Revision1: D-13-0016

Changes in Base Composition Bias of Nuclear and Mitochondrial Genes in Lice (Insecta: Psocodea)

Kazunori Yoshizawa, Kevin P. Johnson

K. Yoshizawa (corresponding author)

Systematic Entomology, Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

e-mail: psocid@res.agr.hokudai.ac.jp

K. P. Johnson

Illinois Natural History Survey, University of Illinois, Champaign, IL 61820, USA

e-mail: kpjohnson@illinois.edu

Abstract While it is well known that changes in the general processes of molecular evolution have occurred on a variety of timescales, the mechanisms underlying these changes are less well understood. Parasitic lice ("Phthiraptera") and their close relatives (infraorder Nanopsocetae of the insect order Psocodea) are a group of insects well known for their unusual features of molecular evolution. We examined changes in base composition across parasitic lice and bark lice. We identified substantial differences in percent GC content between the clade comprising parasitic lice plus closely related bark lice (= Nanopsocetae) versus all other bark lice. These changes occurred for both nuclear and mitochondrial protein coding and ribosomal RNA genes, often in the same direction. To evaluate whether correlations in base composition change also occurred within lineages, we used phylogenetically controlled comparisons, and in this case few significant correlations were identified. Examining more constrained sites (first/second codon positions and rRNA) revealed that, in comparison to the other bark lice, the GC content of parasitic lice and close relatives tended towards 50% either up from less than 50% GC or down from greater than 50% GC. In contrast, less constrained sites (third codon positions) in both nuclear and mitochondrial genes showed less of a consistent change of base composition in parasitic lice and very close relatives. We conclude that relaxed selection on this group of insects is a potential explanation of the change in base composition for both mitochondrial and nuclear genes, which could lead to nucleotide frequencies closer to random expectation (i.e., 50% GC) in the absence of any mutation bias. Evidence suggests this relaxed selection arose once in the non-parasitic common ancestor of Phthiraptera + Nanopsocetae and is not directly related to the evolution of the parasitism in lice.

Keywords: molecular evolution; GC content; relaxed selection; slightly deleterious mutation.

Introduction

Understanding variation in the process of DNA substitution across lineages is fundamental to understanding the factors that influence molecular evolution. For strictly neutral mutations, the DNA substitution rate is predicted to be directly proportional to mutation rate, independent of population size (Kimura 1962). However, not all mutations are strictly neutral, and many DNA mutations can be deleterious or beneficial in which the rate of substitution can be affected by population size (Ohta 1973, 1992). Therefore, it is expected that patterns of molecular evolution will vary across groups of organisms. Identifying whether differences in substitution rates are a result of direct selection or a result of underlying differences in mutation (e.g., non-selective effects such as drift on mutation repair mechanisms) is important for understanding this variation. Increases in DNA substitution rates in small populations might be expected when mutations are slightly deleterious (Ohta, 1973, 1992). For example, more non-synonymous substitutions occurred in island birds compared to mainland ones (Johnson and Seger 2001), providing evidence that slightly deleterious mutations might substitute more rapidly in small populations than in large populations. In contrast, differences in the rate of substitution in neutral mutations between lineages is more likely to be due to differences between them in the underlying mutation rate.

Variation in the process of molecular evolution across lineages is also known to confound phylogenetic analyses. Variation in substitution rates among lineages is a widely studied phylogenetic problem (e.g., Cameron and Crespi 1995; Huelsenbeck 1997). Such variation can lead to the problem of long branch attraction (Felsenstein 1978). Another feature of molecular evolution that can reduce the accuracy of phylogenetic estimation is variation in base composition (Galtier and Gouy 1995; Sheffield et al. 2009). In this case, taxa might be united based on similar base composition rather than shared evolutionary history. While there are methods being developed to deal with these problems, base composition bias continues to remain a difficulty rarely accounted for in phylogenetic studies (Sheffield et al. 2009). Finally, variation among lineages in RNA secondary structure makes aligning ribosomal DNA sequences difficult, which also potentially reduces accuracy of phylogenetic estimation (Rosenberg 2009).

All three of these phenomena occur in parasitic lice and their non-parasitic relatives (=Nanopsocetae of the insect order Psocodea). Parasitic lice and Nanopsocetae both have a dramatically elevated rate of nucleotide substitution in comparison to other insects (Rosenberg 2009; Page et al. 1998; Johnson et al. 2003; Yoshizawa and Johnson 2003; Johnson et al. 2004), substantial variation in

mitochondrial ribosomal secondary structure (Page et al. 2002; Yoshizawa and Johnson 2003), and strong base composition biases of mitochondrial genes (Yoshizawa and Johnson 2003). Potentially related phenomena in the louse genome include rearrangements of mitochondrial gene order (Shao et al. 2001; Cameron et al. 2007), the mitochondrial genome being separated into mini-chromosomes (Shao et al. 2009; Cameron et al. 2011; Wei et al. 2012), and very small nuclear genome size which is almost devoid of introns in each gene (Kirkness et al. 2010). Currently, these unusual evolutionary trends are known only from a limited sample of genes in lice and free-living close relatives (mostly mitochondrial genomes), and it is not known whether these trends also occur in nuclear genes (except for the accelerated substitution in 18S rDNA: Johnson et al. 2004; Yoshizawa & Johnson 2010) and, if so, how they might be correlated with patterns observed in mitochondrial genes.

To evaluate trends in the molecular evolution of a variety of genes, we sequenced both protein coding and ribosomal RNA genes from both the mitochondrial and nuclear genomes of lice and relatives. While also used to estimate phylogenetic relationships (Yoshizawa and Johnson 2010), this data set allows us to examine patterns of molecular evolution across different genes. Here, we examine changes of nucleotide base composition in parasitic lice (Phthiraptera) and their non-parasitic relatives (Nanopsocetae) between nuclear and mitochondrial genes and compare this to other free-living non-parasitic bark lice (order Psocodea). In particular, the mitochondrial genes of lice are less AT biased (Yoshizawa and Johnson, 2003) than other insects, which have highly AT biased mitochondrial genomes (Jermiin et al. 1994). Nucleotide base composition bias is a long recognized and long studied biological phenomenon; however, to our knowledge there has not been a study of base composition in both nuclear and mitochondrial genes simultaneously. Here we examine the base composition of both nuclear and mitochondrial protein coding and ribosomal RNA coding genes (nuclear 18S rDNA and Histone 3 and mitochondrial 16S rDNA and COI) and examine both heavily constrained and more lightly constrained portions of each gene. In the previous studies, nuclear and mitochondrial genomes have largely been analyzed independently for base composition biases (e.g., Page et al. 1998, 2002; Johnson et al. 2003; Cameron et al. 2011). However, if there is a cell wide bias in the availability of nucleotides, we might expect a correlation between nuclear and mitochondrial base composition biases. The present data set gives us an opportunity to explore the nature of any correlated changes between nuclear and mitochondrial base composition.

Materials and Methods

To compare changes in base composition among parasitic lice and bark lice, we used the data and maximum likelihood tree from a previous study (Fig. 1) (Yoshizawa and Johnson 2010). The data matrix included sequences from four genes (nuclear 18S rDNA and Histone 3 (H3) and mitochondrial 16S rDNA and COI) obtained from 69 taxa representing all the infraorders of Psocodea (parasitic lice and bark lice). These genes included both ribosomal and protein-coding genes from nuclear and mitochondrial genomes, and there was no missing data in the matrix.

For ribosomal DNA, we prepared a data set excluding all gapped regions. The gapped regions were excluded because the rDNA data set included a considerable portion of some sequences that could not be compared between groups, and the base composition of gapped regions for sequences with gaps in the alignment is undefined. For protein-coding genes, each gene was divided into two data sets by codon position to represent more and less constrained positions: 1st+2nd codon positions (H3(1st+2nd) and COI(1st+2nd)) and 3rd position (H3(3rd) and COI(3rd)).

The base composition of different gene partitions was calculated from the terminal taxa using MacClade 4.08 (Maddison and Maddison 2005) (Fig. 1). Because of phylogenetic non-independence, we did not analyze these data statistically. Each data set was separated into two categories *a priori*, Nanopsocetae (non-parasitic infraorder including Sphaeropsocidae, Pachytroctiidae and Liposcelididae) + Phthiraptera (parasitic lice) (NP) versus other Psocodea (OP). NP taxa were the groups identified in previous studies that appear to have the greatest difference in molecular evolutionary trends compare to OP taxa, including substitution rate, mitochondrial base composition bias, and rRNA secondary structure (Page et al. 1998, 2002; Johnson et al. 2003, 2004; Yoshizawa and Johnson 2003).

To explore whether correlations of changes in base composition between genes occurred along branches within each group, we controlled for potential phylogenetic correlation in the data by constructing phylogenetically independent comparisons using PDAP package (Midford et al. 2010) for Mesquite (Maddison and Maddison 2010). This software package implements the phylogenetically independent contrast method (PIC) (Felsenstein 1985). We also compared the independent contrast obtained from between the NP clade and its sister taxon (asterisk in Fig. 1) with all other contrasts to see if the change in the ancestor of the NP clade is an outlier (which would indicate a large change in base composition on the ancestral branch of this clade). Statistical tests were performed using JMP v. 8 (SAS 2009).

In addition, to assess whether the differences in base composition might be explained by differences in amino acid substitution rates, we also compared the non-synonymous to synonymous

132 substitution ratio (dN/dS ratio) of the protein-coding genes between two taxon categories using the
133 TestBranchDNDS batch file implemented in HyPhy v2.0 (Kosakovsky et al. 2005). The default setting
134 was selected except for the selection of the complete model for site-by-site variation. Original data
135 (Nexus file of sequence data: Supplementary Data 1; CSV file of GC% table: Supplementary Data 2) are
136 available from the journal web site, or http://kazu.psocodea.org/data/gc/.
137
138 **Results**
139 Strong differences in base composition between Nanopsocetae + Phthiraptera (NP) and other
140 Psocodea (OP) were evident for 18S, H3(1st+2nd), 16S, and COI(1st+2nd), but base composition was
141 less different between the two groups of taxa for third positions of protein coding genes (H3 and COI; Fig.
142 2, Table 1). GC content of 18S, 16S and COI(1st+2nd) in the NP taxa was almost consistently higher than
143 that of OP taxa (7-10 of top 10 taxa were NP taxa), whereas that of H3(1st+2nd) was almost consistently
144 lower in NP than OP taxa (9 of lowest 10 were NP) (Fig. 2).
145 Several genes and partitions, i.e., 18S, H3(1st+2nd), 16S, and COI(1st+2nd), of the OP taxa
146 exhibited a relatively constant range of GC% for that gene (Table 1, Fig. 2 white bars). In contrast, in the
147 variability of GC% of third positions for H3 and COI was large, even among the OP taxa (Table 1, Fig. 2
148 white bars). The variability in base composition of the Nanopsocetae + Phthiraptera (NP) taxa showed
149 similar trends, but results from the Bartlett's test showed that the variability in base composition of 16S
150 and COI(1st+2nd) were significantly higher compared to OP taxa (Table 1). Except for COI(3rd), the
151 standard deviations of base composition is always higher in NP taxa than that of OP. The variability in
152 base compositions of 18S and H3(1st+2nd) in the NP taxa was low across these taxa as well as the OP
153 taxa, while third position base composition of H3 and COI for the NP taxa was highly variable, similar to
154 the OP taxa (Table 1).
155 We used phylogenetically independent contrasts to determine whether the changes in base
156 composition were correlated between genes along branches across the phylogeny and also within each of
157 the major groups. There were no correlations among closely related taxa in the increase or decrease of GC
158 content between different data sets (Table 2). In a few comparisons, significant correlations were
159 identified (asterisks in Table 2) but none of them were significant when accounting for multiple testing.
160 When only the branches in the NP clade were analyzed, a correlation was identified only among
161 mitochondrial data sets, and this was not significant after accounting for multiple testing. Comparisons of
162 phylogenetically independent contrast absolute values revealed that the value for the ancestor of the NP

clade is higher for 16S dataset (1/68) but not for any other gene or partition (Table 3). In other comparisons, the phylogenetically independent contrast value for the ancestor of the NP clade was generally higher than average (8-17/68), but the contrast for the third positions of COI was even lower than the average (41/68) (Table 3).

Comparisons of dN/dS ratios between the NP and OP taxa revealed no significant difference for both H3 (P=0.66448) and COI (P=0.96373).

Discussion

The Nanopsocetae + Phthiraptera clade of parasitic lice and non-parasitic bark lice (Insecta: Psocodea) shows a different pattern of base composition than other members of Psocodea in all the genes examined, including nuclear and mitochondrial genes, both protein-coding and ribosomal RNA (Fig. 2). Analyses that accounted for phylogenetic correlation using independent contrasts, however, only showed this correlation for a limited number of comparisons, and none of these are significant after accounting for multiple testing (Table 2). That is, the overall changes nuclear and mitochondrial base composition in the ancestor of NP clade was not reflected in any additional correlated changes within this group. Such a correlation might be expected, for example, if there is a cell wide bias in the availability of nucleotides. Thus, these results suggest that the major change in base composition across genes between NP and OP taxa occurred once in the evolutionary history of this group and directional change in base composition is not continuing. However, the absolute value of the contrast for the ancestor of the NP clade was only significantly larger than all other contrasts for the 16S gene, and this is also not significant after accounting for multiple testing (Table 3).

A more detailed examination of these differences revealed some other interesting patterns, and may provide a more general mechanism for changes in both nuclear and mitochondrial base composition in this group of insects. First, for more constrained sites, such as first and second codon positions and ribosomal DNA, the GC% is relatively invariant throughout OP taxa compared to NP taxa (Fig. 2, Table 1). First and second positions of H3 are GC biased and first and second positions of COI and rDNA are AT biased. In theory, if mutation and substitution are random, over time an equal frequency of all four nucleotides will result (Li 1997). Therefore, the relatively biased base composition (i.e. away from 50% GC) in the OP taxa suggests either mutational bias (from environmental or cellular constraints) or that selection is operating to maintain base composition in relation to secondary structures of rRNA and mRNA, amino-acid composition, and 3D structure of protein (Foster et al. 1997; Chiusano et al. 1999). In

three out of four data sets of constrained sites, the nucleotide composition of the NP taxa is more GC biased compared to those of the OP taxa (Fig. 2, Table 1). One exception is that nucleotide composition of H3(1st+2nd) in the NP taxa are more AT biased compared to those of OP taxa, showing an opposite trend from other genes. However, in all cases the change in base composition in the NP taxa is in the direction of 50% GC, either up towards 50% for the ribosomal RNA genes and first and second positions of COI, or down towards 50% in the case of first and second positions for H3 (Fig. 2).

For less constrained third codon positions, the base composition is more highly variable even in the OP taxa, and this is the case for both the nuclear H3 and mitochondrial COI genes (Fig. 2, Table 1). Such variability in third position GC content has also been previously detected in *Drosophila* (Matsuo 2003). The differences of GC content between the NP and OP taxa are also much less striking for these third positions compared to more constrained rDNA or 1st and 2nd codon positions (Fig. 2, Table 1). Selection is much weaker on third codon positions regardless of whether proteins are under different selection pressures in different taxa. Therefore, the much weaker difference in GC% for third positions of H3 and COI between the OP and NP taxa may be explained by lower selection pressure on 3rd codon positions for both groups of taxa. Even so, there does appear to be some differences in the GC content of third positions for COI between the OP and NP taxa (Fig. 2). While the GC% for H3 fluctuates above and below 50%, the third codon position of the COI gene is strongly AT biased (Fig. 2), while also being highly variable. The mitochondrial genomes of insects are known to be strongly AT biased, which may be maintained by selective pressure (Jermiin et al. 1994). In the present case, the most strongly biased species contains only 3.5% of GC at the 3rd codon position of COI gene, and it seems likely that selection may be operating to maintain such a strong bias (Jermiin et al. 1994). Therefore, the nucleotide composition for third codon positions of mitochondrial genes is probably subject to selection, and the different GC% (closer to 50%) in third positions of COI for the NP taxa may also indicate a change in selective pressure.

Given the nature of all the biases observed, it seems the best explanation for the changes in base composition across these taxa may be relaxed selection. In particular, the changes in base composition in the NP taxa are in different directions for H3 (1st+2nd) (GC% lowered) compared to other genes (GC% increased). However, these changes converge towards 50% GC in all cases. This result suggests that there is not an overall change in base composition in one direction in both nuclear and mitochondrial genomes. Rather, in the absence of mutation biases, accumulation of random substitutions will tend to lead to a more equal frequency of nucleotide composition (Li 1997). Reduction of base composition bias is also

225 reported from *Drosophila* when selection is relaxed (Shields et al. 1988; Sharp and Li 1989; Moriyama
226 and Gojobori 1992). Therefore, less pronounced base composition bias in NP taxa can be explained by
227 relaxed selection and accumulation of more random substitution.
228 Significantly higher variance of the GC% for rDNA and 1st+2nd codon positions in the NP clade
229 compared to OP taxa also supports the hypothesis of relaxed selection against base composition bias (Fig.
230 2, Table 1). Given that the NP taxa are more closely related to each other than the OP taxa, it is even
231 more unlikely that the NP taxa have higher variance by chance alone. An exception is the GC% of third
232 codon positions (i.e., H3(3^{rd}) and COI(3rd)), which showed high variance in both the OP taxa and NP
233 taxa (Fig. 2, Table 1). The third codon positions of protein-coding genes are largely silent sites. In effect
234 these sites are already under relaxed selection, which would lead to high variance in base composition in
235 both OP and NP taxa. In particular, the GC% of third positions of H3 fluctuates with similar variance
236 around 50% GC for both NP and OP taxa (Table 1), suggesting these positions may be under equally
237 relaxed selection in both groups. Thus, the high variance of base composition at third sites in the OP taxa
238 is consistent with our conclusions about relaxed selection at more constrained sites. In contrast to the
239 nuclear H3 gene, third codon positions of the mitochondrial COI are highly AT biased, particularly in the
240 OP taxa, which cannot be explained by consistently relaxed selection at these sites. However, it is widely
241 known that the insect mitochondrial genome is highly AT biased probably either due to mutation bias or
242 selection favoring this bias (Jermiin et al. 1994). In our analysis, the GC% of COI(3^{rd}) in the NP taxa is
243 less AT biased (Fig. 2, Table 1), which might also be indicative of relaxed selection at these sites, if
244 selection maintains the strong AT bias in most insects.
245 Other unusual evolutionary trends observed in genomes of parasitic lice and non-parasitic close
246 relatives, such as accelerated substitution rates (Page et al. 1998; Johnson et al. 2003; Yoshizawa and
247 Johnson 2003; Johnson et al. 2004), modifications of ribosomal RNA secondary structure (Yoshizawa
248 and Johnson 2003; Page et al. 2002), and minicircularization of mitochondrial chromosome (Shao et al.
249 2009; Cameron et al. 2011; Wei et al. 2012) all involve what would normally be slightly deleterious
250 mutations. Therefore, these phenomena can also be explained by relaxed selection at molecular level. The
251 other explanations previously proposed for these phenomena in lice and their relatives are less applicable
252 than relaxed selection hypothesis. For example, short generation time was proposed as a potential factor
253 that could increase substitution rate (Hafner et al. 1994). However, this effect cannot explain other
254 phenomena including nucleotide base composition biases. The presence of symbionts is also known to
255 affect molecular evolutionary trends of organisms harboring them (Nigro and Prout 1990; Kambhampati

et al. 1992), and the highly AT-biased endosymbionts of parasitic lice (Sasaki-Fukatsu et al. 2006; Fukatsu et al. 2007) may have some relation to GC-bias of many genes in parasitic lice. However, this cannot explain the more AT-biased first and second positions of H3.

Relaxed selection at the molecular level has been identified in parasitic plants (Young and dePamphilis 2005), parasitic wasps (Dowton and Austin 1995; Dowton et al. 2009), and endosymbionts of insects (Clark et al. 1999; Woolfit and Bronham 2003). Thus, the parasitic lifestyle of lice may play a role in maintaining such relaxed selection in this group. One possibility is that relaxed selection is result of small effective population sizes in these insects (Ohta 1973, 1992), which provides a greater potential for the fixation of slightly deleterious mutations. It is generally argued that the effective population size of louse species is small due to highly structured populations with frequent bottlenecks (Page et al. 1998). However, our results indicate relaxed selection originated in the non-parasitic ancestor of the NP clade, and as such is not directly related to the origins of the parasitism in lice. It is currently unknown if such structured populations with frequent bottleneck events occur in Nanopsocetae. Furthermore, comparisons of dN/dS ratios revealed no significant differences between NP and OP taxa. If small effective population size is the major factor causing relaxed selection in NP taxa, we would expect increased accumulation of non-synonymous substitutions compared to synonymous ones, because synonymous substitutions are less likely to be deleterious and thus less affected by population size (Kimura 1962). However, the documented acceleration of substitution rates in parasitic lice is also evident at third codon positions (Johnson et al. 2003), which are mostly synonymous substitutions. No significant differences in dN/dS ratios were detected between NP and OP taxa, indicating a proportional increase in both synonymous and non-synonymous substitutions in NP taxa. Therefore, it seems unlikely that the changes in base composition occur as strictly the accumulation of slightly deleterious mutations in small populations.

Rather, relaxed selection may extend to the mutation repair machinery itself, because the strength of selection for mutation repair is directly proportional to the selection coefficient against deleterious mutations genome wide (Leigh 1970). In the case of parasites and endosymbionts, certain genes may lose function because of overall simplification in body form, in which case the number of genome wide potential deleterious mutations is reduced. This hypothesis predicts increased substitution rates of both neutral and selected sites and thus can account for both an accelerated substitution rate at third codon position (Johnson et al. 2003) and a lack of significant difference in dN/dS ratio between the NP and OP taxa. These unusual molecular evolutionary phenomena, including minicircularization of mitochondrial chromosome (Wei et al. 2012), are also observed in some non-parasitic Nanopsocetae and thus cannot be

attributed to the origins of parasitism in lice. Generally, members of Nanopsocetae have a smaller body size, simpler morphology with many reduced features (e.g., reduced size in eyes, reduction or absence of wings), and a more cryptic lifestyle (e.g. under bark) than other groups of free-living bark lice. Such a reduction in morphological and behavioral complexity may lead relaxed selection at many loci even for these non-parasitic bark lice. The possible features leading to relaxed selection in these insects deserve further study.

One more important point is that the changes in base composition in the parasitic lice and their close relatives may also affect the accuracy of molecular phylogenetic reconstruction. The parasitic lice have long been thought to have a single evolutionary origin and thus classified as a single order, Phthiraptera. However, a molecular phylogeny based on 18S rDNA gene sequences suggests that parasitic lice have evolved twice independently (Johnson et al. 2004). This hypothesis was tested by the data set also used in the current study (Yoshizawa and Johnson 2010), which provided tentative support for the polyphyly of lice hypothesis. However, significant base composition bias may have affected the results. A couple of recent studies, albeit with relatively poor taxon sampling , suggested monophyly of Phthiraptera (Wei et al. 2012; Johnson et al. 2013). Future tests of this hypothesis should adopt the methods that account for differences in base composition, though accurately doing this may be difficult (Sheffield et al. 2009).

Acknowledgments

We thank C. Lienhard, E. L. Mockford, Y. Kamimura, L. Durden, and J. Klicka for specimens. We also thank Joe Felsenstein and Colin Dale for suggestions in data analyses. Thanks are also due to three anonymous reviewers for constructive comments that are helpful in improving the earlier ms. KY thanks E. Hasegawa and M. Ôhara for allowing the use of their molecular facilities. This study was supported by Japan Society for the Promotion of Science grants 18770058 and 21770083 to KY and US National Science Foundation grant DEB-0612938 to KPJ.

References

Cameron D, Crespi BJ (1995) Do long branches attract flies? Nature 373:666.

Cameron SL, Johnson KP, Whiting MF (2007) The mitochondrial genome of the screamer louse *Bothriometopus* (Phthiraptera: Ischnocera): Effects of extensive gene rearrangements on the evolution of the genome as a whole. J Mol Evol 65:589-604

318 Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having
319 extreme base compositions. Mol Biol Evol 16:1586-1598
320 Cameron SL, Yoshizawa K, Mizukoshi A, Whiting MF, Johnson KP (2011) Mitochondrial genome
321 deletions and minicircles are common in lice (Insecta: Phthiraptera). BMC Genom 12:394
322 Chiusano ML, D'Onofrio G, Alvarez-Valin F, Jabbari K, Colonna G, Bernardi, G (1999) Correlations of
323 nucleotide substitution rates and base composition of mammalian coding sequences with protein
324 structure. Gene 238:23-31
325 Dowton M, Austin AD (1995) Increased genetic diversity in mitochondrial genes is correlated with the
326 evolution of parasitism in the Hymenoptera. J Mol Evol 16:298-309
327 Dowton M, Cameron SL, Dowavic JI, Austin AD, Whiting MF (2009) Characterisation of 66
328 mitochondrial gene rearrangements in the Hymenoptera reveal underlying trends in
329 mitochondrial genome evolution. Mol Biol Evol 26:1607-1617
330 Felsenstein J (1978) Case in which parsimony or compatibility methods will be positively misleading.
331 Syst Zool 27:401-410
332 Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1-15
333 Foster PG, Jermiin LS, Hickey DA (1997) Nucleotide composition bias affects amino acid content in
334 protein coded by animal mitochondria. J Mol Evol 44:282-288
335 Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, Sasaki-Fukatsu K, Yoshizawa K, Dale C, Clayton
336 DH (2007) Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to
337 endosymbionts of grain weevils and tsetse flies. App Env Microbiol 72:6660-6668
338 Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions.
339 Proc Natl Acad Sci USA 92:11317-11321
340 Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA (1994) Disparate rates
341 of molecular evolution in cospeciating hosts and parasites. Science 265:1087-1090
342 Huelsenbeck JP (1997) Is the Felsenstein zone a fly trap? Syst Biol 46:69-74.
343 Jermiin LS, Graur D, Lowe RM, Crozier RH (1994) Analysis of directional mutation pressur and
344 nucleotide content in mitochondrial cytochrome *b* genes. J Mol Evol 39:160-173
345 Johnson KP, Cruikshank RH, Adams RJ, Smith VS, Page RDM, Clayton DH (2003) Dramatically
346 elevated rates of mitochondrial substitution in lice (Insecta: Phthiraptera). Mol Phylog Evol
347 26:231-242
348 Johnson KP, Seger J (2001) Elevated rates of nonsynonymous substitution in island birds. Mol Biol Evol

18:874-881

Johnson KP, Walden KKO, Robertson HM (2013) Next-generation phylogenomics using a Target Restricted Assembly Method. Mol Phylogenet Evol 66:417-422.

Johnson KP, Yoshizawa K, Smith VS (2004) Multiple origins of parasitism in lice. Proc R Soc Lond B 271:1771-1776

Kambhampati S, Rai KS, Verleye DM (1992) Frequencies of mitochondrial DNA haplotypes in laboratory cage populations of the mosquito, *Aedes albopictus*. Genetics 132:205-209

Kimura M (1962) On the probability of fixation of mutant genes in a population. Genetics 47:713-719

Leigh EG Jr (1970) Natural selection and mutability. Am Nat 104:301-305

Li W-H (1997) Molecular Evolution. Sinauer Associates, Sunderland

Kirkness EF, Haas BJ, Sun W *et al*. (2010) Genome sequence of the human body louse and primary endosymbiont provide insights into the permanent parasitic lifestyle. PNAS 107:12168-12173

Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676-679

Maddison DR, Maddison WP (2005) MacClade ver 4.08. Sinauer Associates, Sunderland

Maddison WP, Maddison DR (2010) Mesquite:a molecular system for evolutionary analysis. Ver. 2.74. http://mesquiteproject.org. Accessed day Month year

Matsuo Y (2003) Evolution of the GC content of the histone 3 gene in seven *Drosophila* species. Genes & Genetic Systems 78:309-318.Accessed day Month year

Midford PE, Garland Jr T, Maddison WP (2010) PDAP package of Mesquite. Ver. 1.15. http://mesquiteproject.org/pdap_mesquite/index.html.Accessed day Month year

Moriyama EN, Gojobori T (1992) Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics 130:855-864

Nigro L, Prout T (1990) Is there selection on RFLP differences in mitochondrial DNA? Genetics 125:551-555

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96-98

Ohta T (1992) The nearly neutral theory of molecular evolution. Ann Rev Ecol Syst 23:263-286

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

Page RDM, Cruickshank R, Johnson KP (2002) Louse (Insecta: Phthiraptera) mitochondrial 12S rRNA secondary structure is highly variable. Insect Mol. Biol. 11:361-369

Rosenberg MS (ed) (2009) Sequence Alignment. University of California Press, Barkley

SAS Institute Inc (2009) JMP, version 8. SAS Institution, Cray

Sasaki-Fukatsu K, Koga R, Nikoh N, Yoshizawa K, Kasai S, Mihara M, Kobayashi M, Tomita T, Fukatsu T (2006) Symbiotic bacteria associated with stomach discs of human lice. App Env Microbiol 72:7349-7352

Shao R, Campbell NJH, Barker SC (2001) Numerous gene rearrangements in the mitochondrial genome of wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol Biol Evol 18:858-865

Shao R, Kirkness EF, Barker SC (2009) The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse *Pediculus humanus*. Genome Res 19:904-912

Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. J Mol Biol 28:398-402

Sheffield NC, Song H, Cameron SL, Whiting MF (2009) Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. Syst Biol 58:381-394

Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral: Evidence of selection among synonymous codons. Mol Biol Evol 5:704-716

Wei DD, Shao R, Yuan ML, Dou W, Barker SC, Wang JJ (2012) The multiparitite mitochondrial genome of *Liposcelis bostrichophila*: insights into the evolution of mitochondrial genomes in bilateral animals. PLoS ONE 7:e33973

Woolfit M, Bronham L (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol 20:1545-1555

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

Yoshizawa K, Johnson KP (2010) How stable is the "Polyphyly of Lice" hypothesis (Insecta: Psocodea)?: A comparison of phylogenetic signal in multiple genes. Mol Phylog Evol 55:939-951

Young ND, dePamphilis CW (2005) Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. BMC Evol Biol 5:16

Figure Legends

Figure 1. ML tree estimated from all data sets combined (redrawn from Yoshizawa & Johnson, 2010). Relative branch lengths of the tree is shown at the top-left corner. GC% of terminal taxon is shown at the tip of branches (18S, H3(1st,2nd), H3(3rd), 16S, COI(1st,2nd), COI(3rd)) followed by Taxon ID# (See Fig. 2).

Figure 2. Base composition of each taxon for each gene. Black bar indicates samples from Nanopsocetae+parasitic lice (NP taxa), and white bar indicates other Psocodea samples (OP taxa). See Fig. 1 for Taxon ID#.

Supplementary Data

Supplementary data 1. Nexus format data matrix of the gene data analyzed in this paper.

Supplementary data 2. CSV format data file of the GC% of terminal taxa.

Tables

Table 1. Comparisons of GC% between the NP (Nanopsocetae + Phthiraptera) and OP (other Psocodea) taxa. P-value indicates the results from the Bartlett's test for homogeneity of variance. Except for COI(3rd), standard deviation (SD) of NP taxa is always higher than that of OP. Asterisks indicate significant at 1% (**) or 0.1% (***).

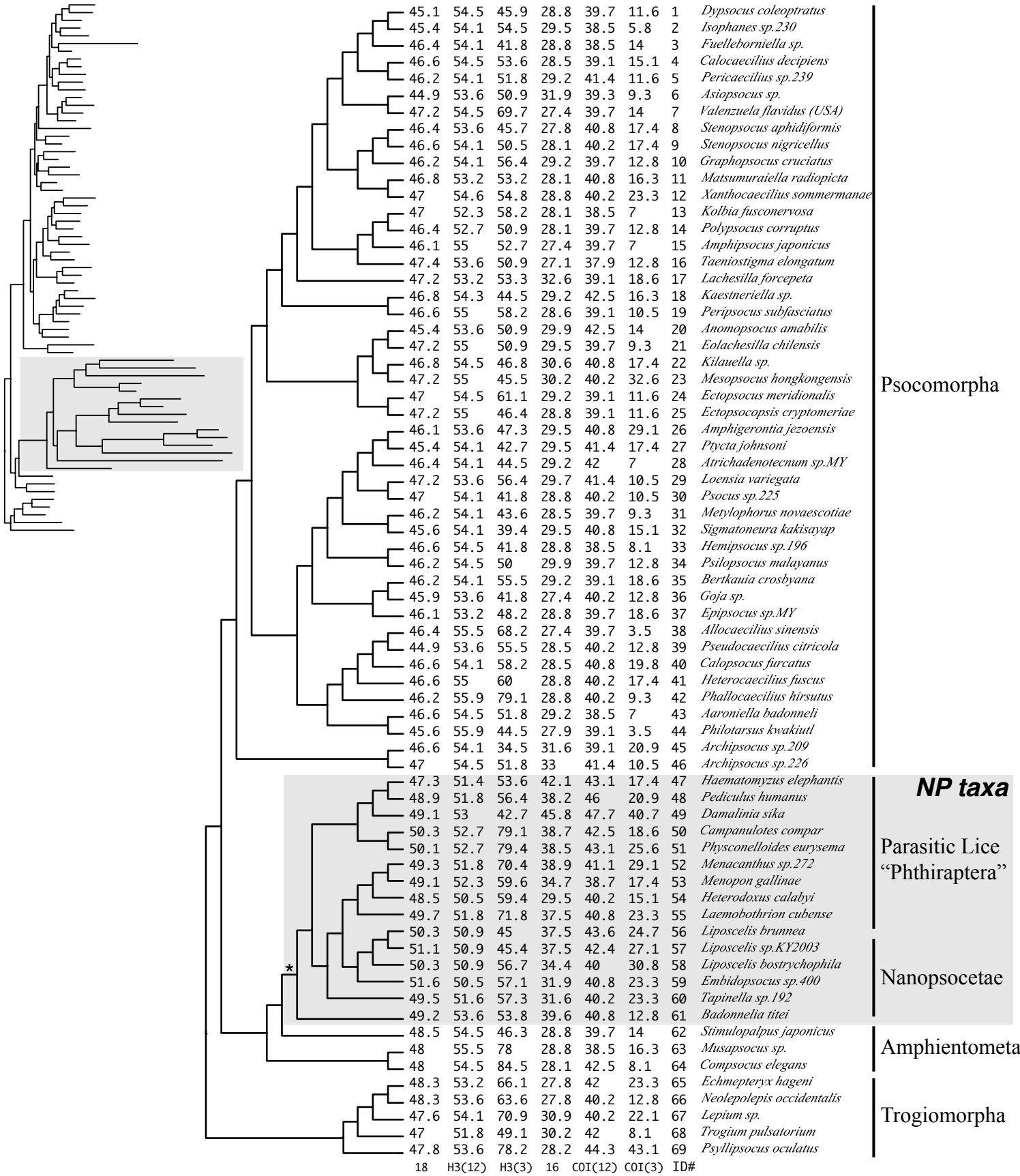
Gene | Result	OP range	OP mean	OP SD	NP range	NP mean	NP SD	Homogeneity of Variances (Bartlett: NP vs OP)
18S	44.9-48.5	46.6	0.828	47.3-51.6	49.6	1.060	P = 0.2299
H3(1st+2nd)	51.8-55.9	54.2	0.794	50.5-53.6	51.8	0.947	P = 0.3986
H3(3rd)	34.5-84.5	53.6	10.735	47.2-79.4	59.2	11.450	P = 0.7615
16S	27.1-33.0	29.0	1.247	29.5-45.8	37.1	4.179	P < 0.0001***
COI(1st+2nd)	37.9-44.3	40.1	1.274	38.7-47.7	42.1	2.395	P = 0.0012**
COI(3rd)	3.5-43.1	14.3	7.050	12.8-40.7	23.3	7.017	P = 0.9827

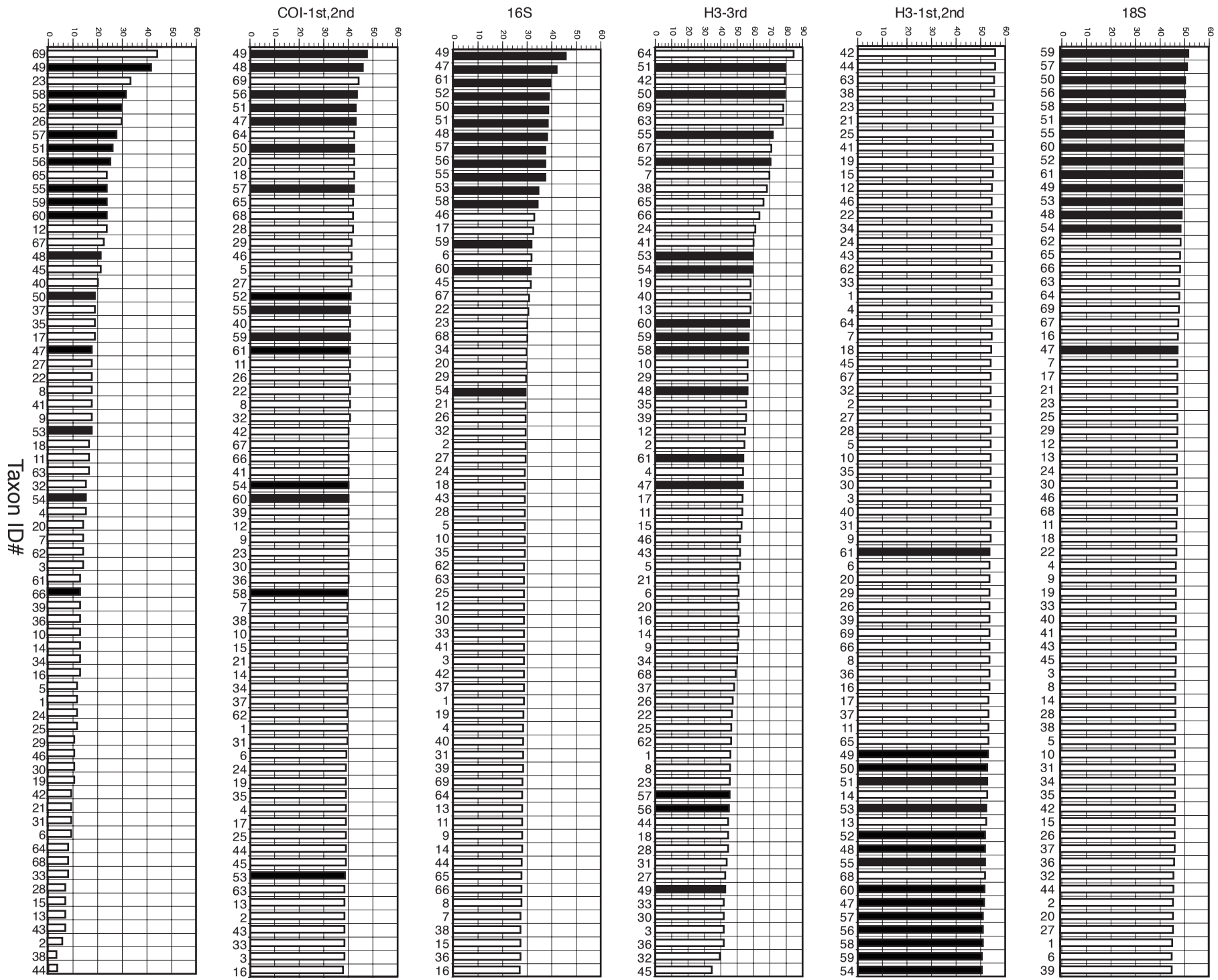
Table 2. Correlations of the phylogenetically independent contrasts (PIC) between genes. P-values were calculated for analyses including all taxa (total) and the Nanopsocetae-Phthiraptera clade analyzed alone (NP). * indicate significant at 5% level, but none of these are significant after correcting for multiple testing.

Comparison	PIC	P(total)	P(NP)
18S vs. H3(1st+2nd)		P=0.9557	P=0.6441
18S vs. H3(3rd)		P=0.0162*	P=0.1643
18S vs. 16S		P=0.8141	P=0.6404
18S vs. COI(1st+2nd)		P=0.3475	P=0.8386
18S vs. COI(3rd)		P=0.1779	P=0.4252
H3(1st+2nd) vs. H3(3rd)		P=0.0535	P=0.8735
H3(1st+2nd) vs. 16S		P=0.4859	P=0.0728
H3(1st+2nd) vs. COI(1st+2nd)		P=0.0571	P=0.9801
H3(1st+2nd) vs. COI(3rd)		P=0.1602	P=0.8863
H3(3rd) vs. 16S		P=0.8788	P=0.5843
H3(3rd) vs. COI(1st+2nd)		P=0.9713	P=0.5169
H3(3rd) vs. COI(3rd)		P=0.8623	P=0.7011
16S vs. COI(1st+2nd)		P=0.0145*	P=0.0313*
16S vs. COI(3rd)		P=0.0204*	P=0.0745
COI(1st+2nd) vs. COI(3rd)		P=0.0173*	P=0.0473*

Table 3. Comparisons of phylogenetically independent contrast absolute values for the ancestor of Nanopsocetae + Phthiraptera clade against all the other contrasts. *indicates significant at $P = 0.05$, but it is not significant after accounting for multiple testing.

Gene	Rank/Total Contrasts
18S	16/68
H3(1st+2nd)	8/68
H3(3rd)	17/68
16S	1/68*
COI(1st+2nd)	15/68
COI(3rd)	41/68


45.1 54.5 45.9 28.8 39.7 11.6 1 Dypsocus coleoptratus
45.4 54.1 54.5 29.5 38.5 5.8 2 Isophanes sp.230
46.4 54.1 41.8 28.8 38.5 14 3 Fuelleborniella sp.
46.6 54.5 53.6 28.5 39.1 15.1 4 Calocaecilius decipiens
46.2 54.1 51.8 29.2 41.4 11.6 5 Pericaecilius sp.239
44.9 53.6 50.9 31.9 39.3 9.3 6 Asiopsocus sp.
47.2 54.5 69.7 27.4 39.7 14 7 Valenzuela flavidus (USA)
46.4 53.6 45.7 27.8 40.8 17.4 8 Stenopsocus aphidiformis
46.6 54.1 50.5 28.1 40.2 17.4 9 Stenopsocus nigricellus
46.2 54.1 56.4 29.2 39.7 12.8 10 Graphopsocus cruciatus
46.8 53.2 53.2 28.1 40.8 16.3 11 Matsumuraiella radiopicta
47 54.6 54.8 28.8 40.2 23.3 12 Xanthocaecilius sommermanae
47 52.3 58.2 28.1 38.5 7 13 Kolbia fusconervosa
46.4 52.7 50.9 28.1 39.7 12.8 14 Polypsocus corruptus
46.1 55 52.7 27.4 39.7 7 15 Amphipsocus japonicus
47.4 53.6 50.9 27.1 37.9 12.8 16 Taeniostigma elongatum
47.2 53.2 53.3 32.6 39.1 18.6 17 Lachesilla forcepeta
46.8 54.3 44.5 29.2 42.5 16.3 18 Kaestneriella sp.
46.6 55 58.2 28.6 39.1 10.5 19 Peripsocus subfasciatus
45.4 53.6 50.9 29.9 42.5 14 20 Anomopsocus amabilis
47.2 55 50.9 29.5 39.7 9.3 21 Eolachesilla chilensis
46.8 54.5 46.8 30.6 40.8 17.4 22 Kilauella sp.
47.2 55 45.5 30.2 40.2 32.6 23 Mesopsocus hongkongensis
47 54.5 61.1 29.2 39.1 11.6 24 Ectopsocus meridionalis
47.2 55 46.4 28.8 39.1 11.6 25 Ectopsocopsis cryptomeriae
46.1 53.6 47.3 29.5 40.8 29.1 26 Amphigerontia jezoensis
45.4 54.1 42.7 29.5 41.4 17.4 27 Ptycta johnsoni
46.4 54.1 44.5 29.2 42 7 28 Atrichadenotecnum sp.MY
47.2 53.6 56.4 29.7 41.4 10.5 29 Loensia variegata
47 54.1 41.8 28.8 40.2 10.5 30 Psocus sp.225
46.2 54.1 43.6 28.5 39.7 9.3 31 Metylophorus novaescotiae
45.6 54.1 39.4 29.5 40.8 15.1 32 Sigmatoneura kakisayap
46.6 54.5 41.8 28.8 38.5 8.1 33 Hemipsocus sp.196
46.2 54.5 50 29.9 39.7 12.8 34 Psilopsocus malayanus
46.2 54.1 55.5 29.2 39.1 18.6 35 Bertkauia crosbyana
45.9 53.6 41.8 27.4 40.2 12.8 36 Goja sp.
46.1 53.2 48.2 28.8 39.7 18.6 37 Epipsocus sp.MY
46.4 55.5 68.2 27.4 39.7 3.5 38 Allocaecilius sinensis
44.9 53.6 55.5 28.5 40.2 12.8 39 Pseudocaecilius citricola
46.6 54.1 58.2 28.5 40.8 19.8 40 Calopsocus furcatus
46.6 55 60 28.8 40.2 17.4 41 Heterocaecilius fuscus
46.2 55.9 79.1 28.8 40.2 9.3 42 Phallocaecilius hirsutus
46.6 54.5 51.8 29.2 38.5 7 43 Aaroniella badonneli
45.6 55.9 44.5 27.9 39.1 3.5 44 Philotarsus kwakiutl
46.6 54.1 34.5 31.6 39.1 20.9 45 Archipsocus sp.209
47 54.5 51.8 33 41.4 10.5 46 Archipsocus sp.226
47.3 51.4 53.6 42.1 43.1 17.4 47 Haematomyzus elephantis
48.9 51.8 56.4 38.2 46 20.9 48 Pediculus humanus
49.1 53 42.7 45.8 47.7 40.7 49 Damalinia sika
50.3 52.7 79.1 38.7 42.5 18.6 50 Campanulotes compar
50.1 52.7 79.4 38.5 43.1 25.6 51 Physconelloides eurysema
49.3 51.8 70.4 38.9 41.1 29.1 52 Menacanthus sp.272
49.1 52.3 59.6 34.7 38.7 17.4 53 Menopon gallinae
48.5 50.5 59.4 29.5 40.2 15.1 54 Heterodoxus calabyi
49.7 51.8 71.8 37.5 40.8 23.3 55 Laemobothrion cubense
50.3 50.9 45 37.5 43.6 24.7 56 Liposcelis brunnea
51.1 50.9 45.4 37.5 42.4 27.1 57 Liposcelis sp.KY2003
50.3 50.9 56.7 34.4 40 30.8 58 Liposcelis bostrychophila
51.6 50.5 57.1 31.9 40.8 23.3 59 Embidopsocus sp.400
49.5 51.6 57.3 31.6 40.2 23.3 60 Tapinella sp.192
49.2 53.6 53.8 39.6 40.8 12.8 61 Badonnelia titei
48.5 54.5 46.3 28.8 39.7 14 62 Stimulopalpus japonicus
48 55.5 78 28.8 38.5 16.3 63 Musapsocus sp.
48 54.5 84.5 28.1 42.5 8.1 64 Compsocus elegans
48.3 53.2 66.1 27.8 42 23.3 65 Echmepteryx hageni
48.3 53.6 63.6 27.8 40.2 12.8 66 Neolepolepis occidentalis
47.6 54.1 70.9 30.9 40.2 22.1 67 Lepium sp.
47 51.8 49.1 30.2 42 8.1 68 Trogium pulsatorium
47.8 53.6 78.2 28.2 44.3 43.1 69 Psyllipsocus oculatus
18 H3(12) H3(3) 16 COI(12) COI(3) ID#
terminal GC%
Psocomorpha
NP taxa
Parasitic Lice
"Phthiraptera"
Nanopsocetae
Amphientometae
Trogiomorpha
*


GC%
COI-1st,2nd
16S
H3-3rd
H3-1st,2nd
18S
Taxon ID#