



Title	Changes in base composition bias of nuclear and mitochondrial genes in lice (Insecta: Psocodea).
Author(s)	Yoshizawa, Kazunori; Johnson, Kevin P
Citation	Genetica, 141(10-12), 491-499 <a href="https://doi.org/10.1007/s10709-013-9748-z">https://doi.org/10.1007/s10709-013-9748-z</a>
Issue Date	2013-12
Doc URL	<a href="http://hdl.handle.net/2115/57601">http://hdl.handle.net/2115/57601</a>
Rights	The final publication is available at <a href="http://link.springer.com">link.springer.com</a>
Type	article (author version)
File Information	2013Gene.pdf



[Instructions for use](#)

1 *Revision1: D-13-0016*

2

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

5

6 **Kazunori Yoshizawa, Kevin P. Johnson**

7

8 K. Yoshizawa (corresponding author)

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

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

11

12 K. P. Johnson

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

14 e-mail: kpjohnson@illinois.edu

15

16

17

18 **Abstract** While it is well known that changes in the general processes of molecular evolution have  
19 occurred on a variety of timescales, the mechanisms underlying these changes are less well understood.  
20 Parasitic lice (“Phthiraptera”) and their close relatives (infraorder Nanopsocetae of the insect order  
21 Psocodea) are a group of insects well known for their unusual features of molecular evolution. We  
22 examined changes in base composition across parasitic lice and bark lice. We identified substantial  
23 differences in percent GC content between the clade comprising parasitic lice plus closely related bark  
24 lice (= Nanopsocetae) versus all other bark lice. These changes occurred for both nuclear and  
25 mitochondrial protein coding and ribosomal RNA genes, often in the same direction. To evaluate whether  
26 correlations in base composition change also occurred within lineages, we used phylogenetically  
27 controlled comparisons, and in this case few significant correlations were identified. Examining more  
28 constrained sites (first/second codon positions and rRNA) revealed that, in comparison to the other bark  
29 lice, the GC content of parasitic lice and close relatives tended towards 50% either up from less than 50%  
30 GC or down from greater than 50% GC. In contrast, less constrained sites (third codon positions) in both  
31 nuclear and mitochondrial genes showed less of a consistent change of base composition in parasitic lice  
32 and very close relatives. We conclude that relaxed selection on this group of insects is a potential  
33 explanation of the change in base composition for both mitochondrial and nuclear genes, which could  
34 lead to nucleotide frequencies closer to random expectation (i.e., 50% GC) in the absence of any mutation  
35 bias. Evidence suggests this relaxed selection arose once in the non-parasitic common ancestor of  
36 Phthiraptera + Nanopsocetae and is not directly related to the evolution of the parasitism in lice.  
37  
38 **Keywords:** molecular evolution; GC content; relaxed selection; slightly deleterious mutation.

39 **Introduction**

40           Understanding variation in the process of DNA substitution across lineages is fundamental to  
41 understanding the factors that influence molecular evolution. For strictly neutral mutations, the DNA  
42 substitution rate is predicted to be directly proportional to mutation rate, independent of population size  
43 (Kimura 1962). However, not all mutations are strictly neutral, and many DNA mutations can be  
44 deleterious or beneficial in which the rate of substitution can be affected by population size (Ohta 1973,  
45 1992). Therefore, it is expected that patterns of molecular evolution will vary across groups of organisms.  
46 Identifying whether differences in substitution rates are a result of direct selection or a result of  
47 underlying differences in mutation (e.g., non-selective effects such as drift on mutation repair  
48 mechanisms) is important for understanding this variation. Increases in DNA substitution rates in small  
49 populations might be expected when mutations are slightly deleterious (Ohta, 1973, 1992). For example,  
50 more non-synonymous substitutions occurred in island birds compared to mainland ones (Johnson and  
51 Seger 2001), providing evidence that slightly deleterious mutations might substitute more rapidly in small  
52 populations than in large populations. In contrast, differences in the rate of substitution in neutral  
53 mutations between lineages is more likely to be due to differences between them in the underlying  
54 mutation rate.

55           Variation in the process of molecular evolution across lineages is also known to confound  
56 phylogenetic analyses. Variation in substitution rates among lineages is a widely studied phylogenetic  
57 problem (e.g., Cameron and Crespi 1995; Huelsenbeck 1997). Such variation can lead to the problem of  
58 long branch attraction (Felsenstein 1978). Another feature of molecular evolution that can reduce the  
59 accuracy of phylogenetic estimation is variation in base composition (Galtier and Gouy 1995; Sheffield et  
60 al. 2009). In this case, taxa might be united based on similar base composition rather than shared  
61 evolutionary history. While there are methods being developed to deal with these problems, base  
62 composition bias continues to remain a difficulty rarely accounted for in phylogenetic studies (Sheffield  
63 et al. 2009). Finally, variation among lineages in RNA secondary structure makes aligning ribosomal  
64 DNA sequences difficult, which also potentially reduces accuracy of phylogenetic estimation (Rosenberg  
65 2009).

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

70 mitochondrial ribosomal secondary structure (Page et al. 2002; Yoshizawa and Johnson 2003), and strong  
71 base composition biases of mitochondrial genes (Yoshizawa and Johnson 2003). Potentially related  
72 phenomena in the louse genome include rearrangements of mitochondrial gene order (Shao et al. 2001;  
73 Cameron et al. 2007), the mitochondrial genome being separated into mini-chromosomes (Shao et al.  
74 2009; Cameron et al. 2011; Wei et al. 2012), and very small nuclear genome size which is almost devoid  
75 of introns in each gene (Kirkness et al. 2010). Currently, these unusual evolutionary trends are known  
76 only from a limited sample of genes in lice and free-living close relatives (mostly mitochondrial  
77 genomes), and it is not known whether these trends also occur in nuclear genes (except for the accelerated  
78 substitution in 18S rDNA: Johnson et al. 2004; Yoshizawa & Johnson 2010) and, if so, how they might  
79 be correlated with patterns observed in mitochondrial genes.

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

99

## 100 **Materials and Methods**

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

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

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

121 To explore whether correlations of changes in base composition between genes occurred along  
122 branches within each group, we controlled for potential phylogenetic correlation in the data by  
123 constructing phylogenetically independent comparisons using PDAP package (Midford et al. 2010) for  
124 Mesquite (Maddison and Maddison 2010). This software package implements the phylogenetically  
125 independent contrast method (PIC) (Felsenstein 1985). We also compared the independent contrast  
126 obtained from between the NP clade and its sister taxon (asterisk in Fig. 1) with all other contrasts to see  
127 if the change in the ancestor of the NP clade is an outlier (which would indicate a large change in base  
128 composition on the ancestral branch of this clade). Statistical tests were performed using JMP v. 8 (SAS  
129 2009).

130 In addition, to assess whether the differences in base composition might be explained by  
131 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

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

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

169

## 170 **Discussion**

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

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



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

200 For less constrained third codon positions, the base composition is more highly variable even in  
201 the OP taxa, and this is the case for both the nuclear H3 and mitochondrial COI genes (Fig. 2, Table 1).  
202 Such variability in third position GC content has also been previously detected in *Drosophila* (Matsuo  
203 2003). The differences of GC content between the NP and OP taxa are also much less striking for these  
204 third positions compared to more constrained rDNA or 1st and 2nd codon positions (Fig. 2, Table 1).  
205 Selection is much weaker on third codon positions regardless of whether proteins are under different  
206 selection pressures in different taxa. Therefore, the much weaker difference in GC% for third positions of  
207 H3 and COI between the OP and NP taxa may be explained by lower selection pressure on 3rd codon  
208 positions for both groups of taxa. Even so, there does appear to be some differences in the GC content of  
209 third positions for COI between the OP and NP taxa (Fig. 2). While the GC% for H3 fluctuates above and  
210 below 50%, the third codon position of the COI gene is strongly AT biased (Fig. 2), while also being  
211 highly variable. The mitochondrial genomes of insects are known to be strongly AT biased, which may  
212 be maintained by selective pressure (Jermini et al. 1994). In the present case, the most strongly biased  
213 species contains only 3.5% of GC at the 3rd codon position of COI gene, and it seems likely that selection  
214 may be operating to maintain such a strong bias (Jermini et al. 1994). Therefore, the nucleotide  
215 composition for third codon positions of mitochondrial genes is probably subject to selection, and the  
216 different GC% (closer to 50%) in third positions of COI for the NP taxa may also indicate a change in  
217 selective pressure.

218 Given the nature of all the biases observed, it seems the best explanation for the changes in base  
219 composition across these taxa may be relaxed selection. In particular, the changes in base composition in  
220 the NP taxa are in different directions for H3 (1st+2nd) (GC% lowered) compared to other genes (GC%  
221 increased). However, these changes converge towards 50% GC in all cases. This result suggests that there  
222 is not an overall change in base composition in one direction in both nuclear and mitochondrial genomes.  
223 Rather, in the absence of mutation biases, accumulation of random substitutions will tend to lead to a  
224 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<sup>rd</sup>) and COI(3<sup>rd</sup>)), 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 (Jermini et al. 1994). In our analysis, the GC% of COI(3<sup>rd</sup>) 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

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

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

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

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

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

304

### 305 **Acknowledgments**

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

312

### 313 **References**

- 314 Cameron D, Crespi BJ (1995) Do long branches attract flies? *Nature* 373:666.
- 315 Cameron SL, Johnson KP, Whiting MF (2007) The mitochondrial genome of the screamer louse  
316 *Bothriometopus* (Phthiraptera: Ischnocera): Effects of extensive gene rearrangements on the  
317 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, Jermini 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 Jermini LS, Graur D, Lowe RM, Crozier RH (1994) Analysis of directional mutation pressure 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*

349 18:874-881

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

352 Johnson KP, Yoshizawa K, Smith VS (2004) Multiple origins of parasitism in lice. *Proc R Soc Lond B*  
353 271:1771-1776

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

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

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

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

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

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

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

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

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

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

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

372 Nigro L, Prout T (1990) Is there selection on RFLP differences in mitochondrial DNA? *Genetics*  
373 125:551-555

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

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

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

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

380 Rosenberg MS (ed) (2009) Sequence Alignment. University of California Press, Barkley  
381 SAS Institute Inc (2009) JMP, version 8. SAS Institution, Cray  
382 Sasaki-Fukatsu K, Koga R, Nikoh N, Yoshizawa K, Kasai S, Mihara M, Kobayashi M, Tomita T,  
383 Fukatsu T (2006) Symbiotic bacteria associated with stomach discs of human lice. App Env  
384 Microbiol 72:7349-7352  
385 Shao R, Campbell NJH, Barker SC (2001) Numerous gene rearrangements in the mitochondrial genome  
386 of wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol Biol Evol 18:858-865  
387 Shao R, Kirkness EF, Barker SC (2009) The single mitochondrial chromosome typical of animals has  
388 evolved into 18 minichromosomes in the human body louse *Pediculus humanus*. Genome Res  
389 19:904-912  
390 Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. J Mol Biol 28:398-402  
391 Sheffield NC, Song H, Cameron SL, Whiting MF (2009) Nonstationary evolution and compositional  
392 heterogeneity in beetle mitochondrial phylogenomics. Syst Biol 58:381-394  
393 Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral:  
394 Evidence of selection among synonymous codons. Mol Biol Evol 5:704-716  
395 Wei DD, Shao R, Yuan ML, Dou W, Barker SC, Wang JJ (2012) The multiparitite mitochondrial genome  
396 of *Liposcelis bostrichophila*: insights into the evolution of mitochondrial genomes in bilateral  
397 animals. PLoS ONE 7:e33973  
398 Woolfit M, Bronham L (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi  
399 with small effective population sizes. Mol Biol Evol 20:1545-1555  
400 Yoshizawa K, Johnson KP (2003) Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and  
401 elevated rate of evolution in mitochondrial 12S and 16S rDNA. Mol Phylog Evol 29:102-114  
402 Yoshizawa K, Johnson KP (2010) How stable is the "Polyphyly of Lice" hypothesis (Insecta: Psocodea)?:  
403 A comparison of phylogenetic signal in multiple genes. Mol Phylog Evol 55:939-951  
404 Young ND, dePamphilis CW (2005) Rate variation in parasitic plants: correlated and uncorrelated  
405 patterns among plastid genes of different function. BMC Evol Biol 5:16  
406  
407

408 **Figure Legends**

409

410 Figure 1. ML tree estimated from all data sets combined (redrawn from Yoshizawa & Johnson, 2010).

411 Relative branch lengths of the tree is shown at the top-left corner. GC% of terminal taxon is shown at the  
412 tip of branches (18S, H3(1<sup>st</sup>,2<sup>nd</sup>), H3(3<sup>rd</sup>), 16S, COI(1<sup>st</sup>,2<sup>nd</sup>), COI(3<sup>rd</sup>)) followed by Taxon ID# (See Fig.  
413 2).

414

415 Figure 2. Base composition of each taxon for each gene. Black bar indicates samples from

416 Nanopsocetae+parasitic lice (NP taxa), and white bar indicates other Psocodea samples (OP taxa). See

417 Fig. 1 for Taxon ID#.

418

419



420 **Supplementary Data**

421

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

423

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

425

426 **Tables**

427

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

432  
 433  
 434

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

441

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

446

447	<u>Comparison   PIC</u>	<u>P(total)</u>	<u>P(NP)</u>
448	18S vs. H3(1st+2nd)	P=0.9557	P=0.6441
449	18S vs. H3(3rd)	P=0.0162*	P=0.1643
450	18S vs. 16S	P=0.8141	P=0.6404
451	18S vs. COI(1st+2nd)	P=0.3475	P=0.8386
452	18S vs. COI(3rd)	P=0.1779	P=0.4252
453	H3(1st+2nd) vs. H3(3rd)	P=0.0535	P=0.8735
454	H3(1st+2nd) vs. 16S	P=0.4859	P=0.0728
455	H3(1st+2nd) vs. COI(1st+2nd)	P=0.0571	P=0.9801
456	H3(1st+2nd) vs. COI(3rd)	P=0.1602	P=0.8863
457	H3(3rd) vs. 16S	P=0.8788	P=0.5843
458	H3(3rd) vs. COI(1st+2nd)	P=0.9713	P=0.5169
459	H3(3rd) vs. COI(3rd)	P=0.8623	P=0.7011
460	16S vs. COI(1st+2nd)	P=0.0145*	P=0.0313*
461	16S vs. COI(3rd)	P=0.0204*	P=0.0745
462	<u>COI(1st+2nd) vs. COI(3rd)</u>	<u>P=0.0173*</u>	<u>P=0.0473*</u>

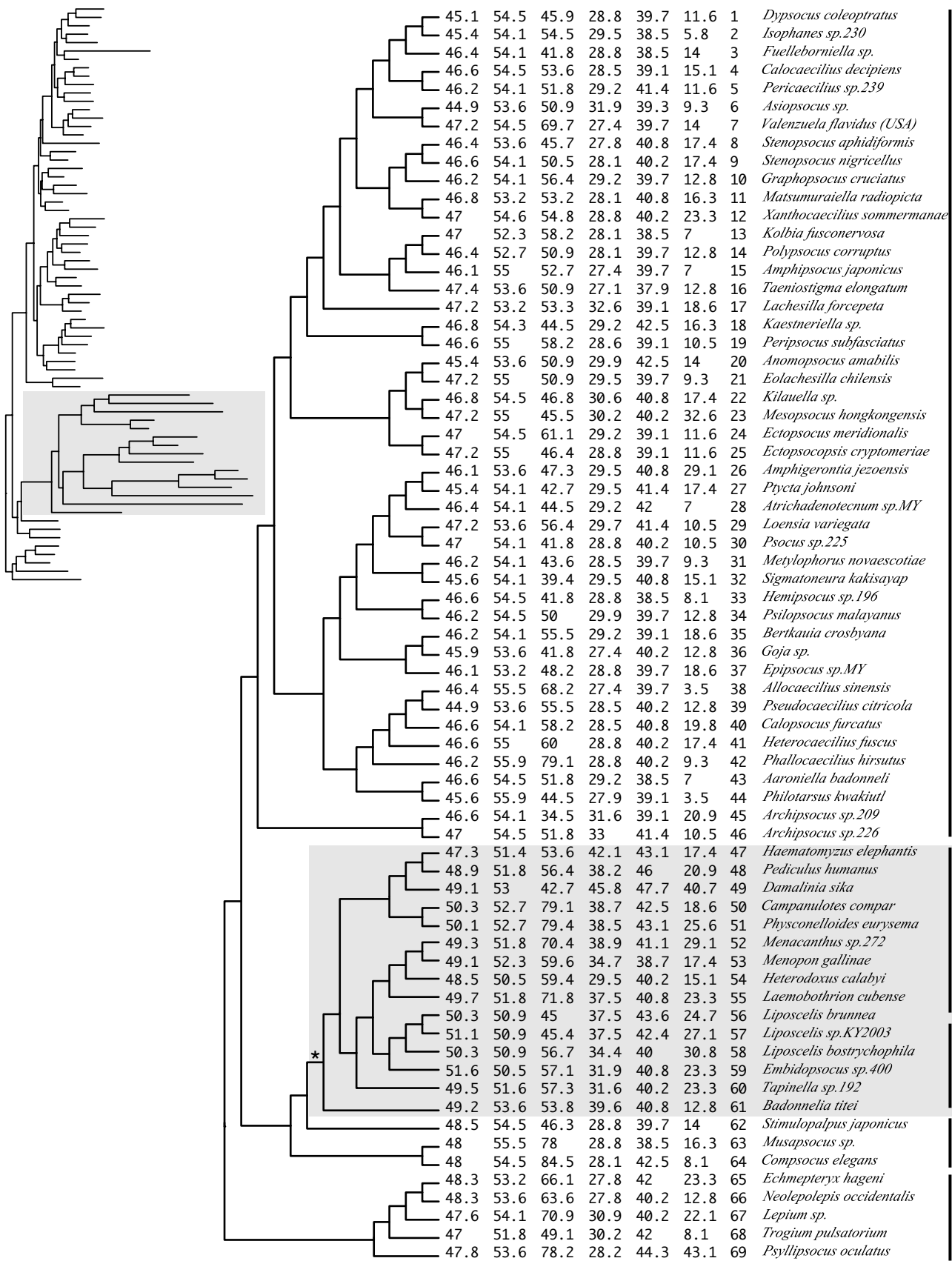
463

464

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

468

469	<u>Gene</u>	<u>Rank/Total Contrasts</u>
470	18S	16/68
471	H3(1st+2nd)	8/68
472	H3(3rd)	17/68
473	16S	1/68*
474	COI(1st+2nd)	15/68
475	<u>COI(3rd)</u>	<u>41/68</u>



Psocomorpha

**NP taxa**

Parasitic Lice  
"Phthiraptera"

Nanopsocetae

Amphientometae

Trogiomorpha

18 H3(12) H3(3) 16 COI(12) COI(3) ID#  
terminal GC%

