Examples of significant cospeciation have, however, been found in systems where host-switching is prevented by the asocial lifestyle of the host and the low mobility of the parasite. Examples include the rodent-lice (Hafner and Nadler, 1988; Hafner and Page, 1995), albatross-lice (Page et al., 2004), and insect-symbiont associations where the bacteria, needed for reproduction, are transmitted maternally (Clark et al., 2000). This study can be added to these few examples of extensive cospeciation as all methodologies agree on the presence of significant cospeciation processes in this host-parasite system, but the results are sensitive to selection of different phylogenetic hypotheses and analytical methods for evaluating cospeciation. Indeed, different methods for cospeciation analysis treat differently the evolutionary events occurring in a host–parasite association and consequently produce different results.

TreeMap provided a large number of feasible historical reconstructions although some of the reconstructions that postulate host switches are unlikely due to the geographically distant hosts, reducing the reconstruction's plausibility. For example, *Anhinga novaehollandiae* is found across Australasia and *Phalacrocorax pygmaeus* is found in Europe around the Mediterranean and Middle East making a host switch unlikely, although it is possible that the hosts once had overlapping distributions. Alternatively, a host switch may have taken place between unsampled species. For example, if we postulate that darters have very closely related lice, a host switch between *Phalacrocorax pygmaeus* and the Asian/Oriental Darter (*Anhinga melanogaster*) could be an equally likely explanation for the TreeMap reconstructions. Thus, unsampled taxa may help to explain some geographically implausible switches and incorporating distributional information of the hosts and parasites into the current cophylogenetic methodologies could help decrease the uncertainties in the estimates of cospeciation.

Lineage sorting in the parasites cannot be ruled out and may be confounding the results. Sorting events occur when parasite species go extinct from a host species or when a founding host population, by chance, does not carry a parasite species found in the parent population ("missing the boat"). In our TreeMap and TreeFitter analyses, sorting events (probably from parasites missing the boat) appear to be quite common. TreeMap found 11 to 31 sorting events using the ML topologies (3 to 19 with MP), whereas TreeFitter found 4 (3 with MP). For example, in the reconstruction shown (Fig. 5), a combination of duplications and lineage sorting is required to explain the presence of *P. bassani* on both gannets species (*M. bassanus* and *M. serrator*). Frequent sorting events in birds and lice is probably due to the patchy distribution of lice and the small host population size during speciation events (Paterson et al., 1999).

Unlike ParaFit, TreeMap and TreeFitter cannot deal with parasites found on several different host species (e.g., *P. bassani*) and thus could underestimate host switching. Both ParaFit and the correlation of coalescence times take into account the divergence of the hosts and parasites as well as their cladistic relationships. These combined approaches appear to be less influenced by alternative hypotheses of host and parasite relationships due to the different reconstruction methods. The ParaFit analyses identified 14 to 16 hosts and parasites that have probably undergone cospeciation and suggested that *P. excornis*, *P. aquaticus*, *P. setosus*, and *P. varius* are most likely to have been subjected to host-switching or sorting events (parasite extinction or primary absence on daughter host lineage). The uncorrected distances computed directly from the raw data provided fewer significant links between hosts and parasites than using the patristic distances for the ParaFit analysis, thus suggesting that taking topology into account can increase the inferred number of cospeciation events (e.g., in the case of the MP patristic distances).

Although we can reject perfect cospeciation as shown by the ILD test of cospeciation and the LRT, it seems that only two host–parasite associations (*Pectinopygus aquaticus–Anhinga novaehollandiae* and *Pectinopygus sulae–Sula sula*) are causing the incongruence as determined by the ILD test. *P. aquaticus* is also implicated in ParaFit as a taxon subject to host-switching, whereas *P. sulae* is not. These differences in host-parasite associations contributing to the global fit (14 in ParaFit using raw distances and 16 in the ILD test) is likely due to the differences in methodologies: the distance method underlying ParaFit versus the character-based approach of the ILD

test of cospeciation. The lack of perfect congruence in the ILD test and LRT likely resulted from biologically meaningful conflict such as lineage sorting events or occasional host switching, although systematic error due to different loci undergoing different evolutionary histories (recombination or different modes of inheritance) cannot be rejected.

This study emphasizes the need to be cautious with topology-based approaches. Whereas the data-based cospeciation analysis (ILD) and ParaFit using uncorrected distances are not affected by the selection of outgroups or the methods of tree building, with the topology-based approaches (TreeMap, TreeFitter, ParaFit with patristic distances) there is often uncertainty in the phylogenetic inferences used for coevolutionary analyses, thus making it difficult to determine the exact number of cospeciations events. However, phylogenies based on multiple data partitions can increase overall support for a topology due to emergent properties not evident in the separate analyses of individual partitions. As a result, the topology-based approaches based on combined data partitions might be less influenced by the selection of genes than the LRT approach that specifically tests for conflict among loci. Thus, aside from the problems associated with outgroup selection and incomplete data sets as a result of difficulties associated with amplification of certain gene fragments or with sampling certain species, the different rates and modes of molecular evolution of the loci in different organisms affects the recovery of congruence between hosts and parasites through the selection of certain genes and methods of tree building.

The significant regression of coalescence times in *Pectinopygus* and Pelecaniformes strongly supports cospeciation despite the effect of the later coalescence time for the basal node of *Pectinopygus*. This relationship permits inferences about the age of the fossilless ischnoceran lice. Cormorants (Phalacrocoracidae) are known from the Upper Oligocene in Germany and France and the Quercy fissure fillings (reviewed in Mayr, 2005). The fossil from France may even be outside the clade Phalacrocoracidae + Anhingidae. Thus, this finding would date the cormorant lice conservatively at approximately 29.3 Mya. Reliable records of extant Pelecaniformes from deposits older than the Oligocene only exist for the Fregatidae described from the Lower Eocene, 50 to 56.5 Mya (Olson, 1977). As diversification of the crown-groups within modern avian families did not take place before the Oligocene (35.4 Mya; Mayr, 2005) and the *Pectinopygus* lice cospeciated with these families, the diversification of the pelican, gannet, booby, cormorant, darter, and frigate lice groups are unlikely to be older than 35.4 Mya, although the genus *Pectinopygus* as a whole is probably older than 50 Mya.

Despite the interest in cospeciation and the development of a range of methodologies, there are still remarkably few unambiguous examples of cospeciating species (e.g., Clark et al., 2000; Hafner and Nadler, 1988; Hafner and Page, 1995; Page et al., 2004). *Pectinopygus*-Pelecaniformes bird assemblages is a paradigm example of a parasite closely tracking its host where cospeciation is influencing the diversification of lice. Nonetheless, the available data suggest that at least some lice have switched host, so that a combination of vertical and horizontal transmission must be postulated. However, the exact number of cospeciation events varies according to the cophylogenetic methodologies and the host and parasite topologies used. Unfortunately, phylogenies are only known by the estimates that have been worked out by researchers and these estimates may be imperfect. We now await improved cophylogenetic methods that take into account these uncertainties in phylogenetic reconstructions.

ACKNOWLEDGMENTS

We thank Jeff Thorne for guidance in the use of MULTIDIVTIME and Pierre Legendre for help with ParaFit. We are also grateful to Vince Smith for information on the ecology and behavior of *Pectinopygus* and to Terry Chesser, R. Cotter, Richard Evans, Robert Furness, Richard Griffiths, Nancy Hoffman, Kevin McCracken, Brent Stephenson, Stefano Volponi, Jan Walker, and Jason Weckstein for assistance in providing specimens. We thank the following people and institutions for supplying bird samples: A. Baker, L. Chadderton, D. Dittmann, C. Duffy, B. Gill, R. Griffiths, S. Hackett, O. Haddrath, K. Orr, S. Volponi, Auckland Institute and Museum, Field Museum, Museum of Natural Science at Louisiana State University, New Zealand Department of Conservation, and Royal Ontario Museum. KPJ was supported by NSF grants DEB-0107891 and DEB-0118794, MK was supported by a New Zealand Foundation for Research, Science and Technology Postdoctoral Fellowship, and JH was supported by NERC grant NE/B000079/1 to RDMP. We also thank Mark Hafner and three anonymous reviewers for their useful comments and suggestions.

REFERENCES

Askew, R. R. 1971. Parasitic insects. American Elsevier Publishing, New York.

Banks, J. C., R. L. Palma, and A. M. Paterson. 2006. Cophylogenetic relationships between penguins and their chewing lice. J. Evol. Biol. 19:156–166.

Brooke, M. and T. Birkhead (eds.). 1991. The Cambridge encyclopedia of birds. Cambridge University Press, New York.

Buckley, T. R., C. Simon, P. K. Flook, and B. Misof. 2000. Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. Insect Mol. Biol. 9:565–980.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

Charleston, M. A. 1998a. Jungles: A new solution to the host/parasite phylogeny reconciliation problem. Math. Biosci. 149:191–223.

Charleston, M. A. 1998b. Spectrum: Spectral analysis of phylogenetic data. Bioinformatics 14:98–9.

Clark, M. A., N. A. Moran, P. Baumann, and J. J. Wernegreen. 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution 54:517–525.

Clay, T. 1964. Geographical distribution of the Mallophaga (Insecta). Bull. Brit. Orni. Club 84:14–16.

Clayton, D. H., R. D. Gregory, and R. D. Price. 1992. Comparative ecology of Neotropical bird lice (Insecta, Phthiraptera). J. Anim. Ecol. 61:781–795.

Cracraft, J., F. K. Barker, M. Braun, J. Harshman, G. J. Dyke, J. Feinstein, S. Stanley, A. Cibois, P. Schikler, P. Beresford, J. García-Moreno, M. D. Sorenson, T. Yuri, and D. P. Mindell. 2004. Phylogenetic relationships among modern birds (Neornithes). Pages 468–489 *in* Assembling the tree of life (J. Cracraft and M. J. Donoghue, eds.). Oxford University Press, New York.

Cruickshank, R. H., K. P. Johnson, V. S. Smith, R. J. Adams, D. H. Clayton, and R. D. M. Page. 2001. Phylogenetic analysis of partial

sequences of elongation factor 1 alpha identifies major groups of lice (Insecta: Phthiraptera). Mol. Phylogen. Evol. 20:326–326.
Danforth, B. N., S. G. Brady, S. D. Sipes, and A. Pearson. 2004. Single-copy nuclear genes recover Cretaceous-age divergences in bees. Syst. Biol. 53:309–326.
De Queiroz, A., M. J. Donoghue, and J. Kim. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.
Dyke, G. J., and M. Van Tuinen. 2004. The evolutionary radiation of modern birds (Neornithes): Reconciling molecules, morphology and the fossil record. Zoo. J. Linn. Soc. 141:153–177.
Ericson, P. G. P., C. L. Anderson, Britton, T., A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon and G. Mayr. 2006. Diversification of Neoaves: Integration of molecular sequence data and fossils. Biol. Lett. 2:543–547.
Estabrook, G. F. 1992. Evaluating undirected positional congruence of individual taxa between 2 estimates of the phylogenetic tree for a group of taxa. Syst. Biol. 41:172–177.
Faith, D. P. 1991. Cladistic permutation tests for monophyly and non-monophyly. Syst. Zool. 40:366–375.
Faith, D. P., and P. S. Cranston. 1991. Could a cladogram this short have arisen by chance alone—On permutation tests for cladistic structure. Cladistics 7:1–28.
Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1994. Testing significance of incongruence. Cladistics 10:315–319.
Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1995. Constructing a significance test for incongruence. Syst. Biol. 44:570–572.
Felsenstein, J. 2003. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts.
Fitch, W. M., and M. Bruschi. 1987. The evolution of prokaryotic ferredoxins with a general method correcting for unobserved substitutions in less branched lineages. Mol. Biol. Evol. 4:381–394.
Garcia-Moreno, J., M. D. Sorenson, and D. P. Mindell. 2003. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. J. Mol. Evol. 57:27–37.
Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652–670.
Hafner, M. S., and S. A. Nadler. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. Nature 332:258–259.
Hafner, M. S., and S. A. Nadler. 1990. Cospeciation in host-parasite assemblages: comparative analysis of rates of evolution and timing of cospeciation events. Syst. Zool. 39:192–204.
Hafner, M. S., and R. D. Page. 1995. Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system. Phil. Trans. R. Soc. Lond. B Biol. Sci. 349:77–83.
Hafner, M. S., P. D. Sudman, F. X. Villablanca, T. A. Spradling, J. W. Demastes, and S. A. Nadler. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265:1087–90.
Harrison, P. 1983. Seabirds—An identification guide. Croom Helm, Beckenham, UK.
Hellenthal, R. A., and R. D. Price. 1991. Biosystematics of the chewing lice of pocket gophers. Annu. Rev. Entomol. 36:185–203.
Hendy, M. D. 1993. Spectral-analysis of phylogenetics data. J. Classif. 10:5–24.
Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83:189–195.
Huelsenbeck, J. P., and J. J. Bull. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. 45:92–98.
Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11:152–158.
Huelsenbeck, J. P., B. Rannala, and B. Larget. 2000. A Bayesian framework for the analysis of cospeciation. Evolution 54:352–364.
Huelsenbeck, J. P., B. Rannala, and B. Larget. 2003. A statistical perspective for reconstructing the history of host-parasite associations. Pages 93–119 *in* Tangle trees: Phylogeny, cospeciation, and coevolution (R. D. M. Page, ed.). The University of Chicago Press, Chicago, Illinois.
Huelsenbeck, J. P., B. Rannala, and Z. H. Yang. 1997. Statistical tests of host-parasite cospeciation. Evolution 51:410–419.
Johnson, K. P., R. J. Adams, and D. H. Clayton. 2002. The phylogeny of the louse genus *Brueelia* does not refect host phylogeny. Biol. J. Linn. Soc. 77: 233–247.
Johnson, K. P., and D. H. Clayton. 2003 Coevolutionary history of ecological replicates: Comparing phylogenies of wing and body lice to columbiform hosts. Pages 262–286 *in* Tangle trees: Phylogeny, cospeciation, and coevolution (R. D. M. Page, ed.). The University of Chicago Press, Chicago, Illinois.
Johnson, K. P., R. H. Cruickshank, R. J. Adams, V. S. Smith, R. D. M. Page, and D. H. Clayton. 2003. Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). Mol. Phylogen. Evol. 26:231–242.
Johnson, K. P., D. M. Drown, and D. H. Clayton. 2001. A data based parsimony method of cophylogenetic analysis. Zool. Scripta 30:79–87.
Kennedy, M., R. D. Gray, and H. G. Spencer. 2000. The phylogenetic relationships of the shags and cormorants: Can sequence data resolve a disagreement between behavior and morphology? Mol. Phylogen. Evol. 17:345–359.
Kennedy, M., B. R. Holland, R. D. Gray, and H. G. Spencer. 2005. Untangling long branches: Identifying conflicting phylogenetic signals using spectral analysis, neighbor-net, and consensus networks. Syst. Biol. 54:620–633.
Kennedy, M., and H. G. Spencer. 2004. Phylogenies of the frigatebirds (Fregatidae) and Tropicbirds (Phaethonidae), two divergent groups of the traditional order Pelecaniformes, inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 31:31–38.
Legendre, P., and M. J. Anderson. 1998. Program DISTPCOA. Département de sciences biologiques, Université de Montréal.
Legendre, P., Y. Desdevises, and E. Bazin. 2002. A statistical test for host-parasite coevolution. Syst. Biol. 51:217–234.
Lento, G. M., R. E. Hickson, G. K. Chambers, and D. Penny. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. Mol. Biol. Evol. 12:28–52.
Marshall, A. G. 1981. The ecology of ectoparasitic insects. Academic Press, London.
Mayr, G. 2005. The Paleogene fossil record of birds in Europe. Biol. Rev. Camb. Philos. Soc. 80:515–542.
Miyamoto, M. M., and W. M. Fitch. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.
Moran, N., and P. Baumann. 1994. Phylogenetics of cytoplasmically inherited microorganisms of arthropods. Trends Ecol. Evol. 9:15–20.
Olson, S. L. 1977. A Lower Eocene frigatebird from the Green River formation of Wyoming (Pelecaniformes: Fregatidae). Smith. Cont. Paleo. 35:1–33.
Page, R. D. M. 1994a. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. Syst. Biol. 43:58–77.
Page, R. D. M. 1994b. Parallel phylogenies—Reconstructing the history of host-parasite assemblages. Cladistics 10:155–173.
Page, R. D. M. 2003. Tangle trees: Phylogeny, cospeciation, and coevolution. The University of Chicago Press, Chicago, Illinois.
Page, R. D. M., and M. A. Charleston. 1998. Trees within trees: Phylogeny and historical associations. Trends Ecol. Evol. 13:356–359.
Page, R. D. M., R. H. Cruickshank, M. Dickens, R. W. Furness, M. Kennedy, R. L. Palma, and V. S. Smith. 2004. Phylogeny of "Philoceanus complex" seabird lice (Phthiraptera: Ischnocera) inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 30:633–652.
Page, R. D. M., R. Cruickshank, and K. P. Johnson. 2002. Louse (Insecta: Phthiraptera) mitochondrial 12S rRNA secondary structure is highly variable. Insect Mol. Biol. 11:361–369.
Page, R. D., P. L. Lee, S. A. Becher, R. Griffiths, and D. H. Clayton. 1998. A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. Mol. Phylogenet. Evol. 9:276–293.
Paterson, A. M., and R. D. Gray. 1996. Cospeciation of birds and ectoparasites. *in* Coevolutionary biology of birds and parasites. (D. H. Clayton, and J. Moore, eds.). Oxford University Press, Oxford, UK.
Paterson, A. M., R. L. Palma and R. D. Gray. 1999. How frequently do avian lice miss the boat? Implications for coevolutionary studies. Syst. Biol. 48: 214–223.
Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51:664–671.
Posada, D., and K. A. Crandall. 1998. ModelTest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

Price, R. D. 1975. The *Menacanthus eurysternus* complex (Mallophaga: Menoponidae) of the Passeriformes and Piciformes (Aves). Ann. Entomol. Soc. Am. 68:617–622.

Price, R. D., R. A. Hellenthal, R. L. Palma, K. P. Johnson, and D. H. Clayton. 2003. The chewing lice: World checklist and biological overview. Illinois Natural History Survey Special Publication 24.

Rambaut, A. and N. C. Grassly. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238

Rannala, B., and Y. Michalakis. 2003. Population genetics and cospeciation: From process to pattern. Pages 120–143 *in* Tangle trees: Phylogeny, cospeciation, and coevolution (R. D. M. Page, ed.). The University of Chicago Press, Chicago, Illinois.

Ronquist, F. 1995. Reconstructing the history of host-parasite associations using generalised parsimony. Cladistics 11:73–89.

Ronquist, F. 2003. Parsimony analysis of coevolving species associations. Pages 22–64 *in* Tangle trees: Phylogeny, cospeciation, and coevolution (R. D. M. Page, ed.). The University of Chicago Press, Chicago, Illinois.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Semple, C., and M. Steel. 2003. Phylogenetics. Oxford University Press, Oxford, UK.

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

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.

Swofford, D. W. 1998. PAUP*: Phylogenetic reconstruction using parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

Thorley, J. L., and R. D. Page. 2000. RadCon: Phylogenetic tree comparison and consensus. Bioinformatics 16:486–487.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Tompkins, D. M., and D. H. Clayton. 1999. Host resources govern the specificity of swiftlet lice: Size matters. J. Anim. Ecol. 68:489–500.

van Tuinen, M., C. G. Sibley, and S. B. Hedges. 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Mol. Biol. Evol. 17:451–457.

Weckstein, J. D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. Syst. Biol. 53:154–164.

Yoshizawa, K., and K. P. Johnson. 2003. Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. Mol. Phylogenet. Evol. 29:102–114.

Zwickl, D. J., and D. M. Hillis. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.

First submitted 16 October 2006; reviews returned 24 October 2006; final acceptance 8 November 2006
Acting Editor-in-Chief: Mark Hafner

APPENDIX 1. Collection, voucher code, collector, and host species the louse was collected on. All samples are vouchered on slides at the Museum of New Zealand, Te Papa Tongarewa, New Zealand, and the DNA extraction is stored in the Lousebase collection at the University of Glasgow, UK, except for the Pgfre.7.25.2005.5 DNA extraction, which is stored at the Illinois Natural History Survey.

Species	Voucher and extraction code	Collection locality	Collector	Host
Pectinopygus sp.	RF-15	Ilhéu Raso, Cape Verde	R. Furness	*Sula leucogaster*
Pectinopygus annulatus	NZ68	Bahía Darwin, Isla Genovesa, Galápagos Islands, Ecuador	R. Palma	*Sula* dactylatra
Pectinopygus aquaticus	GLA909	Australia	K. Johnson	*Anhinga novaehollandiae*
Pectinopygus bassani	GLA651	Cape Kidnappers, Hawkes Bay, New Zealand	B. Stephenson	*Morus serrator*
Pectinopygus bassani	JJ-09	Nolsoy Island, Faroe Islands	J. Jensen	*Morus bassanus*
Pectinopygus brevicornis	V11	Foula, Shetland Islands, UK	R. Furness	*Phalacrocorax aristotelis*
Pectinopygus excornis	GLA508	Punte Alberete, Ravenna, Emilia-Romagna, Italy	S. Volponi	*Phalacrocorax pygmaeus*
Pectinopygus farallonii	Evans02	Laguna Beach, Orange, California	R. Evans	*Phalacrocorax auritus*
Pectinopygus fregatiphagus	Pgfre.7.25.2005.5	Louisiana, Gulf of Mexico	J. Weckstein	*Fregata magnificens*
Pectinopygus gracilicornis	JD01	Johnston Atoll, Hawaii, Pacific Ocean	J. Donahue	*Fregata minor*
Pectinopygus gyricornis	GLA509	Po Delta, Venice, Rovigo, Italy	S. Volponi	*Phalacrocorax carbo*
Pectinopygus minor	GLA904	Isla Rábida, Galapagos, Galápagos Islands, Ecuador	R. Palma	*Sula nebouxii*
Pectinopygus occidentalis	Evans01	Newport Beach, Orange, California	R. Evans	*Pelecanus occidentalis*
Pectinopygus punctatus	GLA524	Birdlings Flat, Lake Ellesmere, Canterbury, New Zealand	J. Walker	*Phalacrocorax punctatus*
Pectinopygus setosus	GLA484	Hutt River Mouth, Wellington, North Island, New Zealand	R. Cotter	*Phalacrocorax sulcirostris*
Pectinopygus sulae	NZ67	Bahía Darwin, Isla Genovesa, Galápagos Islands, Ecuador	R. Palma	*Sula sula*
Pectinopygus tordoffi	KJ01	Louisiana	J. Weckstein	*Pelecanus erythrorhynchos*
Pectinopygus varius	KJ02	Australia	K. McCracken	*Phalacrocorax varius*
Anaticola crassicornis	JJ-08-2	Nolsoy Island, Faroe Islands	J. Jensen	*Somateria mollissima*

APPENDIX 2. Voucher and collection information for the pelecaniform birds used in this study.

Species	Voucher information	Collection location
Australian Darter	Sample 984, Royal Ontario Museum	Australia
Anhinga novaehollandiae	Sample 826, Museum of Victoria	
Magnificent Frigatebird	Samples 82 and 83, Museum of Natural Science at Louisiana State University	Mexico
Fregata magnificens		
Great Frigatebird	Unvouchered sample or DNA extract	(?)
Fregata minor		
Northern Gannet	Samples 8644, 13349, and 18059, Museum of Natural Science at Louisiana State University	USA
Morus bassanus		
Australasian Gannet	Unvouchered samples or DNA extracts	New Zealand
Morus serrator		
American White Pelican	Samples 388009 and 395693, Field Museum	USA
Pelecanus erythrorhynchos		
Brown Pelican	Samples 5797, 10336, and 16337, Museum of Natural Science at Louisiana State University	USA
Pelecanus occidentalis		
European Shag	Unvouchered samples or DNA extracts	UK
Phalacrocorax aristotelis		
Double-crested Cormorant	Samples 166, 793, and 1285, Royal Ontario Museum	Canada
Phalacrocorax auritus		
Great Cormorant	Samples 90–97 and 90–98, Auckland Institute and Museum	Australia and New Zealand
Phalacrocorax carbo	Sample 956, Museum of Victoria	
Spotted Shag	Sample 91–52, Auckland Institute and Museum	New Zealand
Phalacrocorax punctatus	Unvouchered samples or DNA extracts	
Pygmy Cormorant	Unvouchered samples or DNA extracts	Italy
Phalacrocorax pygmaeus		
Little Black Cormorant	Sample 90–80, Auckland Institute and Museum	Australia and New Zealand
Phalacrocorax sulcirostris	Sample 959, Museum of Victoria and unvouchered sample or DNA extract	
Pied Cormorant	Sample 42490, Museum of Victoria	Australia
Phalacrocorax varius		
Masked Booby	Samples 891, 4037, and 15445, Museum of Natural Science at Louisiana State University	USA
Sula dactylatra		
Brown Booby	Samples 15242, 15418, and 15422, Museum of Natural Science at Louisiana State University	(?)
Sula leucogaster		
Blue-footed Booby	Sample 5753, Museum of Natural Science at Louisiana State University	Mexico
Sula nebouxii		
Red-footed Booby	Samples 15446 and 15447 Museum of Natural Science at Louisiana State University	(?)
Sula sula		

APPENDIX 3. Accession numbers for the genes sequences for the parasitic lice. The species in bold were used for the LRT as they have all the genes present.

Species	12S	16S	COI	Wingless	EF1-α
***Pectinopygus* sp.**	DQ482950	DQ463173	DQ482967	DQ482936	DQ482982
Pectinopygus annulatus	DQ482953	DQ463176	DQ482970	DQ482934	DQ482984
Pectinopygus aquaticus	DQ482944		DQ482957		DQ482973
Pectinopygus bassani	DQ482948	DQ463170	DQ482963	DQ482939	DQ482978
Pectinopygus bassani	DQ482954	DQ463178	DQ482972	DQ482940	AF320443
Pectinopygus brevicornis	AF189142	DQ463179	AF497800	DQ482930	AF320442
Pectinopygus excornis	DQ482947	DQ463168	DQ482961	DQ482933	DQ482976
Pectinopygus gracilicornis	DQ482952	DQ463175	DQ482969	DQ482938	DQ482983
Pectinopygus gyricornis		DQ463169	DQ482962	DQ482942	DQ482977
Pectinopygus farallonii	DQ482949	DQ463171	DQ482964	DQ482931	DQ482979
Pectinopygus fregatiphagus		DQ489435	DQ489433		DQ489434
Pectinopygus minor		DQ463172	DQ482966		DQ482981
Pectinopygus occidentalis	DQ482945	DQ463165	DQ482958		
Pectinopygus punctatus	DQ482946	DQ463166	DQ482959	DQ482932	DQ482974
Pectinopygus setosus		DQ463167	DQ482960	DQ482941	DQ482975
Pectinopygus sulae	AY314870	DQ463177	DQ482971		AF320444
Pectinopygus tordoffi	DQ482951	DQ463174	DQ482968	DQ482937	
Pectinopygus varius			DQ482965		DQ482980
Anaticola crassicornis	DQ482943	DQ463163	DQ482955	DQ482929	AF320354

APPENDIX 4. Accession numbers for all bird sequences used in this study. New sequences indicated by (*).

Species	12S	COI	ATPase
Australian Darter *Anhinga novaehollandiae*	AY369047	AY369057	AY369071
Magnificent Frigatebird *Fregata magnificens*	AY369042	AY369052	AY369066
Great Frigatebird *Fregata minor*	AY369043	AY369053	AY369067
Northern Gannet *Morus bassanus*	EF101669*	EF101674*	EF101685*
Australasian Gannet *Morus serrator*	AY009321	AY369058	AY009345
American White Pelican *Pelecanus erythrorhynchos*	EF101668*	EF101673*	EF101684*
Brown Pelican *Pelecanus occidentalis*	AY369048	AY369059	AY369072
Masked Booby *Sula dactylatra*	AY941810	EF101677*	AY941806
Brown Booby *Sula leucogaster*	EF101671*	EF101676*	EF101687*
Blue-footed Booby *Sula nebouxii*	EF101670*	EF101675*	EF101686*
Red-footed Booby *Sula sula*	AY009322	EF101678*	AY009346
European Shag *Phalacrocorax aristotelis*	AY009329	EF101682*	AY009353
Double-crested Cormorant *Phalacrocorax auritus*	AY009328	EF101683*	AY009352
Great Cormorant *Phalacrocorax carbo*	AY009323	EF101679*	AY009347
Spotted Shag *Phalacrocorax punctatus*	AY009343	EF101681*	AY009367
Pygmy Cormorant *Phalacrocorax pygmaeus*	EF101672*		EF101688*
Little Black Cormorant *Phalacrocorax sulcirostris*	AY009332	EF101680*	AY009356
Pied Cormorant *Phalacrocorax varius*	AY009338	AY369060	AY009362