How stable is the "Polyphyly of Lice" hypothesis (Insecta: Psocodea)?: A comparison of phylogenetic signal in multiple genes

Kazunori Yoshizawa 1* & Kevin P. Johnson 2

1 Systematic Entomology, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan
2 Illinois Natural History Survey, Champaign, IL 61820, USA

*corresponding author
K. Yoshizawa
e-mail: psocid@res.agr.hokudai.ac.jp
tel: + 81-11-706-2424
fax: + 81-11-706-4939

Abstract
Recent molecular phylogenetic analyses of 18S rDNA have indicated that parasitic lice (order Phthiraptera) are possibly polyphyletic. These analyses recovered one of the parasitic louse suborders, Amblycera, as the sister group to the free-living booklouse family Liposcelididae. We further tested this hypothesis using DNA sequences from five genes: nuclear 18S rDNA, Histone 3, and wingless and mitochondrial 16S rDNA and COI. Combined analyses of these five genes provided reasonably strong support for the Amblycera + Liposcelididae clade, supporting the polyphyly of lice hypothesis. To explore the robustness of this result, we examined the phylogenetic signal contained in each gene independently (except for wingless, which could not be readily amplified in many target taxa). Analyses of each gene separately and in various combinations with other genes revealed that clear signal supporting Amblycera + Liposcelididae only existed in the 18S data, although no analysis supported monophyly of parasitic lice. Nevertheless, combined analyses of all genes provided stronger support for this relationship than that obtained from 18S data alone. The increase in support for this clade was mostly explained by the stabilization of other parts of the tree and potentially inappropriate substitution modeling. These findings demonstrate that the increased support values provided by combined data set does not always indicate corroboration of the hypothesis.

Introduction

The parasitic lice, generally known as the insect order Phthiraptera, are the only insects that spend their entire life cycle on vertebrate hosts. Based on morphology, a close evolutionary relationship between parasitic lice and booklice (Liposcelididae: a family of Psocoptera) has long been recognized (Lyal, 1985: Fig. 1). This relationship is also well supported by phylogenetic analyses of DNA sequences (Yoshizawa & Johnson, 2003; Johnson et al., 2004; Murrell & Barker, 2005). Together these data indicate that Psocoptera are paraphyletic: thus some authors recognize Psocodea (= Phthiraptera + Psocoptera) as the only valid order (e.g., Hennig, 1981; Kristensen, 1991; Yoshizawa & Johnson, 2006). In contrast, monophyly of Phthiraptera has not been questioned (e.g., Hennig, 1966; Jamieson et al., 1999), because of a number of shared morphological and physiological specializations to a parasitic lifestyle.

However, recent molecular analyses have called into question the monophyly of parasitic lice. Phylogenetic analyses of mitochondrial 12S and 16S rDNA (Yoshizawa & Johnson, 2003) and nuclear 18S rDNA (Johnson et al., 2004; Murrell & Barker, 2005) have not supported the monophyly of parasitic lice. Furthermore, analyses of 18S data provided strong support for two independent origins of parasitism by this group (Fig. 1). In particular, one of four suborders of the parasitic lice (Amblycera) was recovered as the sister taxon of the booklouse family Liposcelididae (Johnson et al., 2004; Murrell & Barker, 2005), in contrast to the traditional hypothesis based on external morphology (see Fig. 1). However, the ribosomal genes used in these previous studies are known to exhibit unusual evolutionary trends in parasitic lice and their relatives, such as accelerated substitution rates, modification of secondary structures, and increased GC content (Page et al., 2002; Johnson et al., 2003; Yoshizawa & Johnson, 2003; Johnson et al., 2004). These properties can potentially reduce the accuracy of phylogenetic estimation (e.g., Felsenstein, 1978; Kjer, 2004; Jermiin et al., 2004). In addition, parasitic lice share many unique adaptations to parasitism that strongly support their monophyly (e.g., Lyal, 1985). Therefore, some authors have questioned the reliability of these molecular phylogenies based on a limited number of potentially unusual genes (Grimaldi & Engel, 2006).

The polyphyly of parasitic lice hypothesis was also recently tested using male genitalic characters (Yoshizawa & Johnson, 2006). This character system is less likely to be effected by selection related to parasitic lifestyle, and thus should provide useful additional data to test the hypothesis of the polyphyly of parasitic lice. A phylogenetic analysis based on male genitalia also provided support for the polyphyly of parasitic lice (i.e. the presence of a novel articulation in the phallosome of Amblycera, Liposcelididae, and a barklouse family Pachytrocticae: Fig. 1). However, the novel articulation was also observed in a few species of the parasitic louse suborder Ischnocera so this morphological study also failed to provide unambiguous support for the polyphyly of parasitic lice hypothesis. Therefore, analyses of multiple additional gene regions are highly desirable.

In the present study, we tested the hypothesis of parasitic louse polyphyly using five genes selected from both the nuclear and mitochondrial genomes: nuclear 18S rDNA, Histone 3, and wingless and mitochondrial 16S rDNA and COI. These include both ribosomal and protein coding genes from each genome. We analyzed these data both separately and in combination to evaluate the contribution of each gene region to the phylogenetic results.

Materials and Methods

Terminology

The monophyly of Psocoptera with respect to Phthiraptera is doubtful (Fig. 1), and we

recognize Psocodea as the only valid order name uniting Psocoptera and Phthiraptera (Yoshizawa & Johnson, 2006). However, in this paper, we use the terms Psocoptera and Phthiraptera in the traditional sense (e.g., CSIRO, 1991) for convenience. In addition, paraphyly of the bark louse suborder Troctomorpha and infraorder Nanopsocetae also appears to be certain (Fig. 1), but we also use these names in the traditional sense (e.g., Lienhard & Smithers, 2002) because of convenience and lack of alternative terms (Bess et al., 2009).

Target Genes and Taxon Sampling

Four genes, mitochondrial protein coding (COI) and ribosomal (16S) genes and nuclear protein coding (Histone 3) and ribosomal (18S) genes were selected as main targets for this study. These genes were selected to maximize both target gene variety (protein coding and ribosomal genes from two different sources) and taxon coverage (i.e., ease of amplification), such that a relatively comprehensive dataset could be constructed without having to include taxa with missing data. As well as testing the monophyly of parasitic lice, our goal was to evaluate the contribution of each gene partition to the phylogenetic results so that preparation of a data set with no missing gene partitions was critical. However, to evaluate the effects of inclusion of different genes and taxa, we also prepared an additional data set containing taxa with missing data and the nuclear protein coding gene, wingless. The wingless gene was amplified from many parasitic lice taxa examined, but the gene was more difficult to amplify for Psocoptera and thus was missing for a number of taxa (Table 1).

Samples used in the present study are listed in Table 1. These samples cover all suborders and infraorders of the order Psocodea (=Psocoptera + Phthiraptera), except for the infraorder Prionoglaridetae of the bark louse suborder Trogiomorpha, a very rare cave-dwelling group (Yoshizawa et al., 2006). Most importantly, the present data set includes a sample of *Embidopsocus*, a representative of the subfamily Embidopsocinae of the family Liposcelididae (one of two subfamilies that sometimes has wings). The subfamily was not included in the previous analyses of lice and their relatives (Yoshizawa & Johnson, 2003; Johnson et al., 2004; Murrell & Barker, 2005). The tree was rooted with the suborder Trogiomorpha because this suborder has been supported as the sister taxon of the rest of Psocodea by previous studies (Johnson et al., 2004; Yoshizawa et al., 2006). Total DNA was extracted using Qiagen DNeasy Tissue Kit (Qiagen) according to the procedure described in Johnson et al. (2004).

Sequence Determination and Alignment

Primer sets L6625 or C1J-1718 + H7005 (COI: Hafner et al., 1994; Simon et al., 1994), 16Sar + 16Sbr (16S: Simon et al., 1994), HexAF + HexAR (H3: Colgan et al., 1998), NS1 + NS2 (18S: Johnson et al., 2004) and LepWg1 + LepWg2 (Wg: Brower & Egan, 1997) were used to amplify and sequence the regions of the genes used in this study. These primer sets provided 476-505 bp (16S), 330 bp (H3), 497-803 bp (18S), and 414-428 bp (Wg) PCR products. Sequences generated by the primer pair C1J-1718 and H7005 could not be resolved for ca. 100 bp on the 3' end of the sequence, so this region was trimmed from all COI sequences, and only the 260 overlapping base pairs between either primer combination were used. This was done to avoid potentially confounding effects of missing data (Wiens, 1998, 2003). When products from the first PCR were too faint to sequence, a second PCR was performed following Yoshizawa & Johnson (2008). PCR products were sequenced by the University of Illinois Core Sequencing Facilities or by using CEQ DNA Analysis System (Beckman Coulter) following manufacture's protocol.

Alignments of the protein coding H3 and COI genes were straightforward, with only one amino-acid deletion in COI in two *Liposcelis* species for which the position of the deletion was unambiguous based on translated amino-acids. The wingless gene included a longer insertion (a four amino-acid insertion) but alignment of the sequences was also straightforward based on translated amino-acids. The 18S and 16S genes were aligned according to a secondary structure model, as identified by Johnson et al. (2004) and Yoshizawa & Johnson (2003), respectively, and poorly aligned regions were excluded from the analyses (see Online Supplement). The final aligned data included 620 bp of 18S, 330 bp of H3, 335 of 16S, 260 bp of COI, and 428 bp of Wg.

Phylogenetic Analyses

We prepared three datasets: (1) an *expanded data set* which included 85 taxa and all five genes, but it also contained a number of missing data partitions. Presence of missing data can potentially destabilize the position of taxa with many missing data (Platnick et al., 1991; Novacek, 1992; Wiens, 1998). Furthermore, the main purpose of this study was to compare the phylogenetic signal in different genes from different sources. Therefore, we prepared (2) a *full data set* that only included 69 taxa and excluded the wingless gene sequences, and this data set had no missing data partitions. We also prepared (3) a smaller *Nanopsocetae-Phthiraptera data set* (NP-data: 18 taxa and four genes) to complete many different analyses within acceptable timeframe (see below). By using this smaller set of taxa, phylogenetic signal from different genes could be compared more comprehensively.

We analyzed each gene separately and in various combinations including 18S + H3, 18S + 16S, 18S + COI, H3 + 16S + COI (NP data set), 18S + H3 + 16S + COI (NP and full data sets), and all five genes together (expanded data set). For all data combinations, both maximum likelihood and Bayesian MCMC analyses were performed.

Using PAUP*4.0b10 (Swofford 2000), maximum likelihood analyses with TBR branch swapping with a NJ starting tree were conducted. The substitution models for ML analyses were estimated using AIC as implemented in Modeltest 3.7 (Posada & Crandall, 1998). We calculated bootstrap support for the branches estimated from the fully combined data set using PAUP* with 100 replicates of NNI branch swapping using a NJ tree as the starting tree, because TBR branch swapping was too time consuming for the analysis of 85 or 69 taxa. However, searches using NNI branch swapping are insufficient when the topology of ML tree is not similar to that of the NJ tree. When the full taxon data set was analyzed separately for each gene, the trees estimated by PAUP* with NNI branch swapping were substantially different from those estimated by PAUP* with TBR branch swapping (e.g., trees from 16S alone: tree not shown). Therefore, we constructed a data set containing only Nanopsocetae and Phthiraptera (Fig. 1), with Amphientometae as an outgroup (referred to as the NP data set hereafter). Monophyly of Nanopsocetae + Phthiraptera was supported by almost all analyses of each gene separately and in combination, except for the small extremely homoplasious COI data set alone (tree not shown). The NP data set was used to examine the contribution of each gene to relationships within this assemblage. For the analyses of the NP data set, 100 replicates of TBR bootstrapping were performed with PAUP* using a NJ tree as starting tree. However, elimination of many distantly related taxa in the NP data set may provide significantly different result from the data with full taxon set. If this is the case, then the difference could attribute to two factors, different branch swapping method and/or different taxon sampling. To determine the effects of the difference in tree search strategies, we also performed a TBR bootstrapping (100 replicates) for the full data

set, but 15 branches receiving > 95% NNI bootstrap and 100% Bayesian support were constrained to finish the bootstrap search within a reasonable time. The statistical support for some groupings of interests were also tested using the AU test (Shimodaira, 2002) with Consel (Shimodaira & Hasegawa, 2001) and Partitioned Likelihood Support scores (Lee & Hugall, 2003).

We used MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) for Bayesian MCMC analyses. We performed two runs each with four chains for 10,000,000 generations and trees were sampled every 1,000 generations. A previous study showed that, for a large data set including lice and their relatives, it took a very large number of generations to burn-in all parameters in Bayesian MCMC search (Johnson et al., 2004). Therefore, to estimate Bayesian posterior probabilities, we computed a 50% majority consensus tree of the 2,000 trees from the final 2,000,000 generations (i.e., first 98% of trees were excluded for burn-in). Because the NP-data set contained a much smaller number of species (19 species including outgroups), the number of MCMC generations was reduced to 1,000,000 for the data set, and the first 500 trees were excluded for burn-in to calculate the posterior probabilities. The substitution models for all Bayesian analyses were estimated using AIC as implemented in MrModeltest 2.3 (Nylander, 2004).

To explore the phylogenetic signal contained in the homoplasious genes and compare this with randomized data, we also constructed data combination of the real 18S data with randomized H3, 16S, or COI data constructed using Shuffle option in MacClade 4.08 (Maddison & Maddison, 2000). Because any phylogenetic signal in a data set is erased by randomization, hidden phylogenetic signal contained in more homoplasious data set can be identified by comparing the results from the 18S + real data and 18S + randomized data (Archie, 1989). To compare average values of bootstrap supports obtained from different data sets, 10 bootstrapping runs were performed for the NP data sets. For comparisons, T-test was adopted using JMP v. 8. (SAS, 2009).

All the aligned data matrices, outputs from model selections, and resulting tree files are available from http://kazu.psocodea.org/data and the journal's web site.

Results

Plots of uncorrected pairwise distances of each gene, except for 16S, against that of 18S (Fig. 2) indicate considerable multiple substitution for each gene relative to 18S. Plotted against 18S divergence, divergences of protein-coding genes leveled off at around 20-30%, whereas maximum divergence of 16S was 42.3%. However, when closely related species were compared (e.g., two barklouse species of the same genus, *Stenopsocus*: Table 1), pairwise distances was 0.19% for 18S, 8.94% for H3, 1.90% for 16S and 14.6% for COI. This showed that the substitution rate is fastest in COI, followed by H3, 16S, and slowest in 18S. Although the substitution rate appears slower, more overall divergence was observed in 16S than other protein coding genes, probably because the protein coding genes were under strong constraint from amino-acid level selection, limiting the maximum divergences. The wingless gene data set contained a number of missing taxa and was difficult to compare directly to other genes, but the two protein coding nuclear genes (wingless and H3) appeared to be similar in the rate of accumulation of substitution (Fig. 2). These trends agree with those identified by Yoshizawa & Johnson (2008).

The topology for Psocodea from the ML analysis of the (1) *expanded data set* is well-resolved and generally well-supported (Fig. 3). The tree also agreed very well with the previous morphology- and molecular-based phylogenetic hypotheses. Bayesian analyses resulted in a

nearly identical topology except for rearrangements of some weakly supported branches (available online). However, unlike some previous studies, the monophyly of the louse suborder Ischnocera was not recovered, because the suborders Rhynchophthirina + Anoplura were imbedded within Ischnocera. However, the AU test showed that the likelihood score of the ML tree with monophyly of Ischnocera constrained was not significantly worse from that of the unconstrained ML tree (P=0.141). The suborders Ischnocera, Rhynchophthirina and Anoplura together formed a monophyletic group receiving high support (79% bootstrap, 100% posterior probability), and a sister relationship between Rhynchophthirina and Anoplura was also reasonably supported (54% bootstrap and 100% posterior probability). The suborder Anoplura formed monophyletic group with reasonable support values (52% bootstrap and 100% posterior probability). The very small suborder Rhynchopthirina was represented by only one species so that monophyly of the suborder could not be tested by the present data set. Like previous studies based on only 18S monophyly of Phthiraptera was not recovered, and the booklouse family Liposcelididae was placed as the sister taxon of the louse suborder Amblycera (70% bootstrap, 100% posterior probability). However, even though the polyphyly of Phthiraptera was supported by bootstrapping and posterior probability, the likelihood score of the ML tree with constrained monophyly of Phthiraptera was not significantly worse than that of the unconstrained ML tree (P=0.156 under the AU test). Partitioned Likelihood Support (PLS) values were not calculated for the expanded data set because of many missing data present in the data set.

The topology from the ML and Bayesian analysis of the (2) *full data set* is generally in accordance to that obtained from the expanded data set (Fig. 4). All the major relationships recovered by the expanded data set were also reasonably supported by the full data set. In particular, monophyly of Amblycera + Liposcelididae (i.e. polyphyletic parasitic lice, Phthiraptera) was also recovered and received strong statistical support (75% bootstrap by NNI and 100% posterior probability). Using constrained TBR bootstrapping, the support value for this clade increased to 90%. PLS values suggested that all gene partitions contained positive signal for this clade (18S: 4.858; H3: 1.397; 16S: 0.988; COI: 1.999). Results of the AU test indicated that the likelihood scores of the constrained trees were not quite significantly worse (P=0.071 for constrained monophyly of Phthiraptera and P=0.073 for constrained monophyly of Ischnocera) than those of the unconstrained ML tree.

The trees resulting from the (3) *NP-data set* with four genes combined (18S, H3, 16S, COI) were concordant with those estimated from full taxon set (Fig. 5). Bootstrap support obtained from the combined NP data was analogous to that obtained from the full data set. Therefore, the effect of exclusion of distant outgroups is not evident in the NP-data set. Separate analyses of each gene partition revealed that the sister relationship between Lipocelididae and Amblycera was only recovered by the 18S data (Fig. 5), and the support values from 18S data alone were not as high (Table 2) as the combined data. Again, the AU test showed that the result is not quite significantly better than monophyletic Phthiraptera (P=0.094). None of the other gene partitions recovered this relationship, but PLS scores from the combined data showed positive support from all examined genes for this relationship except for H3 (18S: 8.384; H3: -3.374; 16S: 0.852; COI: 1.278). The 16S data failed to recover monophyly of Lipocelididae, and the COI data failed to recover monophyly of both Lipocelididae and Amblycera. Monophyly of Ischnocera was not recovered by any of the separate or combined analyses of the NP-data (Fig. 5).

In examining various gene combinations of the NP-data, a sister relationship between Amblycera and Liposcelididae was recovered only when the data combination included 18S.

Support for this relationship usually increased in these partially combined analyses over those obtained from 18S alone (Table 2). When 18S data were excluded from the analyses, this relationship was never recovered (Fig. 5). However, the AU test of the partially combined H3 + 16S + COI data showed that the likelihood scores of the ML trees with either the constrained monophyly of parasitic lice (P=0.195) or polyphyletic parasitic lice (i.e. Amblycera + Liposcelididae, P=0.407) were not significantly worse from that of the unconstrained ML tree.

Support values for Amblycera + Liposcelididae were increased even if the extremely homoplasious COI was combined with 18S (Table 2). Therefore, to test contribution of the non-18S data for this relationship, we performed the following two analyses: (1) 18S alone or various combinations of H3, 16S, and COI with branches receiving 100% Bayesian support value from the other data set constrained to be monophyletic (i.e., the 100% Bayesian tree from 18S was used to constrain H3, 16S, COI and their combination; the 100% tree from H3+16S+COI was used to constrain 18S); and (2) H3, 16S and COI data were randomized using the Shuffle option as implemented in MacClade 4 (Madditon & Maddison, 2000) and each randomized data was combined with 18S and analyzed with likelihood. The Shuffle option removes completely the phylogenetic signal from the data matrix but keeps some parameters constant, such as the number of sites, codon base composition, and number of invariant sites.

A constrained analysis of 18S alone provided stronger support for Amblycera + Liposcelididae than did unconstrained analyses (Table 2). In contrast, when the other genes were analyzed with constraints, no data combination recovered the Amblycera + Liposcelididae clade. By combining randomized data from the other genes with the true 18S data, support for Amblycera + Liposcelididae increased in all cases. Support for Amblycera + Liposcelididae was originally 68% in unconstrained analysis of 18S and in combination with randomized data from other genes increased to 87% (+ randomized H3), 94% (+ randomized 16S), and 91% (+randomized COI).

To clarify the degree of differences in support values and evaluate which factor likely caused the increased support value by adding phylogenetically uninformative data, we performed 10 bootstrapping (TBR) analyses each for 18S and 18S + randomized COI of the NP data sets as follows (Fig. 6): (1) likelihood analyses adopting the best fit models estimated from 18S and 18S+COI, respectively; (2) likelihood analyses adopting JC model; and (3) parsimony analyses (TBR branch swapping with 100 random starting trees). When the best-fit models were adopted, support values for Amblycera + Liposcelididae obtained from the 18S + randomized COI data set were significantly higher than those obtained from the 18S alone ($P<0.0001$ by t-test). The PLS values exhibited positive signal even from the randomized COI (4.39) as well as 18S (5.88). In contrast, application of JC model and parsimony criterion reduced bootstrap support values of Amblycera + Liposcelididae significantly by inclusion of randomized data (P=0.0315 by JC and <0.001 by parsimony).

We also tested the effects of the addition of highly homoplasious or random data by examining the bootstrap support for some clades well supported by 18S alone, i.e., *Menacanthus + Menopon* and *Physconelloides + Campanulotes*. These clades were strongly supported by 18S data (99-100% bootstrap), and weak signal for these grouping were also detected from separately analyzed COI (55-79% bootstrap: Fig. 5). By combining COI data with more consistent 18S data, bootstrap support for *Menacanthus + Menopon* and *Physconelloides + Campanulotes* decreased (Table 2). Addition of randomized data also decreased likelihood bootstrap value for these groups significantly (figure not shown: average bootstrap value decrease from 100% to 91.2% for *Menacanthus + Menopon* [$P<0.0001$] and from 99.7% to 97.9% for *Physconelloides +*

Campanulotes [P=0.0031]).

Discussion

Combined analyses of five or four gene regions provided a generally well-resolved and supported tree for barklice and parasitic lice (Psocodea). In general, support for various clades of interest was improved over previous studies on a smaller number of genes (Figs 3,4) (Yoshizawa & Johnson, 2003; Johnson et al., 2004; Murrell & Barker, 2005). In particular, monophyly of Troctomorpha + Phthiraptera received stronger support (56-60% bootstrap, 94-99% posterior probability) compared to less than 50% in previous studies. Monophyly of Nanopsocetae + Phthiraptera was very strongly supported (99-100% bootstrap & posterior probability) in comparison to generally less than 50% in most previous analyses. These results are concordant with morphologically-based hypotheses (Lyal, 1985; Yoshizawa & Johnson, 2006) and suggest that clearer phylogenetic signal is obtained by combination of multiple genes. Also of note, the present set of taxa contains *Embidopsocus* (Liposcelididae: Embidopsocinae), a key subfamily not included by Yoshizawa & Johnson (2003), Johnson et al. (2004), or Murrell & Barker (2005). *Embidopsocus*, an occasionally winged free-living barklouse, divided the long ancestral branch of *Liposcelis* (Figs 3,4), which would likely help alleviate the potentially negative effects of long branch attraction. Not all branches in the tree improved and, like previous studies, the placements of *Tapinella* and *Badonnelia* within the Nanopsocetae + Phthiraptera clade were unstable (Figs 3,4).

A key remaining issue is the polyphyly of parasitic lice, as was first suggested by the extensive analyses of 18S data by Johnson et al. (2004). However, 18S exhibits an accelerated substitution rate in lice and their relatives, and it has been suggested that this result might be an artifact caused by long branches relative to other Psocodea (Grimaldi & Engel, 2006). The additional genes included in the current analyses may avoid the accelerated substitution rates and secondary structure anomalies of 18S. Combined analyses of five or four gene regions still support the polyphyly of parasitic lice (Figs 3,4), even though the 18S sequences in the present data set include only about 1/3 of entire sequence, unlike previous studies which used almost the entire gene (Johnson et al., 2004; Murrell & Barker, 2005). As a result, the 18S data comprised only 32-38% of the sites in the present combined data set. PLS values also suggested that H3, 16S, and even COI also contained positive signal for this relationship. Thus, the analyses of multiple genes seemingly corroborate the polyphyly of lice hypothesis.

However, separate analyses of each gene region were not as clear. A sister relationship between Amblycera and Liposcelididae (i.e. parasitic louse polyphyly) was never supported by separate or partially combined analyses of any data set that did not contain the 18S data (Fig. 5). Furthermore, a sister relationship between Amblycera + Liposcelididae was not supported even by a constrained analysis (i.e. constraining some branches receiving 100% posterior probability using 18S alone) of any combination of the H3, 16S, and COI data. These analyses indicate that, unlike 18S, the other genes contain little phylogenetic signal supporting Amblycera + Liposcelididae. It should also be noted, however, that no analysis supported monophyly of parasitic lice (Phthiraptera).

Nevertheless, when data from other genes were combined with 18S, support values for Amblycera + Liposcelididae (i.e. parasitic louse polyphyly) increased in most cases (Table 2). By constraining some branches that were strongly supported by analyses of H3 + 16S + COI (i.e., Bayesian posterior probability 100%), support for Amblycera + Liposcelididae from 18S of the NP data increased from 68% to 76% (Table 2). These analyses indicate that the 18S data

contain even more signal supporting Amblycera + Liposcelididae than revealed by unconstrained ML analysis of 18S data alone, and this increased value can be interpreted as signal hidden by instability of other parts of the tree in the separate analysis of 18S. However, support values for Amblycera + Liposcelididae improved dramatically even when 18S data was combined with the very homoplasious COI gene. This result is somewhat peculiar because separate analyses suggested that the COI data are almost completely useless in resolving deep relationships of lice and their relatives (Fig. 5) and did not contribute for stabilization of the analyses. PLS values are also unexpected because the results suggested that the most homoplasious COI contained more signal supporting the relationship (1.999) than that in H3 (1.397) and 16S (0.988). Therefore, to evaluate whether the improved support value was from true phylogenetic signal in COI data or not, we constructed a data combination of 18S data with randomized COI data constructed using Shuffle option in MacClade (Maddison & Maddison, 2000). Because any phylogenetic signal in the COI gene partition is erased by this randomization, hidden phylogenetic signal contained in the real COI data set can be identified by comparing the results from the 18S + randomized COI and 18S + real COI data (Archie, 1989). Surprisingly, the ML bootstrap support value for the Amblycera + Liposcelididae clade was still improved (over 18S alone) by addition of the randomized COI data set: the value is close to that obtained from 18S + real COI data set (Table 2). In addition, the randomized COI data also resulted in a positive PLS value. This result indicates that the higher support value obtained from analysis of the 18S + real COI data set for Amblycera + Liposcelididae was not a result of hidden signal contained in the COI gene. A similar effect was also identified when randomized H3 and 16S data were each combined with 18S.

Because the randomized data contain no phylogenetic signal, normally such data should not strengthen the support for branches. In fact, addition of randomized data decrease likelihood bootstrap value significantly for some clades strongly supported by 18S data. For example, average bootstrap values of *Menacanthus* + *Menopon* decrease from 100% (18S alone) to 91.2% (18S + randomized COI) ($P<0.0001$) and that of *Physconelloides* + *Campanulotes* reduced from 99.7% to 97.9% ($P=0.0031$). In addition, reduction of support values was also detected for these clades by adding highly homoplasious COI data to more consistent 18S data (Table 2). Interestingly, the reduction of these support values was detected even though these groups were also supported by independent COI data (Fig. 5). A similar effect was also demonstrated in previous studies that used parsimony bootstrapping (Johnson & Clayton, 2000; Johnson & Whiting, 2002). These results suggest that the increased support value by the addition of random or extremely homoplasious data is a phenomenon specific to the Amblycera + Liposcelididae clade.

One possible cause of this unusual result may be the use of differing evolutionary substitution models in different analyses. When multiple data sets are combined into one data matrix and analyzed under the likelihood criterion, most tree-searching software does not allow application of multiple substitution models for different gene partitions (e.g. PAUP*). In this case, the substitution model applied to the combined analysis will usually differ from that adopted for the analyses of separate gene regions. When substitution models of two gene partitions are extremely different, application of a single substitution model to the combined data set may cause unexpected effects. To test this possibility, we also analyzed the 18S data alone, but the best-fit models estimated from the 18S + H3, 18S + 16S, and 18S + COI data sets were adopted. Even in this situation, a substantial improvement of support values for the Amblycera + Liposcelididae clade (over 18S alone) still occurred (68% [18S model] to 68% [18S + H3

model], 83% [18S + 16S model] and 76% [18S + COI model]). In contrast, when 18S alone and 18S + randomized COI data were analyzed with a JC-model or parsimony criterion, the addition of random data provided lower support values for the Amblycera + Liposcelididae clade (Fig. 6). Under the JC-model and parsimony criterion, substitution patterns and rates are set to be equal among sites and gene partitions (Swofford et al., 1996). Therefore, under these parameters, the power to extract phylogenetic signal from each partition of combined genes is likely not significantly altered even if genes with significantly different substitution rate/pattern are combined into a single matrix. In contrast, under more complex substitution models, weighting for a site can be changed by application of a different substitution rate matrix, gamma shape parameter, or proportion of invariant sites, and increased support values could potentially result from these factors even if the amount of signal is unchanged. When no phylogenetic signal exists in a data set, likelihood analysis with a JC-model generally results in an unresolved polytomy (Swofford et al., 2001). Many empirical studies have shown that the addition of noisy data reduces the support values of parsimony bootstrapping (e.g., Johnson & Clayton 2000; Johnson & Whiting 2002; Damgaard & Congato, 2003), although a simulation study showed that addition of noisy data not always reduces the accuracy of parsimony estimation (Wenzel & Siddal, 1999). These behaviors of likelihood and parsimony analyses for random or homoplasious data probably explain why bootstrap support values for the Amblycera + Liposcelididae were only slightly versus greatly decreased by the addition of randomized data under the JC-modeling and parsimony criterion, respectively.

Another possible problem in the substitution modeling for the likelihood analysis is that nearly all phylogenetic models (including those implemented in PAUP* and MrBayes) adopt a single substitution model for all branches in the tree. If the substitution properties along certain branches is significantly different from those of the other tree parts, then the global best fit model might fit worse for the branches with significantly different substitution patterns and rates (Galtier, 2001). Judging from the trees estimated from the 18S data alone (Fig. 5), Liposcelididae represents the longest branches in the entire tree. In particular, the branch leading to *Embidopsocus* is extremely long even among the long liposcelidid branches. In the genes analyzed in this study, the substitution rates of the 18S gene are slowest, and the substitution rates of the other genes are much faster than 18S (Fig. 2). Therefore, the substitution models estimated from the combination of 18S and more rapidly evolving genes might fit better for some local branches such as Liposcelididae and Amblycera than the global best-fit model. Although the exact cause of this phenomenon is not completely understood and is beyond the scope of this study, the present findings show that, even if the addition of more gene regions provides more support for a certain branch, this increase might not be caused by phylogenetic signal in the additional gene(s).

In summary, analyses of multiple genes from different sources to test the polyphyly of parasitic lice hypothesis recovers a sister relationship between Amblycera and Liposcelididae (i.e. parasitic louse polyphyly). However, results from AU test showed that the tree supporting monophyletic parasitic lice (Phthiraptera) is not significantly worse (though marginally so) from the ML tree supporting Amblycera + Liposcelididae. In addition, most of the phylogenetic signal supporting Amblycera + Liposcelididae is only in the 18S gene partition, and other genes contain little detectable signal supporting this relationship. We also demonstrate that the addition of random data or adopting a different substitution model could even increase the support value for this relationship. Although we did not test how general this phenomenon is, this example shows that caution should be used in interpreting the results of strictly combined multigene analysis. As

a result, the present data set failed to corroborate the polyphyly of lice hypothesis from data other than that originally used to propose the hypothesis (i.e., 18S: Johnson et al., 2004; Murrell & Barker, 2005). However it should be noted that **no** data partition supported the monophyly of parasitic lice (Phthiraptera) and also almost no signal contradictory to this hypothesis could be detected. Our current study also added a representative of the subfamily Embidopsocinae (of Liposcelidae), and this did not alter the Amblycera + Liposcelididae relationship. Thus, this result is robust to taxon inclusion, because the long branch of Liposcelididae was broken by the inclusion of *Embidopsocus* (Figs 3, 4). Apart from the parasitic louse polyphyly question, the present data set successfully clarified some uncertainty that remained in previous studies, i.e., monophyly of Nanopsocetae + Phthiraptera and monophyly of Troctomorpha + Phthiraptera.

Acknowledgments

We thank C. Lienhard, E. L. Mockford, Y. Kamimura, L. Durden, and J. Klicka for specimens and Vince Smith, Alfried Vogler, and two anonymous reviewers for their constructive comments to the ms. KY thanks E. Hasegawa and M. Ôhara for allowing the use of their molecular facilities. This study was supported by JSPS Grant Nos. 18770058 and 21770083 to KY and NSF DEB-0612938 to KPJ. KY's field trips to Malaysia and USA were supported by JSPS Grant No. 17255001 (project leader: Osamu Yata) and MEXT Japan, respectively.

References

Archie, J.W., 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38, 239-252.

Bess, E., Smith, V.S., Johnson, K.P., 2009. Troctomorpha. Parasitic lice and book lice. Version 25. Mar. 2009 (under construction). In *The Tree of Life Web Project.* URL at: http://www.tolweb.org/Troctomorpha/30257.

Brower, A.V.Z., Egan, M.G., 1997. Cladistic analysis of *Heliconius* butterflies and relatives (Nymphalidae: Heliconiiti): a revised phylogenetic position for *Eueides* based on sequences from mtDNA and a nuclear gene. Proc. R. Soc. Lond. (B) 264, 969-977.

Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis, G., Gray, M.R., 1998. Histone H3 and U2 sn RNA DNA sequences and arthropod molecular evolution. Aust. J. Zool. 46, 419-437.

Cruickshank, R. H., Johnson, K. P., Smith, V. S, Adams, R, J., Clayton, D. H. & Page, R. D. M. 2001. Phylogenetic analysis of partial sequences of elongation factor 1 alpha identifies major groups of lice (Insecta: Phthiraptera). Molecular Phylogenetics and Evolution 19, 202-215.

CSIRO, 1991. The Insects of Australia, a Textbook for Students and Research Workers. Cornell Univ. Press., NY.

Damgaard, J., Cognato, A.I., 2003. Sources of character conflict in a clade of water striders (Heteroptera: Gerridae). Cladistics 19, 512-526.

Felsenstein, J., 1978. Cases in which parsimoy or compatibility methods will be positively misleading. Syst. Zool. 27, 401-410.

Galtier, N., 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol. Biol. Evol. 18, 866-873.

Grimaldi, D., Engel, M.S., 2006. Fossil Liposcelididae and the lice ages (Insecta: Psocodea). Proc. R. Soc. B. 273, 625-633.
Hafner, M.S., Sudman, P.D., Villablanca, F.X., Spradling, T.A., Demastes, J.W., Nadler, S.A., 1994. Disparate rates of molecular evolution in cospeciating host and parasites. Science 265, 1087-1090.
Hennig, W., 1966. Phylogenetic Systematics, University of Illinois Pr., Urbana, IL.
Hennig, W., 1981. Insect Phylogeny. Wiley & Suns, Chichester, UK.
Jamieson, B.G.M., Dallai, R., Afzelius, B.A., 1999. Insects: Their Spermatozoa and Phylogeny. Sciences Publishers, Inc., NH.
Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, AnW.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53, 638-643.
Johnson, K. P. & Clayton, D. H. 2000. Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). Molecular Phylogenetics and Evolution 14, 141-151.
Johnson, K.P., Cruickshank, R.H., Adams, R.J., Smith, V.S., Page, R.D.M., Clayton, D.H., 2003. Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). Mol. Phylogenet. Evol., 26, 231-242.
Johnson, K.P., Whiting, M.F., 2002. Multiple genes and the monophyly of Ischnocera (Insecta: Phthiraptera). Mol. Phylogenet. Evol. 22, 101-110.
Johnson, K.P., Yoshizawa, K. & Smith, V.S., 2004. Multiple origins of parasitism in lice. Proc. R. Soc. Lond. (B) 271, 1771-1776.
Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506-514.
Lee, M.S., Hugall, A.F., 2003. Partitioned likelihood support and the evaluation of data set conflict. Syst. Biol. 52, 15-22.
Lienhard, C., Smithers, C.N., 2002. Psocoptera (Insecta): World Catalogue and Bibliography. Instrumenta Biodiversitatis V, Muséum d'Histoire naturelle, Geneva.
Lyal, C.H.C., 1985. Phylogeny and classification of the Psocodea, with particular reference to the lice (Psocodea: Phthiraptera. Syst. Entomol. 10, 145-165.
Maddison, D.R., Maddison, W.P., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Assoc., Sunderland, MA.
Murrell, A., Barker, S.C., 2005. Multiple origins of parasitism in lice: phylogenetic analysis of SSU rDNA indicates that Phthiraptera and Psocoptera are not monophyletic. Parasitol. Res. 97, 274-280.
Novacek, M. J., 1992. Fossils as critica data for phylogeny. In: Wheeler, Q.D., Novacek, M.J. (Eds). Extinction and Phylogeny. Columbia Univ. Press, NY, 46-88.
Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
Page, R.D.M., Cruickshank, R., Johnson, K.P., 2002. Louse (Insecta: Phthiraptera) mitochondrial 12S rRNA secondary structure is highly variable. Insects Mol. Biol. 11, 361-369.
Platnick, N.I., Griswold, G.E., Coddington, J.A., 1991. On missing entries in cladistic analysis. Cladistics 7, 337-343.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817-818.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under

mixed models. Bioinformatics 19, 1572-1574.
SAS Institute Inc., 2009. JMP, version 8. SAS Institute Inc., Cray, NC.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492-508.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinfo. 17, 1246-1247.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a comparison of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 81-144.
Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4, Sinauer Assoc., Sunderland, MA.
Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inverence. In *Molecular Systematics* (Hillis, D.M., Moritz, C., Mable, B.K. eds), pp. 407-514. Sinauer Assoc., Sunderland, MA.
Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choise between parsimony and likelihood methods. Syst. Biol. 50, 525-539.
Wenzel, J.W., Siddall, M.E., 1999. Noise. Cladistics 15, 51-64.
Wiens, J.J., 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? Syst. Biol. 47, 625-640.
Wiens, J.J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol. 52, 528-538.
Yoshizawa, K. & Johnson, K.P., 2003. Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. Mol. Phylogenet. Evol. 29, 102-114.
Yoshizawa, K. & Johnson, K.P., 2006. Morphology of male genitalia in lice and their relatives and phylogenetic implications. Syst. Entomol. 31, 350-361.
Yoshizawa, K. & Johnson, K. P. 2008. Molecular systematics of the barklouse family Psocidae (Insecta: Psocodea: “Psocoptera”) and implications for morphological and behavioral evolution. Molecular Phylogenetics and Evolution 46, 547-559.

Table and Figure Captions

Table 1. Taxa included in the study.

Table 2. Statistical support values (ML bootstrap/posterior probability or ML bootstrap) for some clades of interest obtained by the various gene combinations or substitution model application of the NP-data set.

Fig. 1. Two alternative hypotheses on the phylogeny of parasitic lice and their relatives. Corresponding common names are also shown.

Fig. 2. Plot of uncorrected pairwise distances of genes against 18S. Taxa with missing gene sequences are not included.

Fig. 3. The maximum likelihood tree estimated from the *expanded data set* (all genes for 85 taxa, with many missing data partitions). Branch lengths are proportional to substitutions per site estimated by likelihood. Numbers associated branches are likelihood bootstrap support (NNI search)/Bayesian posterior probability values. Support value lower than 50% bootstrap and 90% posterior probability are not shown. The Nanopsocetae + Phthiraptera clade is highlighted with pale gray and Amblycera + Liposcelididae is highlighted with dark gray. Suborder names of the Phthiraptera are highlighted with thick font.

Fig. 4. The maximum likelihood tree estimated from the *full data set* (four genes for 69 taxa: no missing data). Branch lengths are proportional to substitutions per site estimated by likelihood. Numbers associated branches are likelihood bootstrap support (NNI search)/Bayesian posterior probability values (top) and likelihood bootstrap support obtained by constrained TBR search (bottom). Support value lower than 50% bootstrap and 90% posterior probability are not shown. The Nanopsocetae + Phthiraptera clade is highlighted with pale gray and Amblycera + Liposcelididae is highlighted with dark gray. Suborder names of the Phthiraptera are highlighted with thick font.

Fig. 5. The maximum likelihood trees estimated from combined and separately analyzed *Nanopsocetae-Phthiraptera data sets*. Branch lengths are proportional to substitutions per site estimated by likelihood. Numbers associated branches are likelihood bootstrap support (above) and Bayesian posterior probability (below) values. Support value lower than 50% bootstrap and 90% posterior probability are not shown. Amblycera and Lipocelididae are highlighted with gray.

Fig. 6. Variation in bootstrap support values obtained from 10 independent bootstrap analyses of 18S data alone and combined 18S + randomized COI. Plots indicate bootstrap value obtained from each analysis, and diamonds indicate mean value (middle line), 95% confidence intervals (upper and lower ends), and overlap lines (near upper and lower ends).

Figure1

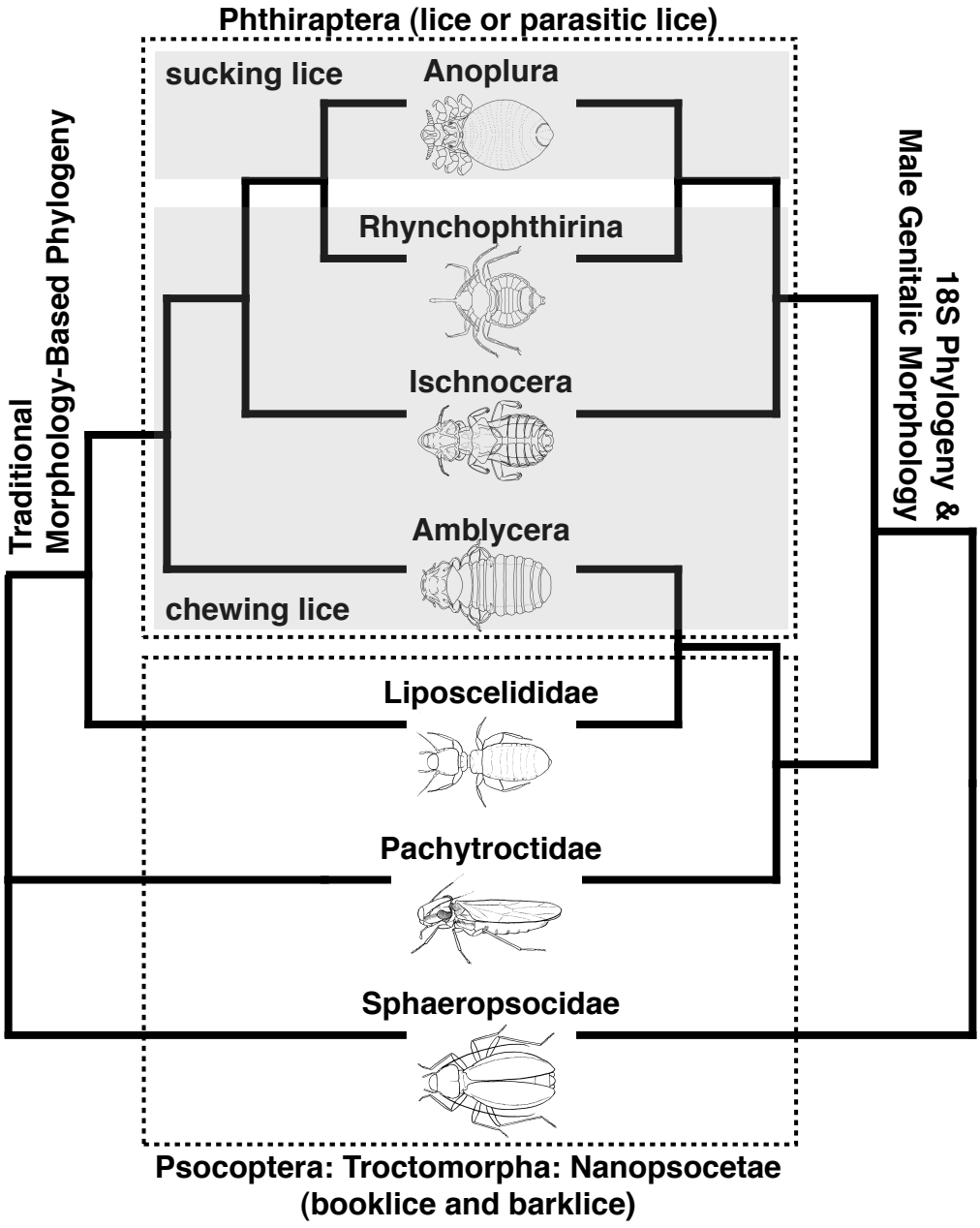

Phthiraptera (lice or parasitic lice)
sucking lice
Anoplura
Rhynchophthirina
Ischnocera
Amblycera
chewing lice
Traditional Morphology-Based Phylogeny
18S Phylogeny & Male Genitalic Morphology
Liposcelididae
Pachytroctidae
Sphaeropsocidae
Psocoptera: Troctomorpha: Nanopsocetae
(booklice and barklice)

Figure2

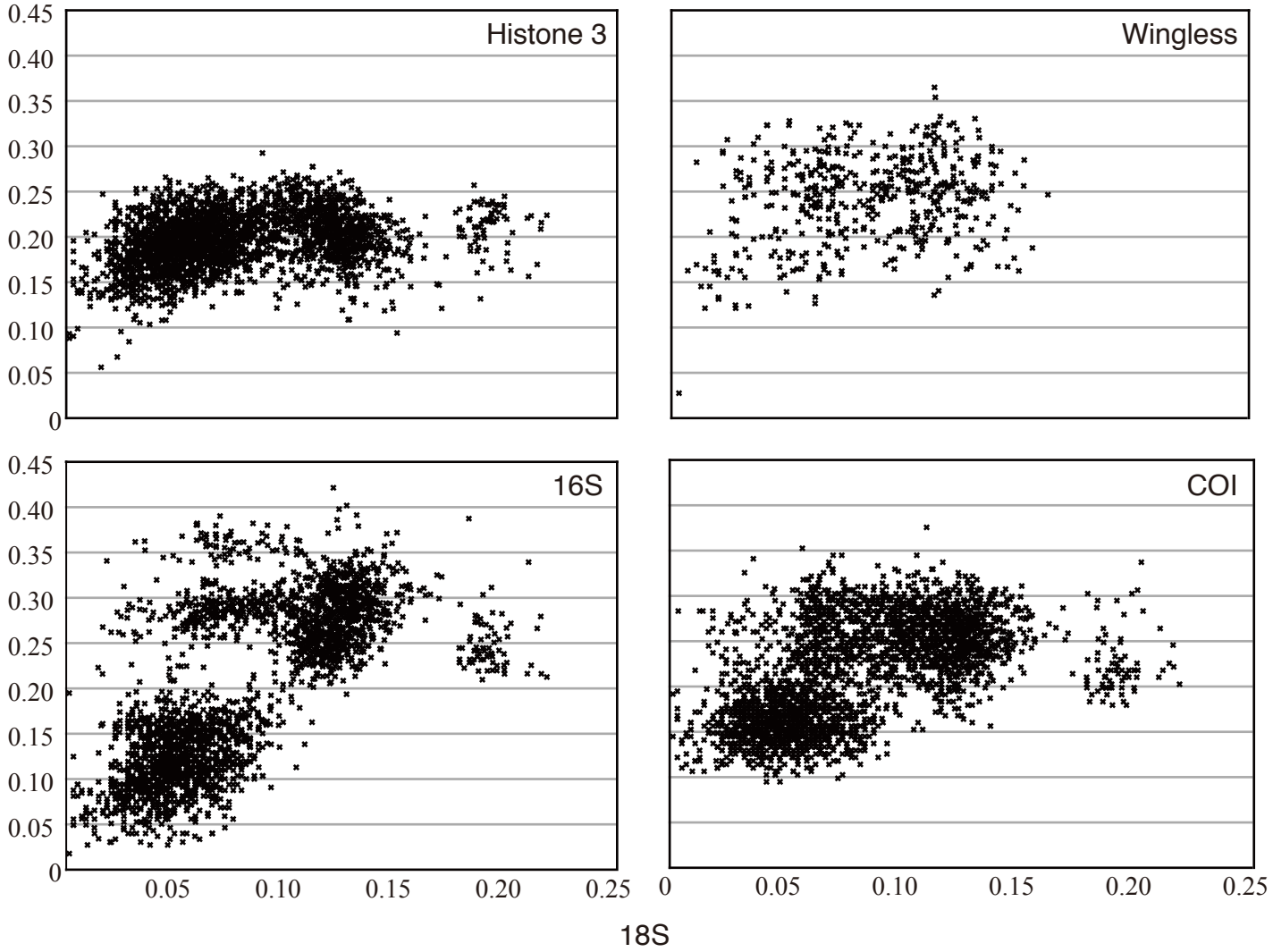

Histone 3
Wingless
16S
COI
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
0.05
0.10
0.15
0.20
0.25
18S

Figure3

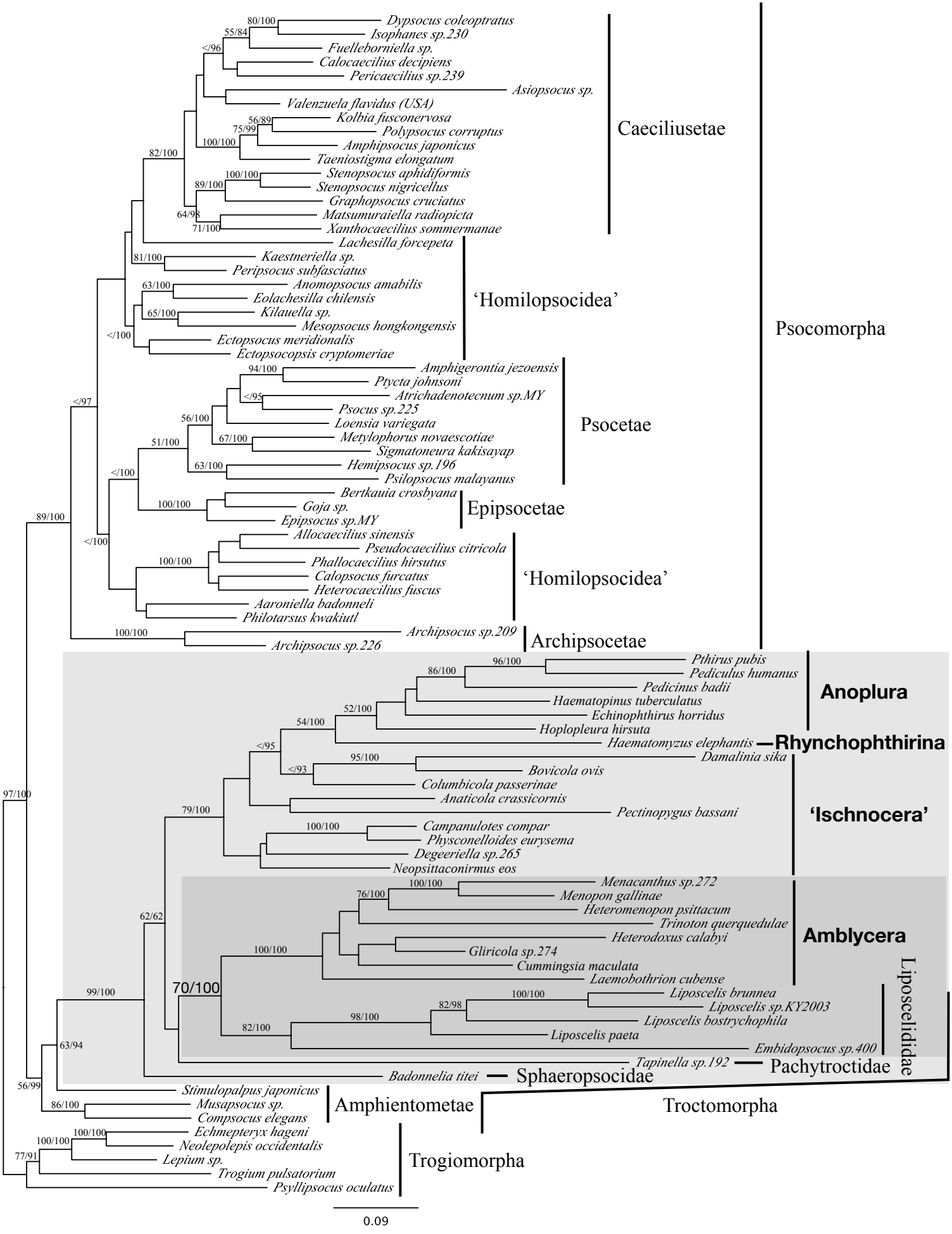

Dypsocus coleoptratus
Isophanes sp.230
Fuelleborniella sp.
Calocaecilius decipiens
Pericaecilius sp.239
Asiopsocus sp.
Valenzuela flavidus (USA)
Kolbia fusconervosa
Polypsocus corruptus
Amphipsocus japonicus
Taeniostigma elongatum
Stenopsocus aphidiformis
Stenopsocus nigricellus
Graphopsocus cruciatus
Matsumuraiella radiopicta
Xanthocaecilius sommermanae
Caeciliusetae
Lachesilla forcepeta
Kaestneriella sp.
Peripsocus subfasciatus
Anomopsocus amabilis
Eolachesilla chilensis
Kilauella sp.
Mesopsocus hongkongensis
Ectopsocus meridionalis
Ectopsocopsis cryptomeriae
'Homilopsocidea'
Amphigerontia jezoensis
Ptycta johnsoni
Atrichadenotecnum sp.MY
Psocus sp.225
Loensia variegata
Metylophorus novaescotiae
Sigmatoneura kakisayap
Hemipsocus sp.196
Psilopsocus malayanus
Psocetae
Bertkauia crosbyana
Goja sp.
Epipsocus sp.MY
Epipsocetae
Allocaecilius sinensis
Pseudocaecilius citricola
Phallocaecilius hirsutus
Calopsocus furcatus
Heterocaecilius fuscus
Aaroniella badonneli
Philotarsus kwakiutl
'Homilopsocidea'
Archipsocus sp.209
Archipsocus sp.226
Archipsocetae
Psocomorpha
Pthirus pubis
Pediculus humanus
Pedicinus badii
Haematopinus tuberculatus
Echinophthirus horridus
Hoplopleura hirsuta
Haematomyzus elephantis
Anoplura
Rhynchophthirina
Damalinia sika
Bovicola ovis
Columbicola passerinae
Anaticola crassicornis
Pectinopygus bassani
Campanulotes compar
Physconelloides eurysema
Degeeriella sp.265
Neopsittaconirmus eos
'Ischnocera'
Menacanthus sp.272
Menopon gallinae
Heteromenopon psittacum
Trinoton querquedulae
Heterodoxus calabyi
Gliricola sp.274
Cummingsia maculata
Laemobothrion cubense
Amblycera
Liposcelis brunnea
Liposcelis sp.KY2003
Liposcelis bostrychophila
Liposcelis paeta
Embidopsocus sp.400
Liposcelididae
Tapinella sp.192
Pachytroctidae
Badonnelia titei
Sphaeropsocidae
Troctomorpha
Stimulopalpus japonicus
Musapsocus sp.
Compsocus elegans
Amphientometae
Echmepteryx hageni
Neolepolepis occidentalis
Lepium sp.
Trogium pulsatorium
Psyllipsocus oculatus
Trogiomorpha
0.09

Figure4

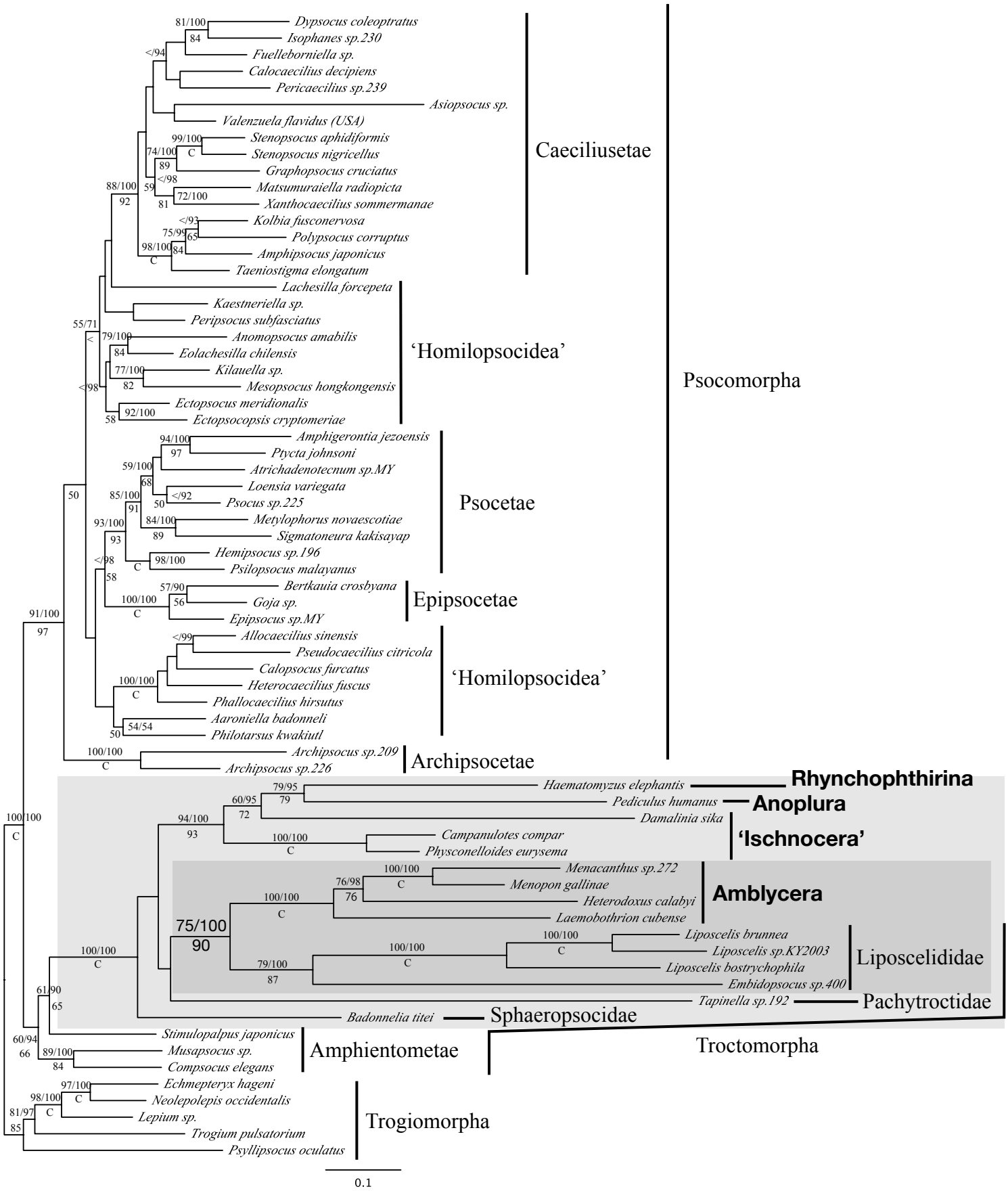

Dypsocus coleoptratus
Isophanes sp.230
Fuelleborniella sp.
Calocaecilius decipiens
Pericaecilius sp.239
Asiopsocus sp.
Valenzuela flavidus (USA)
Stenopsocus aphidiformis
Stenopsocus nigricellus
Graphopsocus cruciatus
Matsumuraiella radiopicta
Xanthocaecilius sommermanae
Kolbia fusconervosa
Polypsocus corruptus
Amphipsocus japonicus
Taeniostigma elongatum
Caeciliusetae
Lachesilla forcepeta
Kaestneriella sp.
Peripsocus subfasciatus
Anomopsocus amabilis
Eolachesilla chilensis
Kilauella sp.
Mesopsocus hongkongensis
Ectopsocus meridionalis
Ectopsocopsis cryptomeriae
'Homilopsocidea'
Amphigerontia jezoensis
Ptycta johnsoni
Atrichadenotecnum sp.MY
Loensia variegata
Psocus sp.225
Metylophorus novaescotiae
Sigmatoneura kakisayap
Hemipsocus sp.196
Psilopsocus malayanus
Psocetae
Bertkauia crosbyana
Goja sp.
Epipsocus sp.MY
Epipsocetae
Allocaecilius sinensis
Pseudocaecilius citricola
Calopsocus furcatus
Heterocaecilius fuscus
Phallocaecilius hirsutus
Aaroniella badonneli
Philotarsus kwakiutl
'Homilopsocidea'
Archipsocus sp.209
Archipsocus sp.226
Archipsocetae
Psocomorpha
Haematomyzus elephantis
Pediculus humanus
Rhynchophthirina
Anoplura
Damalinia sika
Campanulotes compar
Physconelloides eurysema
'Ischnocera'
Menacanthus sp.272
Menopon gallinae
Heterodoxus calabyi
Laemobothrion cubense
Amblycera
Liposcelis brunnea
Liposcelis sp.KY2003
Liposcelis bostrychophila
Embidopsocus sp.400
Liposcelididae
Tapinella sp.192
Pachytroctidae
Badonnelia titei
Sphaeropsocidae
Troctomorpha
Stimulopalpus japonicus
Musapsocus sp.
Compsocus elegans
Amphientometae
Echmepteryx hageni
Neolepolepis occidentalis
Lepium sp.
Trogium pulsatorium
Psyllipsocus oculatus
Trogiomorpha
0.1

Figure5

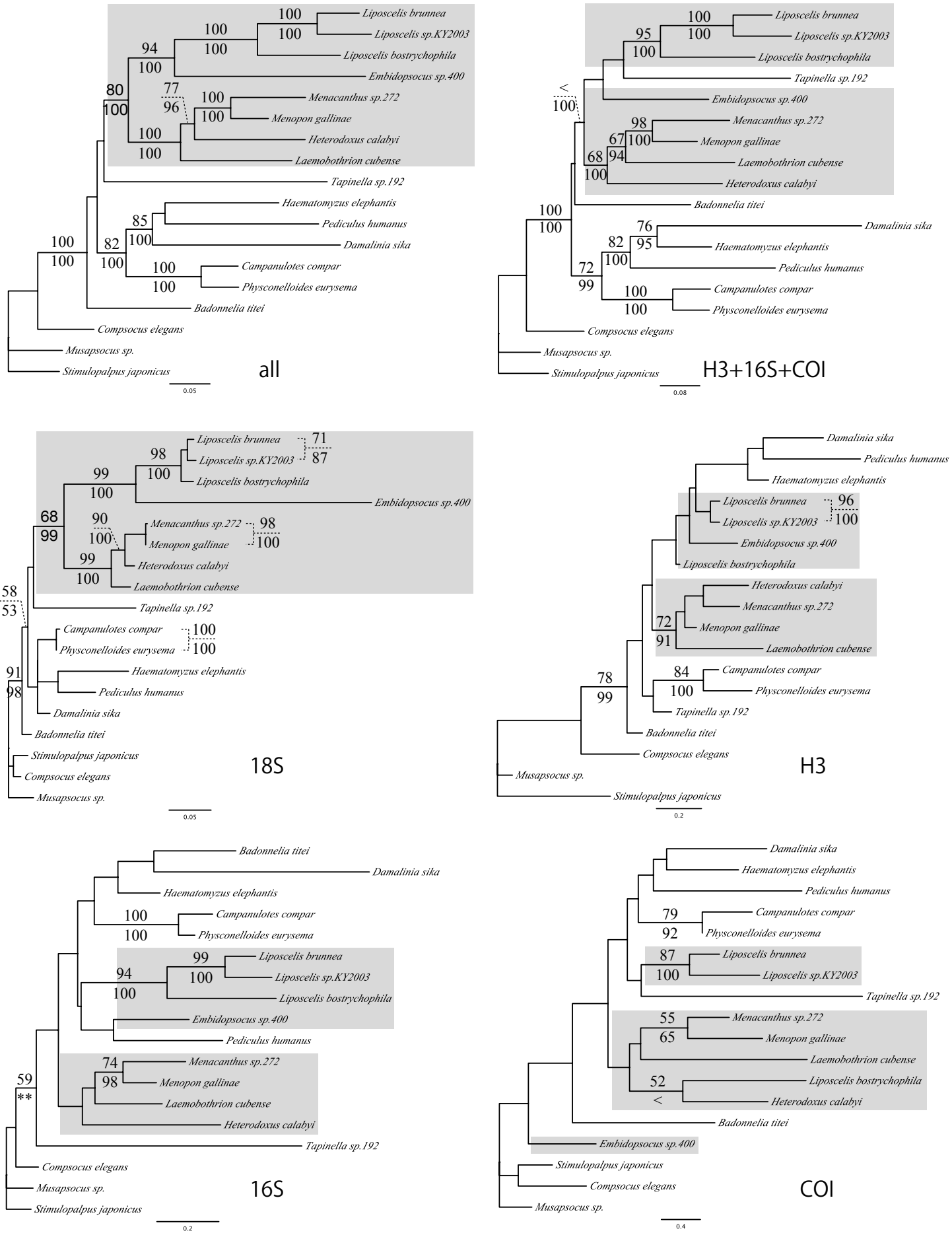
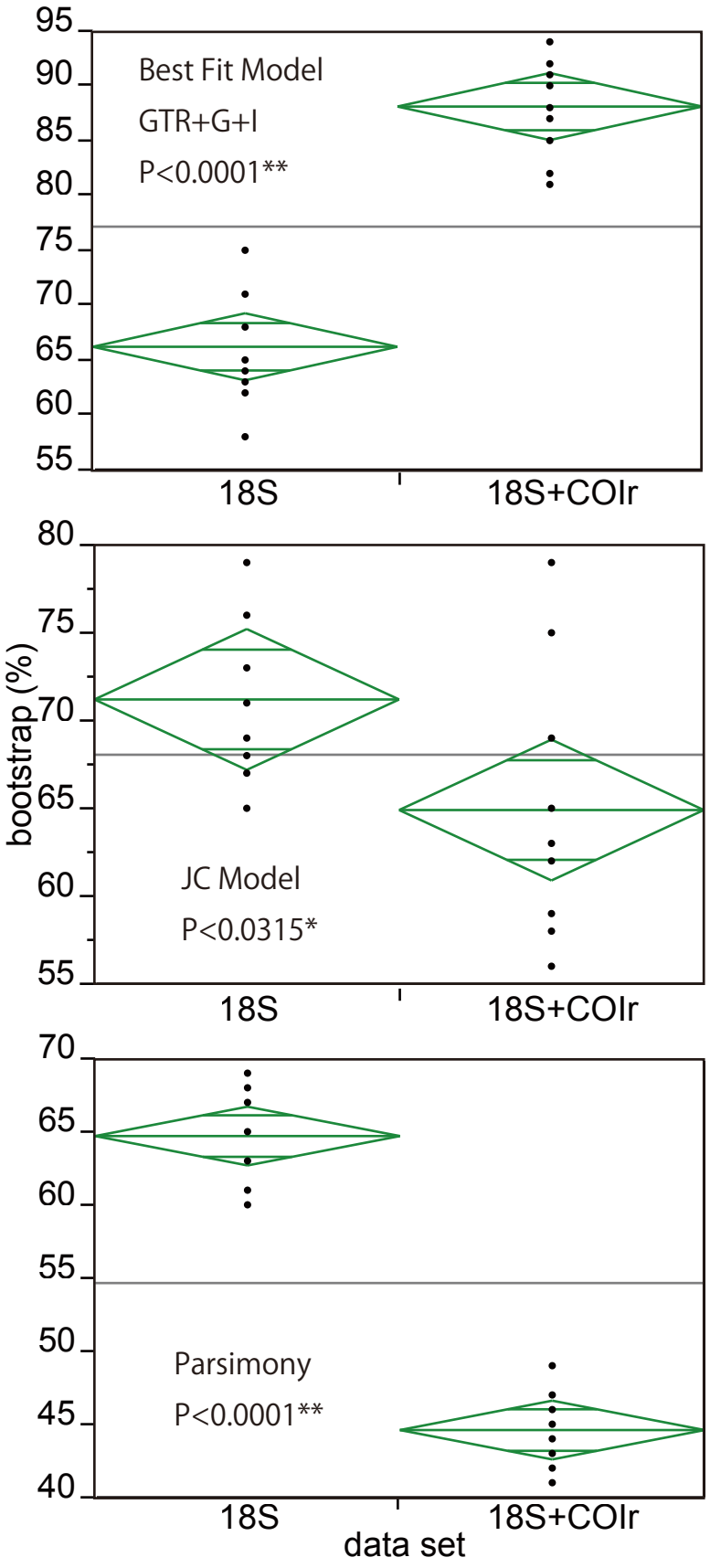

Liposcelis brunnea
Liposcelis sp.KY2003
Liposcelis bostrychophila
Embidopsocus sp.400
Menacanthus sp.272
Menopon gallinae
Heterodoxus calabyi
Laemobothrion cubense
Tapinella sp.192
Haematomyzus elephantis
Pediculus humanus
Damalinia sika
Campanulotes compar
Physconelloides eurysema
Badonnelia titei
Compsocus elegans
Musapsocus sp.
Stimulopalpus japonicus
all
H3+16S+COI
18S
H3
16S
COI

Figure6


95
90
85
80
75
70
65
60
55
Best Fit Model
GTR+G+I
P<0.0001**
18S
18S+COIr
80
75
70
65
60
55
bootstrap (%)
JC Model
P<0.0315*
18S
18S+COIr
70
65
60
55
50
45
40
Parsimony
P<0.0001**
18S
18S+COIr
data set

Tables1

Suborder	Family	*Species*	Voucher ID	Locality or Host	GenBank 18S	GenBank H3	GenBank 16S	GenBank COI	GenBank WG
Trogiomorpha	Psyllipsocidae	*Psyllipsocus oculatus*	Psocu.2.4.2002.12	Mexico	AY630455	DQ104776	DQ104748	GU569242	missing
Trogiomorpha	Trogiidae	*Lepium sp.*	Lpsp.11.17.2003.11	USA	AY630451	GU569312	GU569187	GU569244	missing
Trogiomorpha	Trogiidae	*Trogium pulsatorium*	Tgpul.11.17.2003.4	UK	AY630453	DQ104786	DQ104759	GU569243	missing
Trogiomorpha	Lepidopsocidae	*Neolepolepis occidentalis*	Neoocc.8.31.2001.13	USA	AY630446	DQ104779	DQ104751	GU569246	missing
Trogiomorpha	Lepidopsocidae	*Echmepteryx hageni*	Echag.1.16.2001.1	USA	AY630448	DQ104782	DQ104754	GU569245	missing
Psocomorpha	Archipsocidae	*Archipsocus sp.226*	KY226	Malaysia	AY630478	DQ104791	DQ104764	GU569248	missing
Psocomorpha	Archipsocidae	*Archipsocus sp.209*	KY209	Malaysia	GU569164	GU569313	GU569188	GU569247	GU569365
Psocomorpha	Epipsocidae	*Bertkauia crosbyana*	Becro.8.31.2991.14	USA	AY630537	DQ104793	DQ104766	GU569250	missing
Psocomorpha	Epipsocidae	*Goja sp.*	Gosp.12.4.2003.3	Costa Rica	AY630538	GU569315	GU569191	GU569251	missing
Psocomorpha	Epipsocidae	*Epipsocus sp.MY*	KY205	Malaysia	AY630539	GU569314	GU569189	GU569249	missing
Psocomorpha	Heipsocidae	*Hemipsocus sp.196*	KY196	Malaysia	AY630543	EF662139	EF662100	GU569252	GU569366
Psocomorpha	Psilopsocidae	*Psilopsocus malayanus*	KY195	Malaysia	AY630541	EF662140	EF662101	EF662064	GU569367
Psocomorpha	Psocidae	*Amphigerontia jezoensis*	KY213	Japan	AY630546	EF662143	EF662104	EF662067	GU569368
Psocomorpha	Psocidae	*Ptycta johnsoni*	KY235	Japan	AY630553	EF662175	AY139954	EF662093	missing
Psocomorpha	Psocidae	*Loensia variegata*	KY179	France	AY630549	EF662170	AY139953	AY374556	GU569369
Psocomorpha	Psocidae	*Atrichadenotecnum sp.MY*	KY238	Malaysia	AY630551	EF662156	EF662116	EF662079	GU569370
Psocomorpha	Psocidae	*Psocus sp.225*	KY225	Japan	AY630555	EF662162	EF662121	EF662084	GU569371
Psocomorpha	Psocidae	*Metylophorus novaescotiae*	Menov.2.3.2001.3	USA	AY630558	EF662154	AY275361	AY275286	missing
Psocomorpha	Psocidae	*Sigmatoneura kakisayap*	KY240	Malaysia	AY630557	GU569316	EF662112	EF662076	GU569372
Psocomorpha	Philotarsidae	*Aaroniella badonneli*	Aabad.8.31.2001.8	USA	AY630532	GU569317	GU569192	GU569253	missing
Psocomorpha	Philotarsidae	*Philotarsus kwakiutl*	Phkwa.11.17.2003.10	USA	AY630530	GU569318	GU569193	GU569254	missing
Psocomorpha	Calopsocidae	*Calopsocus furcatus*	KY199	Malaysia	AY630519	GU569319	GU569194	GU569255	missing
Psocomorpha	Pseudocaeciliidae	*Allocaecilius sinensis*	KY232	Japan	AY630526	DQ104796	DQ104769	GU569258	missing
Psocomorpha	Pseudocaeciliidae	*Heterocaecilius fuscus*	KY237	Japan	AY630520	DQ104795	DQ104768	GU569257	GU569374
Psocomorpha	Pseudocaeciliidae	*Phallocaecilius hirsutus*	KY217	Japan	AY630523	GU569320	GU569195	GU569256	GU569373
Psocomorpha	Pseudocaeciliidae	*Pseudocaecilius citricola*	Pccit.11.17.2003.12	Australia	AY630527	GU569321	GU569196	GU569259	missing
Psocomorpha	Elipsocidae	*Kilauella sp.*	Kisp.11.24.2003.10	Hawaii	AY630517	GU569329	GU569204	GU569267	missing
Psocomorpha	Mesopsocidae	*Mesopsocus hongkongensis*	KY224	Japan	AY630516	DQ104794	DQ104767	GU569268	missing
Psocomorpha	Lachesillidae	*Anomopsocus amabilis*	Anama.11.17.2003.9	USA	AY630509	GU569326	GU569201	GU569264	missing
Psocomorpha	Lachesillidae	*Eolachesilla chilensis*	KY214	Chile	AY630514	GU569328	GU569203	GU569266	GU569375
Psocomorpha	Lachesillidae	*Lachesilla forcepeta*	Lafor.8.31.2001.10	USA	AY630503	GU569327	GU569202	GU569265	missing
Psocomorpha	Ectopsocidae	*Ectopsocus meridionalis*	Epmer.2.3.2001.4	USA	AY630512	GU569322	GU569197	GU569260	missing
Psocomorpha	Ectopsocidae	*Ectopsocopsis cryptomeriae*	Etcry.11.17.2003.2	USA	AY630511	GU569323	GU569198	GU569261	missing
Psocomorpha	Peripsocidae	*Kaestneriella sp.*	Kasp.11.24.2003.5	USA	AY630506	GU569324	GU569199	GU569262	missing
Psocomorpha	Peripsocidae	*Peripsocus subfasciatus*	Pesub.2.3.2001.2	USA	AY630507	GU569325	GU569200	GU569263	missing
Psocomorpha	Asiopsocidae	*Asiopsocus sp.*	Assp.11.17.2003.3	USA	AY630481	GU569330	GU569205	GU569269	missing
Psocomorpha	Amphipsocidae	*Amphipsocus japonicus*	KY211	Japan	AF630489	GU569331	GU569206	GU569270	GU569376
Psocomorpha	Amphipsocidae	*Calocaecilius decipiens*	KY201	Malaysia	AY630485	GU569332	GU569207	GU569271	missing
Psocomorpha	Amphipsocidae	*Kolbia fusconervosa*	KY208	Japan	AY630487	GU569333	GU569208	GU569272	missing
Psocomorpha	Amphipsocidae	*Polypsocus corruptus*	Pocor.8.31.2001.6	USA	AY630488	GU569334	GU569209	GU569273	missing
Psocomorpha	Amphipsocidae	*Taeniostigma elongatum*	KY257	Malaysia	AY630486	GU569335	GU569210	GU569274	missing
Psocomorpha	Dasydemellidae	*Matsumuraiella radiopicta*	KY236	Japan	AY630493	DQ104797	DQ104770	GU569275	GU569377
Psocomorpha	Stenopsocidae	*Graphopsocus cruciatus*	Grcru.11.2.2001.5	USA	AY630490	GU569336	GU569211	GU569276	missing
Psocomorpha	Stenopsocidae	*Stenopsocus nigricellus*	KY241	Japan	AY630492	GU569338	GU569213	GU569278	missing
Psocomorpha	Stenopsocidae	*Stenopsocus aphidiformis*	KY219	Japan	AY630491	GU569337	GU569212	GU569277	GU569378
Psocomorpha	Caeciliusidae	*Dypsocus coleoptratus*	KY202	Japan	AY630482	GU569341	GU569216	GU569281	GU569379
Psocomorpha	Caeciliusidae	*Fuelleborniella sp.*	Fusp.11.24.2003.6	Gahna	AY630496	GU569339	GU569214	GU569279	missing
Psocomorpha	Caeciliusidae	*Isophanes sp.230*	KY230	Japan	AY630483	GU569342	GU569217	GU569282	missing
Psocomorpha	Caeciliusidae	*Pericaecilius sp.239*	KY239	Taiwan	AY630495	GU569340	GU569215	GU569280	missing
Psocomorpha	Caeciliusidae	*Valenzuela flavidus (USA)*	Vafla.8.31.2001.5	USA	AY630499	GU569343	GU569218	GU569283	missing
Psocomorpha	Caeciliusidae	*Xanthocaecilius sommermanae*	Xasom.8.31.2001.4	USA	AY630500	GU569344	GU569219	GU569284	missing
Troctomorpha	Amphientomidae	*Stimulopalpus japonicus*	Stjap.8.31.2001.15	USA	AY630459	GU569345	GU569220	GU569286	missing
Troctomorpha	Compsocidae	*Compsocus elegans*	Coele.3.24.2001.14	Costa Rica	AY630462	DQ104790	DQ104763	GU569287	missing
Troctomorpha	Musapsocidae	*Musapsocus sp.*	Musp.2.4.2002.13	Mexico	AY630461	DQ104789	DQ104762	GU569285	missing
Troctomorpha	Sphaeropsocidae	*Badonnelia titei*	Batit.12.4.2003.12	Switzerland	AY630464	GU569346	GU569221	GU569288	missing
Troctomorpha	Pachytroctidae	*Tapinella sp.192*	KY192	Malaysia	AY630466	GU569347	GU569222	GU569289	GU569380
Troctomorpha	Liposcelididae	*Embidopsocus sp.400*	KY400	Japan	GU569165	GU569348	GU569223	GU569290	missing
Troctomorpha	Liposcelididae	*Liposcelis brunnea*	KY245	Czech Rep.	AY630473	GU569349	GU569224	GU569291	missing
Troctomorpha	Liposcelididae	*Liposcelis sp.KY2003*	Lisp.11.2.2001.11	USA	AY630474	GU569350	GU569225	GU569292	missing
Troctomorpha	Liposcelididae	*Liposcelis bostrychophila*	Libos.8.31.2001.1	USA	AY630476	GU569351	GU569226	GU569293	missing
Amblycera	Boopidae	*Heterodoxus spiniger*	KY282	*Canis lupus*	GU569166	GU569352	GU569227	GU569294	missing
Amblycera	Laemobothriidae	*Laemobothrion cubense*	KY263	*Aramus guarana*	GU569167	GU569353	GU569228	GU569295	missing
Amblycera	Menoponidae	*Menacanthus sp.272*	KY272	*Corvinella corvina*	GU569168	GU569354	GU569229	GU569296	missing
Amblycera	Menoponidae	*Menopon gallinae*	KY237	*Gallus gallus*	GU569169	GU569355	GU569230	GU569297	missing
Ischnocera	Philopteridae	*Campanulotes compar*	KY269	*Columba livia*	GU569170	GU569356	GU569231	GU569298	GU569381
Ischnocera	Philopteridae	*Physconelloides eurysema*	KY261	*Columbina passerina*	GU569171	GU569357	GU569232	GU569299	GU569382
Ischnocera	Trichodectidae	*Damalinia sika*	KY397	*Cervus nippin*	GU569172	GU569358	GU569233	GU569300	missing
Rhynchophthirina	Haematomyzidae	*Haematomyzus elephantis*	KY281	*Elephas maximus*	AY077778	GU569361	GU569236	GU569302	missing
Anoplura	Pediculidae	*Pediculus humanus*	KY278	*Homo sapiens*	GU569174	GU569360	GU569235	GU569301	GU569384
Taxa only included in the expanded data set									
Troctomorpha	Liposcelididae	*Liposcelis paeta*	KY244	Czech Rep.	GU569173	GU569359	GU569234	missing	GU569383
Amblycera	Gyropidae	*Gliricola sp.274*	KY274	*Proechimys cuvieri*	GU569176	missing	GU569238	missing	GU569386
Amblycera	Trimenoponidae	*Cummingsia maculata*	KY275	*Lestoros inca*	GU569178	missing	GU569240	missing	GU569388
Amblycera	Menoponidae	*Trinoton querquedulae*	KY270	*Anas platyrhynchos*	GU569175	missing	GU569237	GU569303	GU569385
Amblycera	Menoponidae	*Hetenomenopon psittacum*	KY264	*Platycercus elegans*	GU569177	missing	GU569239	GU569304	GU569387
Ischnocera	Philotarsidae	*Anaticola crassicornis*	KY260	*Anas platyrhynchos*	GU569186	missing	GU569241	GU569311	GU569394
Ischnocera	Philotarsidae	*Columbicola passerinae*	KY262	*Columbina passerina*	GU569183	GU569364	missing	GU569308	GU569392
Ischnocera	Philotarsidae	*Degeeriella sp.265*	KY265	*Falco berigora*	GU569182	GU569363	missing	missing	GU569391
Ischnocera	Philopteridae	*Neopsittaconirmus eos*	KY271	*Eolophus roseicapillus*	GU569179	GU569362	missing	GU569305	missing
Ischnocera	Philotarsidae	*Pectinopygus bassani*	GenBank	*Morus serrator*	missing	missing	DQ463170	AF545743	DQ482939
Ischnocera	Trichodectidae	*Bovicola ovis*	KY267	*Ovis avies*	GU569184	missing	missing	GU569309	GU569393
Anoplura	Echinophthiriidae	*Echinophthirius horridus*	KY279	*Phoca vitulina*	GU569185	missing	missing	GU569310	missing
Anoplura	Haematopinidae	*Haematopinus tuberculatus*	KY280	*Bubalus bubalis*	GU569180	missing	missing	GU569306	GU569389
Anoplura	Hoplopleuridae	*Hoplopleura hirsuta*	KY277	*Sigmodon hispidus*	GU569181	missing	missing	GU569307	GU569390
Anoplura	Pedicinidae	*Pedicinus badii*	GenBank	*Procolobus gadii*	FJ267403	FJ267452	missing	EF152556	FJ267475
Anoplura	Pediculidae	*Phthirus pubis*	GenBank	*Homo sapiens*	AY077776	FJ267450	missing	EF152554	FJ267473

Tables2

	18S	const. 18S	18S+H3	+16S	+COI
Ambl + Lipo	68|99	76	67|93	89|100	87|100
Mena + Meno	98|100	Con.	97|100	100|100	96|100
Phys + Camp	100|100	Con.	99|100	100|100	97|100