Genetic insights into family group co-occurrence in Cryptocercus punctulatus, a sub-social woodroach from the southern Appalachian Mountains

Ryan C Garrick Corresp. 1

1 Department of Biology, University of Mississippi, Oxford, Mississippi 38677, United States of America

Corresponding Author: Ryan C Garrick
Email address: rgarrick@olemiss.edu

The wood-feeding cockroach *Cryptocercus punctulatus* Scudder (Blattodea: Cryptocercidae) is an important member of the dead wood (saproxylic) community in montane forests of the southeastern United States. However, its population biology remains poorly understood. Here, aspects of family group co-occurrence were characterized in order to provide basic information that can be extended by studies into the evolution and maintenance of sub-sociality. Broad sampling across the species' range was coupled with molecular data [mitochondrial DNA (mtDNA) sequences]. The primary questions were: (1) what proportion of rotting logs contain two or more different mtDNA haplotypes and how often can this be attributed to multiple families inhabiting the same log, (2) are multi-family logs spatially clustered, and (3) what levels of genetic differentiation among haplotypes exist within a log, and how genetically similar are matrilines of co-occurring family groups? Multi-family logs were identified on the premise that three different mtDNA haplotypes, or two different haplotypes among adult females, is inconsistent with a single family group founded by one male-female pair. Results showed that of the 88 rotting logs from which multiple adult *C. punctulatus* were sampled, 41 logs (47%) contained two or more mtDNA haplotypes, and at least19 of these logs (22% overall) were inferred to be inhabited by multiple families. There was no strong evidence for spatial clustering of the latter class of logs. The frequency distribution of nucleotide differences between co-occurring haplotypes was strongly right-skewed, such that most haplotypes were only one or two mutations apart, but more substantial divergences (up to 18 mutations, or 1.6% uncorrected sequence divergence) do occasionally occur within logs. This work represents the first explicit investigation of family group co-occurrence in *C. punctulatus*, providing a valuable baseline for follow-up studies.

1 **Genetic insights into family group co-occurrence in *Cryptocercus punctulatus*,**

2 **a sub-social woodroach from the southern Appalachian Mountains**

4 Ryan C. Garrick

6 Department of Biology, University of Mississippi, Oxford, MS 38677, USA

8 Correspondence: Ryan C. Garrick, Department of Biology, University of Mississippi, Oxford,

9 MS 38677, USA E-mail: rgarrick@olemiss.edu; Tel.: +1-662-915-2341; Fax: +1-662-915-5144

11 Running title: Woodroach family group co-occurrence

13 Subjects: Entomology; Genetics; Zoology

14   **Abstract.** The wood-feeding cockroach *Cryptocercus punctulatus* Scudder (Blattodea:

15   Cryptocercidae) is an important member of the dead wood (saproxylic) community in montane

16   forests of the southeastern United States. However, its population biology remains poorly

17   understood. Here, aspects of family group co-occurrence were characterized in order to provide

18   basic information that can be extended by studies into the evolution and maintenance of sub-

19   sociality. Broad sampling across the species' range was coupled with molecular data

20   [mitochondrial DNA (mtDNA) sequences]. The primary questions were: (1) what proportion of

21   rotting logs contain two or more different mtDNA haplotypes and how often can this be

22   attributed to multiple families inhabiting the same log, (2) are multi-family logs spatially

23   clustered, and (3) what levels of genetic differentiation among haplotypes exist within a log, and

24   how genetically similar are matrilines of co-occurring family groups? Multi-family logs were

25   identified on the premise that three different mtDNA haplotypes, or two different haplotypes

26   among adult females, is inconsistent with a single family group founded by one male-female

27   pair. Results showed that of the 88 rotting logs from which multiple adult *C. punctulatus* were

28   sampled, 41 logs (47%) contained two or more mtDNA haplotypes, and at least19 of these logs

29   (22% overall) were inferred to be inhabited by multiple families. There was no strong evidence

30   for spatial clustering of the latter class of logs. The frequency distribution of nucleotide

31   differences between co-occurring haplotypes was strongly right-skewed, such that most

32   haplotypes were only one or two mutations apart, but more substantial divergences (up to 18

33   mutations, or 1.6% uncorrected sequence divergence) do occasionally occur within logs. This

34   work represents the first explicit investigation of family group co-occurrence in *C. punctulatus*,

35   providing a valuable baseline for follow-up studies.

36

Introduction

In sub-social invertebrates, offspring stay with parents for extended periods of time, but usually disperse before reproducing themselves (Yip & Rayor 2014). Numerous studies—particularly those focusing on spiders—have sought to understand how this form of social organization impacts genetic structure within species, and how cooperative group living evolves (e.g., Johannesen *et al*. 1998; Johannesen & Lubin 1999; Duncan *et al*. 2010; Yip *et al*. 2012). The deepest insights into costs and benefits associated with transitions from sub-sociality to eusociality have been gained by studying closely related lineages that represent different stages along this gradient (Bilde *et al*. 2005; Helantera *et al*. 2013). Accordingly, identifying sets of taxa that are suitable for comparative analyses, and characterizing their basic population biology, is of considerable value.

Sub-social *Cryptocercus* woodroaches are the closest living relatives of extant termites (Lo *et al*. 2000). As a consequence of the phylogenetic position of *Cryptocercus* within Blattodea, members of this genus are key evolutionary links for understanding transitions to eusociality (Klass *et al*. 2008; Nalepa 2015). The best studied members of *Cryptocercus* are the southern Appalachian Mountain lineages (i.e., the *C. punctulatus* complex) from the southeastern United States (Bell *et al*. 2007 and references therein). Yet, owing to their cryptic log-dwelling (saproxylic) nature, little is known about family group formation and co-occurrence in *C. punctulatus*, as direct observation is not possible in most cases. To extend our understanding of the ecological and microevolutionary processes that affect genetic structure and local persistence of sub-social invertebrates, these basic knowledge gaps need to be filled.

60

61   Previous work has revealed that *C. punctulatus* form mate pairs upon reaching maturity, and

62   produce their first—and usually only—clutch of offspring approximately one year later. In

63   addition to excavating galleries and maintaining the nest, parents provide intensive brood care.

64   This includes feeding young instars on hindgut fluids via proctodeal trophallaxis, with care

65   continuing after nymphs become nutritionally independent until the parents die, typically at least

66   three years after the birth of their young (Nalepa 1984, 2015). At this point, offspring are no

67   longer fragile; they are at least half grown and have a robust cuticle (Nalepa & Grayson 2011).

68   Given that copulation between adults likely takes place within the same gallery of the rotting log

69   in which the pair later raise their family, mate pairs have been considered monogamous since

70   opportunities for extra-pair copulation are few (but see Nalepa & Grayson 2011). Although the

71   spatial demarcation of family groups is difficult, sampling strategies employed in several

72   molecular studies of *C. punctulatus* have used a single random sample (e.g., Steinmiller *et al.*

73   2001; Aldrich *et al.* 2004a,b). Depending on the question at hand, this may be adequate.

74   However, it is worth noting that large logs typically have numerous gallery systems (Nalepa

75   1984) and so it is possible that multiple genetically divergent family groups do co-occur within a

76   single log. But it is as-yet unknown if this occurs, how often, and from where multiple family

77   groups tend to originate.

78

79   Recent work on *C. punctulatus* at Mountain Lake Biological Station in West Virginia has shed

80   new light on dispersal and colonization processes, and mate-pair composition. A pitfall trapping

81   study by Nalepa and Grayson (2011) confirmed that large nymphs, sub-adults, and adults do

82   occasionally move between logs. Population genetic studies have repeatedly shown that dispersal

83 distances of wingless saproxylic invertebrates are often very short (e.g., Sunnucks *et al*. 2006;

84 Garrick *et al*. 2007, 2008; Leschen *et al*. 2008; Marske *et al*. 2009; Walker *et al*. 2009; Bull *et al*.

85 2013), and this is also likely to be true of *C. punctulatus* (Nalepa *et al.* 2002). Accordingly,

86 following colonization of an uninhabited log by a male or female woodroach, potential mates

87 probably arrive only from logs within close proximity. This suggests that *C. punctulatus* mate-

88 pairs may be very close relatives. However, genetic estimates of relatedness between mate-pairs

89 from 36 different logs at Mountain Lake showed that this is rarely the case (Yaguchi *et al.* in

90 press). As a group, studies at this particular Biological Station (Nalepa 1984; Nalepa & Grayson

91 2011; Yaguchi *et al.* in press) have provided critical baseline data on the population biology *C.*

92 *punctulatus*, but information about family group co-occurrence, and how this might vary across

93 the species' range, is still lacking.

94

95 DNA sequence data from maternally-inherited mitochondrial DNA (mtDNA) have enabled

96 inferences about the number and composition of family groups in diverse social animal species

97 (Möller *et al.* 2001; Faulkes *et al.* 2003), including insects (Vargo 2003; Holzer *et al.* 2009). The

98 approach is most powerful when used in conjunction with information on the sex of sampled

99 individuals. In the present study, range-wide geographic sampling of *C. punctulatus* was coupled

100 with sexing and mtDNA sequencing of multiple adult cockroaches per log to address the

101 following questions: (1) what proportion of rotting logs contain two or more different mtDNA

102 haplotypes and how often can this be attributed to multiple families inhabiting the same log, (2)

103 are multi-family logs spatially clustered, and (3) what levels of genetic differentiation among

104 haplotypes exist within a log, and how genetically similar are matrilines of co-occurring family

105   groups? Although these questions focus on basic characteristics of population biology, this work

106   represents the first study to explicitly investigate family group co-occurrence in *C. punctulatus*.

107

108   **Materials and Methods**

109

110   **Taxonomy**

111

112   The taxonomic status of southern Appalachian lineages of *Cryptocercus* is contentious. Briefly,

113   following Kambhampati *et al.*'s (1996) discovery of four chromosomal races within the group,

114   Burnside *et al.* (1999) described and named each of them as separate species. However, since no

115   reliable morphological differences were apparent, the only diagnostic characters presented by the

116   authors were mtDNA nucleotides (originally seven species-specific mutations, but subsequently

117   reduced to four by Steinmiller *et al.* 2001), and the descriptions were based on few reliably

118   classified individuals (i.e., those for which both karyotype and mtDNA sequence were

119   determined). Furthermore, species diagnosis on the basis of mtDNA has been applied

120   inconsistently (e.g., Aldrich *et al.* 2004a), and the geographic origin of some type material is

121   unclear (Nalepa *et al.* 2002). Since these issues remain unresolved, the original taxon name, *C.*

122   *punctulatus*, is used here.

123

124   **Sampling and rotting log classification**

125

126   Adult cockroaches ($n = 245$) were sampled from 88 rotting logs spanning the southern

127   Appalachian Mountains and surrounding areas (Appendix 1), under scientific collecting permits

128   issued by the Alabama DCNR, Georgia DNR (29-WBH-12-16), USDA Forest Service, and US

129   National Park Service (GRSM-2012-SCI-2242; SHEN-2012-SCI-0015). Sampling involved

130   breaking open the hard outer shell of each log with a hatchet and then carefully dismantling the

131   interior woody material, which usually exhibited advanced-stage brown rot decay, with a small

132   pry bar. Although it was not feasible to follow individual galleries, cockroaches were collected

133   from the same general location within the log, until at least three adults had been caught. In most

134   cases, broods were present with the sampled adults, but were usually not collected. Specimens

135   will be lodged in the University of Mississippi Insect Collection (UMIC) following completion

136   of an on-going project in which they are being used. Two or three adults per log (mean = 2.78)

137   were used for molecular analyses. DNA extraction and polymerase chain reaction amplification,

138   sequencing, alignment and validation of data from mtDNA *cytochrome oxidase subunit I* (COI)

139   and *subunit II* (COII) genes followed Garrick (2016), and characteristics of the molecular dataset

140   are summarized in Table 1. For each individual, COI+COII sequences were concatenated (1125-

141   bp), and each rotting log was classified as containing cockroaches with the same mtDNA

142   haplotype *vs*. two or more different mtDNA haplotypes (i.e., single-haplotype *vs*. multi-

143   haplotype logs). For the latter group, the sex of each cockroach was determined based on

144   presence (♂) *vs*. absence (♀) of styli on the ventral surface of the 9th abdominal segment

145   (subgenital plate) via examination under 10× magnification. The null hypothesis of a 1:1 sex

146   ratio was then assessed using a χ^2 test.

147

148   The first goal of this study was to determine what proportion of rotting logs contain two or more

149   different mtDNA haplotypes, and how often this can be attributed to multiple families inhabiting

150   the same log. This information would provide new insights into the basic population biology of

151   *C. punctulatus*. To achieve this goal, individual-based information on mtDNA haplotype and sex

152   (Appendix 2) was used to distinguish multi-haplotype logs that contained two or more different

153   family groups (i.e., multi-family logs) from those that were consistent with expectations for only

154   a single family (i.e., other multi-haplotype logs). The inference framework was based on the

155   premise that mtDNA is strictly maternally inherited, and non-recombining. Furthermore, it was

156   assumed that *de novo* mutations are sufficiently rare that their probability of occurrence within

157   the 1125-bp region sequenced in this study was zero. However, it was not assumed that all

158   sampled adults from a given log were from the same age cohort. On this basis, a multi-family log

159   was defined as any log that contained three different haplotypes, or where two females each had

160   a different haplotype (Table 2). All other situations can be attributed to the existence of only a

161   single family derived from a monogamous pair of adults (e.g., via sampling a combination of

162   mother, father, daughters and/or sons), and thus were designated as other multi-haplotype logs.

163   These three log classes (Table 2) were the basis of subsequent analyses.

164

165   **Spatial clustering**

166

167   The second goal of this study was to explore whether multi-family logs show a non-random

168   spatial distribution; if so, this might indicate that particular environmental conditions facilitate

169   (or inhibit) family group co-occurrence. For this purpose, Cuzick and Edwards' (1990) test was

170   used to assess the null hypothesis of a random geographic distribution of multi-family logs, as

171   this test can detect global spatial clusters in individual-level data. By treating multi-family logs

172   as cases and assessing clustering relative to controls (i.e., single haplotype logs, other multi-

173   haplotype logs, or both classes combined), this method has the advantage of using only those

174   geo-spatial coordinates that were actually sampled. The procedure involved calculating the test

175   statistic Tk (i.e., the number of cases that neighbor other cases, where k is the number of nearest

176   neighbors to consider) from the empirical data, and then comparing Tk to a null distribution

177   generated via Monte Carlo randomizations of case-control labels for each of the spatial locations

178   (9,999 iterations), with significance assessed at the upper tail. Given that the most appropriate

179   neighborhood size is not known *a priori*, iterations of Cuzick and Edwards' (1990) test was run

180   for values of $k = 1$ to 5, and a Bonferroni correction was used to account for multiple testing. The

181   test was performed for the following data partitions: multi-family logs ($n = 19$ cases) *vs.* other

182   multi-haplotype logs ($n = 22$ controls), single haplotype logs ($n = 47$ controls), or both classes

183   combined ($n = 69$ controls). All spatial clustering analyses were implemented in

184   CLUSTERSEER v2.5.2 (BioMedware, USA). To examine the influence of topography on family

185   group co-occurrence, the null hypothesis of no difference in mean elevation between multi-

186   family logs *vs.* other multi-haplotype logs, single haplotype logs, or both classes combined, was

187   assessed via two-tailed *t*-tests assuming equal variances (as determined using *F*-tests).

188

189   **Genetic differentiation among haplotypes**

190

191   The third goal of this study was to quantify levels of genetic differentiation among different

192   family group matrilines that co-occur within a log, as this would provide new insights into how

193   sub-socialilty could impact genetic structure within *C. punctulatus*. By extension, outcomes of

194   this component could point towards potential benefits (or costs) of multi-family living. To

195   quantify levels of genetic divergence among co-occurring haplotypes, MEGA v6.06 (Tamura *et*

196   *al.* 2013) was used to calculate the number of nucleotide differences between individuals within

197   each multi-family and other multi-haplotype log, and these values were then plotted as a

198   frequency distribution. Within logs, redundant haplotypes were omitted, so that only non-zero

199   nucleotide differences were tallied. In the case of multi-family logs, comparisons that could be

200   clearly identified as representing between-family comparisons (e.g., when two co-occurring

201   females each had a different mtDNA haplotype) were partitioned from those that could not (e.g.,

202   when two males and a female each had a different haplotype). To assess whether multi-family *vs*.

203   other multi-haplotype logs exhibited similar distributions of pairwise nucleotide differences, they

204   were compared via a paired two-sample *t*-test.

205

206   **Results**

207

208   **Sampling and rotting log classification**

209

210   A total of 155 unique mtDNA haplotypes were identified. Of the 88 rotting logs from which *C.*

211   *punctulatus* were sampled, 47 (53%) contained only a single haplotype, whereas 41 logs had two

212   or more different haplotypes. Of the latter group, 19 logs unambiguously contained multiple

213   families (i.e., overall 22% were multi-family logs), and 22 logs (25% overall) were classified as

214   other multi-haplotype logs (Figure 1, Appendix 1). There was a significant female bias in multi-

215   family logs (71% females, $n = 56$ observations, $\chi^2 = 10.286$, *d.f.* $= 1$, $P = 0.001$). Conversely,

216   there was no meaningful departure from a 1:1 sex ratio in other multi-haplotype logs (54%

217   females, $n = 116$ observations, $\chi^2 = 0.862$, *d.f.* $= 1$, $P = 0.353$).

218

219   **Spatial clustering**

220

221   Cuzick and Edwards' (1990) tests showed that there was no clearly detectable departure from the

222   null hypothesis that multi-family logs have a random geographic distribution when the control

223   group was represented by other multi-haplotype logs, or single haplotype logs. Conversely, when

224   single haplotype and other multi-haplotype logs were combined to form the control group, the

225   test statistic was significant after Bonferroni correction ($P = 0.045$). However, only one iteration

226   of the test generated this result (i.e., when a neighborhood size of $k = 2$ was assumed). All other

227   neighborhood sizes that were examined (i.e., $k = 1, 3, 4$ and 5) yielded non-significant test

228   statistic values. There was no evidence for elevational partitioning multi-family *vs*. single

229   haplotype logs, other multi-haplotype logs, or both classes combined ($t = 0.031$, *d.f.* $= 64$, $P =$

230   0.975; $t = -1.042$, *d.f.* $= 39$, $P = 0.304$; and $t = -0.430$, *d.f.* $= 86$, $P = 0.668$, respectively).

231

232   **Genetic differentiation among haplotypes**

233

234   The sequence alignment of 155 unique mtDNA haplotypes contained 308 polymorphic sites,

235   with a maximum 116 nucleotide differences between a pair of haplotypes (i.e., 10.3%

236   uncorrected sequence divergence). Of the 41 logs with multiple haplotypes, 38 logs contained

237   two, whereas three logs had three different haplotypes. Accordingly, the frequency distribution

238   of nucleotide differences between haplotypes sampled from within the same rotting log was

239   calculated from a total of 47 pairwise comparisons. This frequency distribution showed that

240   while co-occurring haplotypes most often (66% of the time) differ from one another by only one

241   or two mutations, this is not always the case—modest differences within logs (up to 18

242   mutations, or 1.6 % uncorrected sequence divergence) do also occur (Figure 2). Within multi-

243 family logs, data points that were not clearly attributable to between-family comparisons were

244 relatively few (20%; Figure 2). Furthermore, there was no strong discord in the distribution of

245 pairwise nucleotide differences between multi-family *vs*. other multi-haplotype logs ($t = 0.410$,

246 *d.f.* = 17, $P = 0.343$).

247

248   **Discussion**

249

250 This paper represents the first explicit investigation of family group co-occurrence in *C.*

251 *punctulatus*—an evolutionarily important woodroach from a montane forest biodiversity hotspot

252 (Garrick 2011). Broad geographic sampling was coupled with sequencing of multiple individuals

253 per rotting in order to address the following questions: (1) what proportion of rotting logs contain

254 two or more different mtDNA haplotypes and how often can this be attributed to multiple

255 families inhabiting the same log, (2) are multi-family logs spatially clustered, and (3) what levels

256 of genetic differentiation among haplotypes exist within a log, and how genetically similar are

257 matrilines of co-occurring family groups? The first question provides insights into an as-yet

258 unknown aspect of the basic population biology of *C. punctulatus*, whereas the second question

259 is exploratory, with the potential to indicate whether environmental factors might influence

260 family group co-occurrence. The third question yields information into how sub-socialilty could

261 impact genetic structure within this species. Together, they provide a framework for subsequent

262 studies. Below, major findings are summarized, and limitations of the present work and

263 recommendations for future studies are also highlighted.

264

265 **Family group co-occurrence**

266

267   Approximately half (47%) of the rotting logs from which *C. punctulatus* were sampled contained

268   multiple mtDNA haplotypes (Figure 1, Appendix 1). In contrast, Kambhampati *et al.* (1996)

269   sequenced portions of two mtDNA genes (12S and 16S rRNA; 440-bp; and 406- to 414-bp,

270   respectively) for two *C. punctulatus* individuals from each of five sites across Virginia, North

271   Carolina, Georgia and Alabama, but found no within-site genetic differences. However, given

272   that the latter study was limited by small sample sizes, it does not provide a strong comparison

273   here. Similarly, whereas the present study found that almost half of the confirmed multi-

274   haplotype logs (22% overall) clearly contained representatives of two or more family groups

275   (Figure 1, Appendix 1), other molecular studies have routinely pooled *C. punctulatus* individuals

276   sampled from different logs (e.g., Burnside *et al.* 1999; Hossain & Kambhampati 2001) and so

277   they too provide no useable comparative data.

278

279   One notable exception is a study by Aldrich *et al.* (2005), who used allozyme markers to screen

280   genetic variation in ~40 *C. punctulatus* nymphs from each of 23 sites across the southern

281   Appalachian Mountains. Although samples were often pooled across logs, for 10 sites they were

282   not, and so in these particular cases a single rotting log represented the basic unit of analysis. For

283   this subset of logs, the authors reported that genotype frequencies deviated significantly from

284   Hardy-Weinberg expectations. In nine out of 10 logs the deviation was in the direction of

285   homozygote excess—an outcome that is consistent with Wahlund effect (i.e., genetic

286   substructure, potentially caused by sampling different family groups). However, homozygote

287   excess is also compatible with inbreeding. Notably, recent work by Yaguchi *et al.* (in press)

288   suggested that mating among close relatives is relatively uncommon in *C. punctulatus*. Those

289   authors found that in a sample of 36 mate-pairs genotyped with a set of nuclear microsatellite

290   loci, 72% were unrelated (the remaining 28% had parent-offspring, full-sib, or half-sib

291   relationships). Given this, Aldrich *et al.*'s (2005) data can be explained by frequent sampling of

292   two or more family groups per log, in agreement with conclusions of the present study.

293

294   The finding that multi-family logs are quite common in *C. punctulatus* is interesting, as there are

295   reasons why single family logs could be advantageous. For example, inhabitants of single family

296   logs would be released from intraspecific competition, including energetically expensive defense

297   of galleries against intruders (Nalepa 2015). Also, high relatedness among all individuals in a

298   single family log could potentially enhance cooperative behaviors (e.g., dislodgment and

299   reduction of dead wood into small particles that can be ingested; Watanabe & Tokuda 2010),

300   since benefits are received exclusively by kin. However, despite the potential costs of sharing a

301   log with other families, in *C. punctulatus* this appears to be relatively common.

302

303   **Spatial distribution of multi-family logs**

304

305   Rotting logs that contained representatives of multiple family groups were randomly arrayed

306   across elevational strata, and also across the geographic area that was sampled in this study.

307   Although one of several iterations of Cuzick and Edwards' (1990) test for spatial clustering did

308   indicate that some structure may exist, this outcome was confined to a narrow portion of the

309   parameter space, and so it is generally poorly supported. Overall, the analyses presented here

310   suggest that if there are habitat characteristics that promote (or limit) family group co-

311   occurrence, they probably operate over finer scales than those examined in this study.

312   Previously, it has been suggested that large logs may have greater potential to harbor multiple

313   mate-pairs of *C. punctulatus* than small logs (Yaguchi *et al.* in press). Log diameter, length, and

314   other microhabitat characteristics that affect saproxylic invertebrates such as decomposition class

315   and moisture content can certainly vary over fine geographic scales in montane areas (e.g.,

316   Barclay *et al.* 2000; Woodman *et al.* 2006). Accordingly, an examination of factors that

317   influence family group co-occurrence in *C. punctulatus* should incorporate local environmental

318   variables. As these data are generally not available at the appropriate resolution from remote

319   sensing and other GIS databases, techniques for measuring very fine-scale ecological data would

320   need to be employed (e.g., Barclay *et al.* 2000; Grove 2002).

321

322   **Genetic differentiation**

323

324   The frequency distribution of nucleotide differences between unique haplotypes sampled from

325   the same rotting log was strongly right-skewed, with most co-occurring haplotypes only one or

326   two mutations apart, but with some occurrences of more substantial divergences (up to 18

327   mutations, or 1.6% uncorrected sequence divergence; Figure 2). Also, there was no strong

328   difference between frequency distributions derived from logs that clearly contained multiple

329   families and those that did not (i.e., multi-family *vs*. other multi-haplotype logs). This suggests

330   that most co-occurring families share a recent common ancestor. Furthermore, for multi-

331   haplotype logs that potentially contained only a single family (i.e., where a single mate-pair was

332   sampled), the right-skewed frequency distribution indicates that, based on mtDNA data, parental

333   individuals are not all that distantly related. Although Yaguchi *et al.* (in press) showed that 72%

334   of mate pairs were unrelated (i.e., they shared no more microsatellite alleles per locus than was

335   expected by chance, given the local frequency of each allele), nuclear genotypic data are most

336   informative over relatively short generation-to-generation timescales. Conversely, DNA

337   sequence data can provide insights over deeper time scales (Sunnucks 2000; Garrick *et al.* 2006,

338   2010, 2015). For instance, mate-pairs that are second cousins would probably not register as

339   being related on the basis of a small set of microsatellite loci due to several generations of

340   gametic recombination, but they could nonetheless be identified as close relatives on the basis of

341   their more slowly evolving mtDNA haplotype. Thus, the notion that co-occurring families share

342   a recent common ancestor is not incompatible with Yaguchi *et al.*'s (in press) findings.

343

344   Generally speaking, high genetic diversity among individuals that co-occur at a site promotes

345   long-term persistence of the local population (Frankham 2005, and references therein). However,

346   in circumstances where antagonistic and competitive intraspecific interactions mostly occur

347   among non-siblings, high within-site relatedness may be advantageous (Caesar *et al*. 2010).

348   Baseline data from the present study indicate that compared to genetically polymorphic single

349   family logs, there is no meaningful increase in within-site mtDNA-based diversity among *C.*

350   *punctulatus* that share their rotting log with other families. Although mtDNA sequence

351   divergence may be only loosely correlated with genome-wide nuclear genetic diversity (Zhang &

352   Hewitt 2003), the empirical data nonetheless suggest that if there are benefits to living in a multi-

353   family log, increased local genetic diversity is unlikely to contribute to this.

354

355   In addition to characterizing aspects of family group co-occurrence for the purpose of

356   understanding sub-sociality, the present study also provides an opportunity to examine the

357   adequacy of randomly sampling a single *C. punctulatus* individual per rotting log (e.g.,

358   Steinmiller *et al.* 2001; Aldrich *et al.* 2004a,b). Based on levels of genetic differentiation among

359   co-occurring haplotypes seen here, the one-sample-per-log strategy would fail to fully capture

360   within-log diversity about half of the time. Even though the majority of co-occurring haplotypes

361   have few mutational differences, some divergences are more substantial (Figure 2). Indeed, *C.*

362   *punctulatus* from highly divergent genetic lineages—including those that likely differ in

363   chromosome number—may occasionally co-inhabit the same log (Garrick 2016). Accordingly,

364   depending on the goals of the study, it may be prudent to sample multiple individuals from a log.

365

366   **Limitations and future directions**

367

368   Several limitations of this study warrant consideration. First, while the overall sample size and

369   geographic coverage was large, sample sizes per log were quite small. Accordingly, it was not

370   possible to have a *confirmed absence* of multiple families (or haplotypes) within a given log.

371   This may have impacted the ability to detect spatial clustering, as Cuzick and Edwards' (1990)

372   test assumes that controls were correctly classified. This also means that the reported frequencies

373   of co-occurrence of multiple haplotypes (47% of logs) and family groups (22% of logs) represent

374   lower bounds, not point estimates. This consideration may partly reconcile the much higher

375   frequency of multi-family logs (90% of relevant logs) suggested by Aldrich *et al.*'s (2005)

376   allozyme data. However, testing of Hardy-Weinberg equilibrium is not an ideal framework for

377   making inferences about family group co-occurrence as there are several possible causes for

378   departures from expected genotype frequencies, and so Aldrich *et al.*'s (2005) data may provide

379   an overestimate. A second limitation of the present study is that the mtDNA-based inference

380   framework has a reduced ability to detect family group co-occurrence in logs from which few

381 females are sampled (e.g., a significantly female-biased sex ratio was detected within multi-

382 family logs, but not within other multi-haplotype logs; also see Table 1). This means that

383 resolution of the approach used here may not be equal across all logs. Third, the simplifying

384 assumption of no *de novo* mutations may not hold true. However, the mtDNA mutation rate

385 would need to be exceptionally high to overturn inferences of multiple families within a log, and

386 so this limitation is a source of noise (cf. positively misleading).

387

388 For understanding family group co-occurrence in *C. punctulatus*, neither a mtDNA-only

389 approach, nor a nuclear marker-only approach, is flawless. The clearest insights would be gained

390 by coupling genotypic data from multiple bi-parentally inherited nuclear markers with haplotypic

391 data from maternally inherited mtDNA, together with information on the sex (and also age

392 cohort) of each individual. Dense sampling—without pooling across logs—is also important, and

393 it would be beneficial to quantify characteristics of the log itself so that potential influences of

394 microhabitat on family group co-occurrence could be investigated. Thus, the present study

395 represents a baseline against which follow-up studies can be compared, and the current findings

396 should be considered a working hypothesis, to be refined with additional data.

397

398 **Acknowledgments**

399

400 Sequencing was performed by staff at Yale University's DNA Analysis Facility on Science Hill.

401 R.J. Worthington assisted with cockroach sex identifications, R.J. Dyer, C.Q. Hyseni and R.E.

402 Symula assisted with some fieldwork, and R.N. Yi contributed to some laboratory work. This

403   paper also benefited from constructive comments from D.P.W. Huber and two anonymous

404   reviewers.

405

406   **References**

407

408   Aldrich BT, Krafsur E, Kambhampati S (2004a) Species-specific allozyme markers for

409   Appalachian wood-feeding cockroaches (Dictyoptera: Cryptocercidae). *Biochemical Genetics*,

410   42, 149–164.

411

412   Aldrich BT, Zolnerowich G, Kambhampati S (2004b) Interspecific morphological variation in

413   the wood-feeding cockroach, *Cryptocercus* (Dictyoptera: Cryptocercidae). *Arthropod Structure*

414   *and Development*, 33, 443–451.

415

416   Aldrich BT, Kambhampati S, Krafsur ES (2005) Population genetics of wood-feeding

417   cockroaches in the genus *Cryptocercus*. *Journal of Heredity*, 96, 572–575.

418

419   Barclay S, Ash JE, Rowell DM (2000) Environmental factors influencing the presence and

420   abundance of a log-dwelling invertebrate, *Euperipatoides rowelli* (Onychophora:

421   Peripatopsidae). *Journal of the Zoological Society of London*, 250, 425–436.

422

423   Bell WJ, Roth LM, Nalepa CA (2007) *Cockroaches: Ecology, Behavior, and Natural History*.

424   Pp. 1–230. The Johns Hopkins University Press, Baltimore MD, USA.

425

426   Bilde T, Lubin Y, Smith D, Schneider JM, Maklakov AA (2005) The transition to social inbred

427   mating systems in spiders: Role of inbreeding tolerance in a subsocial predecessor. *Evolution*,

428   59, 160–174.

429

430   Bull, JK, Sands CJ, Garrick RC, Gardner MG, Tait NN, Briscoe DA, Rowell DM, Sunnucks P

431   (2013) Environmental complexity and biodiversity: The multi-layered evolutionary history of a

432   log-dwelling velvet worm in montane temperate Australia. *PLoS ONE*, 8, e84559.

433

434   Burnside CA, Smith PT, Kambhampati S (1999) Three new species of the wood roach,

435   *Cryptocercus* (Blattodea: Cryptocercidae), from the Eastern United States. *Journal of the Kansas*

436   *Entomological Society*, 72, 361–378.

437

438   Caesar S, Karlsson M, Forsman A (2010) Diversity and relatedness enhance survival in colour

439   polymorphic grasshoppers. *PLoS ONE*, 5, e10880.

440

441   Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations. *Journal of the*

442   *Royal Statistical Society Series B*, 52, 73–104.

443

444   Duncan SI, Riechert SE, Fitzpatrick BM, Fordyce JA (2010) Relatedness and genetic structure in

445   a socially polymorphic population of the spider *Anelosimus studiosus*. *Molecular Ecology*, 19,

446   810–818.

447

448   Faulkes CG, Arruda MF, Monteiro Da Cruz MAO (2003) Matrilineal genetic structure within

449   and among populations of the cooperatively breeding common marmoset, *Callithrix jacchus*.

450   *Molecular Ecology*, 12, 1101–1108.

451

452   Frankham R (2005) Genetics and extinction. *Biological Conservation*, 126, 131–140.

453

454   Garrick RC (2011) Montane refuges and topographic complexity generate and maintain

455   invertebrate biodiversity: Recurring themes across space and time. *Journal of Insect*

456   *Conservation*, 15, 469–478.

457

458   Garrick RC (2016) True syntopy between chromosomal races of the *Cryptocercus punctulatus*

459   Scudder (Blattodea: Cryptocercidae) wood-roach species complex. *Insectes Sociaux*, 3, 353–355.

460

461   Garrick RC, Caccone A, Sunnucks P (2010) Inference of population history by coupling

462   exploratory and model-driven phylogeographic analyses. *International Journal of Molecular*

463   *Sciences*, 11, 1190–1227.

464

465   Garrick RC, Kajdacsi B, Russello MA, Benavides E, Hyseni C, Gibbs JP, Tapia W, Caccone A

466   (2015) Naturally rare versus newly rare: Demographic inferences on two timescales inform

467   conservation of Galápagos giant tortoises. *Ecology and Evolution*, 5, 676–694.

468

469 Garrick RC, Rowell DM, Simmons CS, Hillis DM, Sunnucks P (2008) Fine-scale

470 phylogeographic congruence despite demographic incongruence in two low-mobility saproxylic

471 springtails. *Evolution*, 62, 1103–1118.

472

473 Grove SJ (2002) Tree basal area and dead wood as surrogate indicators of saproxylic insect

474 faunal integrity: A case study from the Australian lowland tropics. *Ecological Indicators*, 1, 171–

475 188.

476

477 Helantera H, Aehle O, Roux M, Heinze J, D'Ettorre P (2013) Family-based guilds in the ant

478 *Pachycondyla inversa*. *Biology Letters*, 9, 20130125.

479

480 Holzer B, Keller L, Chapuisat M (2009) Genetic clusters and sex-biased gene flow in a

481 unicolonial *Formica* ant. *BMC Evolutionary Biology*, 9, 69.

482

483 Hossain S, Kambhampati S (2001) Phylogeny of *Cryptocercus* species (Blattodea:

484 Cryptocercidae) inferred from nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution*,

485 21, 162–165.

486

487 Johannesen J, Baumann T, Seitz A, Veith M (1998) The significance of relatedness and gene

488 flow on population genetic structure in the subsocial spider *Eresus cinnaberinus* (Araneae:

489 Eresidae). *Biological Journal of the Linnean Society*, 63, 81–98.

490

491 Johannesen J, Lubin Y (1999) Group founding and breeding structure in the subsocial spider

492 *Stegodyphus lineatus* (Eresidae). *Heredity*, 82, 677–686.

493

494 Kambhampati S, Luykx P, Nalepa CA (1996) Evidence for sibling species in *Cryptocercus*

495 *punctulatus*, the wood roach, from variation in mitochondrial DNA and karyotype. *Heredity*, 76,

496 485–496.

497

498 Klass K-D, Nalepa C, Lo N (2008) Wood-feeding cockroaches as models for termite evolution

499 (Insecta: Dictyoptera): *Cryptocercus* vs. *Parasphaeria boleiriana*. *Molecular Phylogenetics and*

500 *Evolution*, 46, 809–817.

501

502 Leschen RAB, Buckley TR, Harman HM, Shulmeister J (2008) Determining the origin and age

503 of the Westland beech (*Nothofagus*) gap, New Zealand, using fungus beetle genetics. *Molecular*

504 *Ecology*, 17, 1256–1276.

505

506 Marske KA, Leschen RAB, Barker GM, Buckley TR (2009) Phylogeography and ecological

507 niche modelling implicate coastal refugia and trans-alpine dispersal of a New Zealand fungus

508 beetle. *Molecular Ecology*, 18, 5126–5142.

509

510 Möller LM, Beheregaray LB, Harcourt RG, Krützen M (2001) Alliance membership and kinship

511 in wild male bottlenose dolphins (*Tursiops aduncus*) of southeastern Australia. *Proceedings of*

512 *the Royal Society B: Biological Sciences*, 268, 1941–1947.

513

514   Nalepa CA (1984) Colony composition, protozoan transfer and some life history characteristics

515   of the woodroach *Cryptocercus punctulatus* Scudder (Dictyoptera: Cryptocercidae). *Behavioral*

516   *Ecology and Sociobiology*, 14, 273–279.

517

518   Nalepa CA (2015) Origin of termite eusociality: Trophallaxis integrates the social, nutritional,

519   and microbial environments. *Ecological Entomology*, 40, 323–335.

520

521   Nalepa CA, Grayson KL (2011) Surface activity of the xylophagous cockroach *Cryptocercus*

522   *punctulatus* (Dictyoptera: Cryptocercidae) based on collections from pitfall traps. *Annals of the*

523   *Entomological Society of America*, 104, 364–368.

524

525   Nalepa CA, Luykx P, Klass K-D, Deitz LL (2002) Distribution of karyotypes of the

526   *Cryptocercus punctulatus* species complex (Dictyoptera: Cryptocercidae) in the Southern

527   Appalachians: Relation to habitat and history. *Annals of the Entomological Society of America*,

528   95, 276–287.

529

530   Steinmiller B, Kambhampati S, Brock BL (2001) Geographic distribution of, and genetic

531   variation in, the wood roach *Cryptocercus* (Dictyoptera: Cryptocercidae) in the United States.

532   *Annals of the Entomological Society of America*, 94, 732–742.

533

534   Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and*

535   *Evolution*, 15, 199–203.

536

537   Sunnucks P, Blacket MJ, Taylor JM, Sands CJ, Ciavaglia SA, Garrick RC, Tait NN, Rowell DM,

538   Pavlova A (2006) A tale of two flatties: Different responses of two terrestrial flatworms to past

539   environmental climatic fluctuations at Tallaganda in montane southeastern Australia. *Molecular*

540   *Ecology*, 15, 4513–4531.

541

542   Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary

543   Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.

544

545   Vargo EL (2003) Hierarchical analysis of colony and population genetic structure of the eastern

546   subterranean termite, *Reticulitermes flavipes*, using two classes of molecular markers. *Evolution*,

547   57, 2805–2818.

548

549   Walker MJ, Stockman AK, Marek PE, Bond JE (2009) Pleistocene glacial refugia across the

550   Appalachian Mountains and coastal plain in the millipede genus *Narceus*: Evidence from

551   population genetic, phylogeographic, and paleoclimatic data. *BMC Evolutionary Biology*, 9, 25.

552

553   Watanabe H, Tokuda G (2010) Cellulolytic systems in insects. *Annual Review of Entomology*,

554   55, 609–632.

555

556   Woodman JD, Ash JE, Rowell DM (2006) Population structure in a saproxylic funnelweb spider

557   (Hexathelidae: *Hadronyche*) along a forested rainfall gradient. *Journal of Zoology*, 268, 325–

558   333.

559

560   Yaguchi H, Hayashi Y, Tohoku T, Nalepa C, Maekawa K (in press) Genetic data indicate that

561   most field-collected woodroach pairs are unrelated. *Insect Science*.

562

563   Yip EC, Rayor LS (2014) Maternal care and subsocial behaviour in spiders. *Biological Reviews*,

564   89, 427–449.

565

566   Yip EC, Rowell DM, Rayor LS (2012) Behavioural and molecular evidence for selective

567   immigration and group regulation in the social huntsman spider, *Delena cancerides*. *Biological*

568   *Journal of the Linnean Society*, 106, 749–762.

569

570   Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations:

571   Practice, problems and prospects. *Molecular Ecology*, 12, 563–584.

Figure 1

Map showing spatial distributions of classified rotting logs

Figure 1. Map of the southeastern USA showing spatial distributions of rotting logs classified as single haplotype (green circle), multi-family (red star), or other multi-haplotype (blue diamond) logs, based on mtDNA sequence data coupled with information on the sex of *C. punctulatus* individuals. State abbreviations are: Alabama, AL; Georgia, GA, Kentucky, KY; North Carolina, NC; South Carolina, SC: Tennessee, TN; Virginia, VA; and West Virginia, WV.

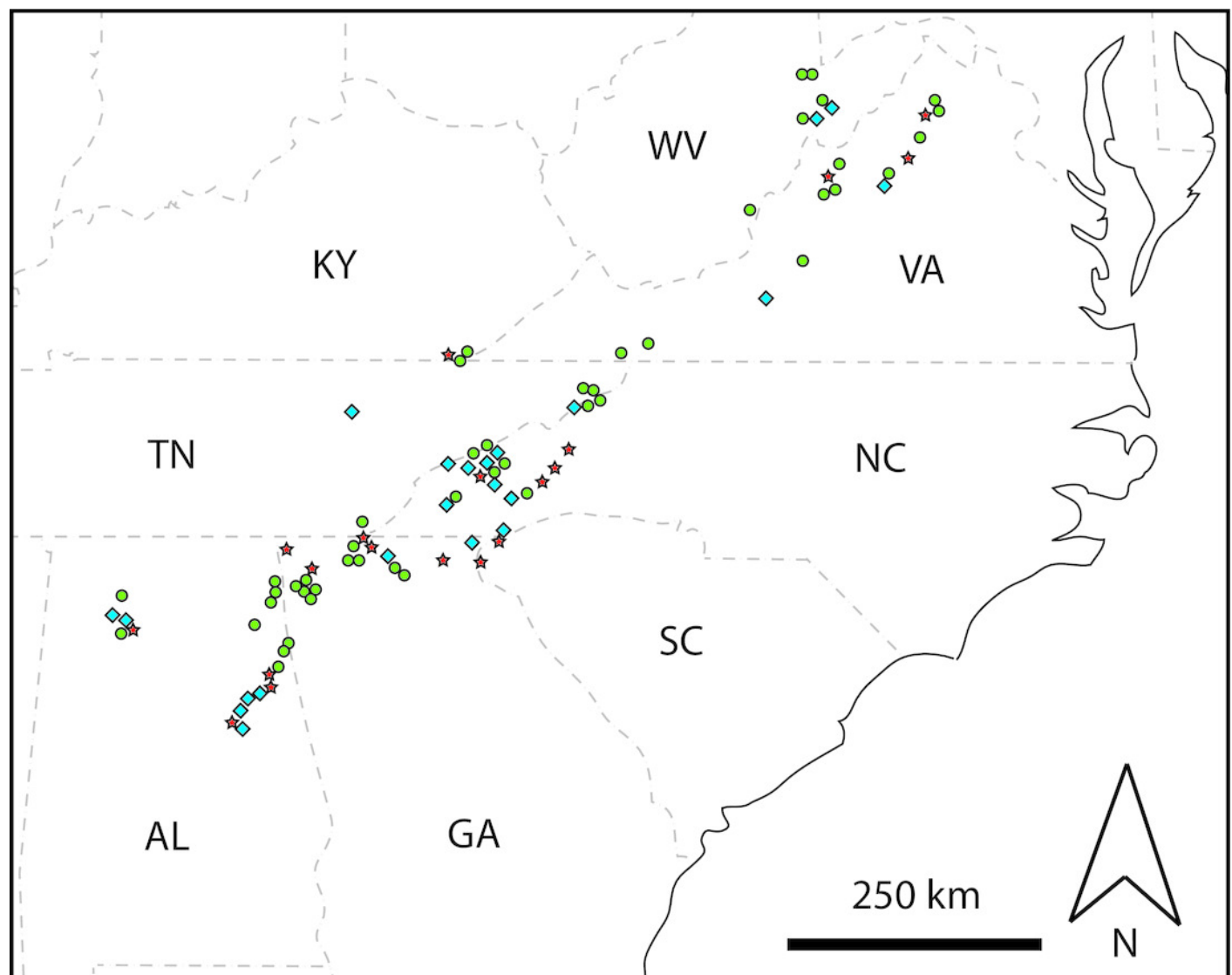



Figure 2

Frequency distribution of the number of nucleotide differences between pairs of non-redundant mtDNA haplotypes from *C. punctulatus* individuals sampled from the same rotting log.

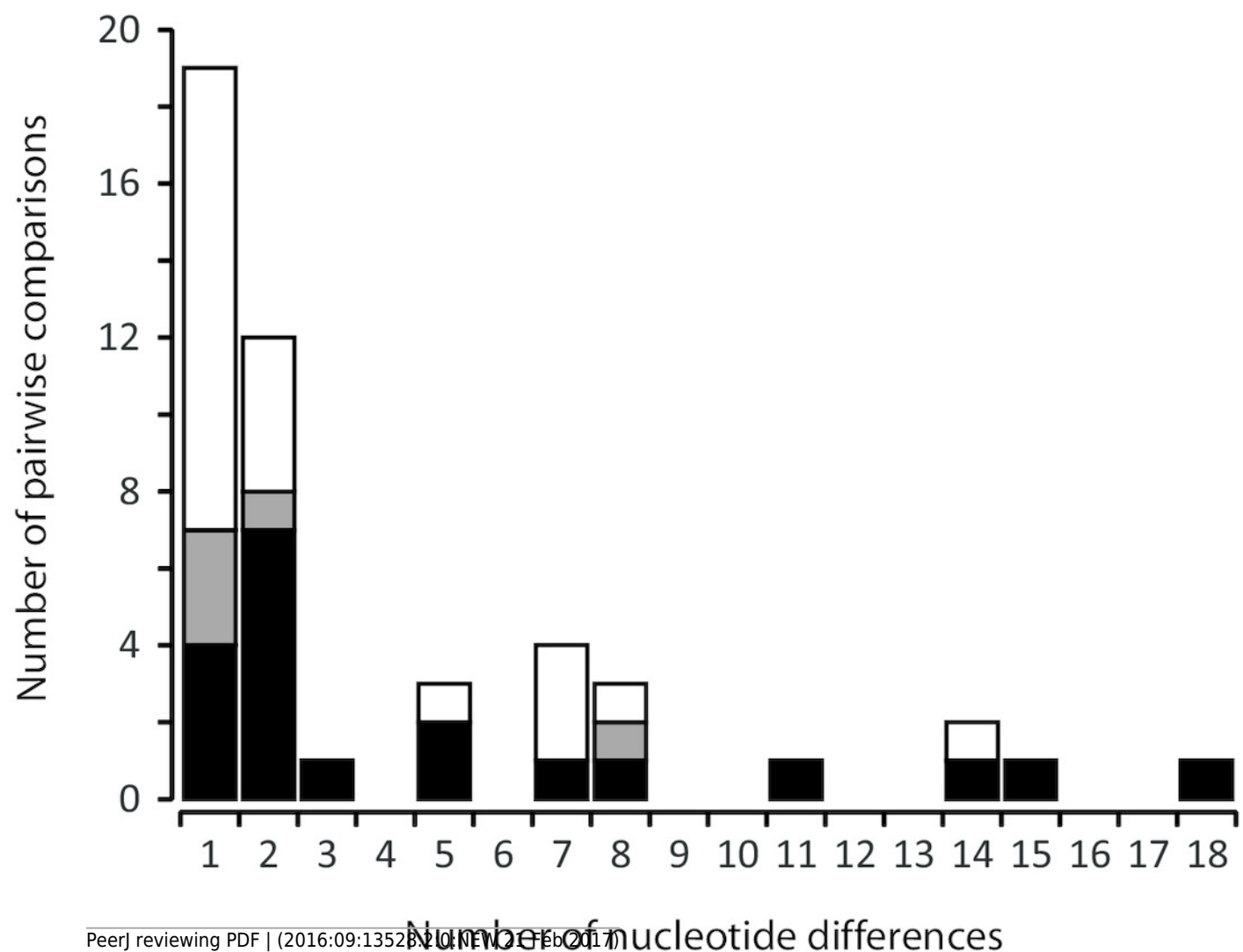
Figure 2. Frequency distribution of the number of nucleotide differences between pairs of non-redundant mtDNA haplotypes from *C. punctulatus* individuals sampled from the same rotting log. Shading of bars represents the proportional contribution of multi-family (black plus grey) vs. other multi-haplotype (white) logs to the overall tally. The bi-partitioning of multi-family logs delineates comparisons that could (black squares) vs. could not (pale grey) be clearly identified as representing between-family comparisons.

Characteristics of mtDNA sequence data

Table 1. Characteristics of mtDNA sequence data generated from 245 *C. punctulatus* individuals sampled in the present study. Summary statistics are presented for each gene separately, and also for the concatenated dataset. Abbreviations are as follows: number of base pairs, bp; proportion of guanine plus cytosine nucleotides, G+C%; nucleotide composition maximum-likelihood estimate (GTR+I+G model) of transition / transversion ratio, ts/tv; number of segregating sites, S; and number of unique haplotypes, N_{hap}.

mtDNA gene region	Alignment length (bp)	G+C%	ts/tv	S	N_{hap}	GenBank accessions
COI	762	34.4	4.2	199	142	KX944872 - KX945114; and KU609620 - KU609621
COII	363	25.7	6.7	109	64	KX945115 - KX945357; and KU609623 - KU609624
Concatenated	1125	33.8	4.9	308	155	

Hypothetical combinations of mtDNA haplotype × individual sex, for a rotting log from which three adults were sampled

Table 2. All possible hypothetical combinations of mtDNA haplotype × individual sex, for a rotting log from which three adults were sampled. For each row, different typeface (italics, underlined, or normal) represents different mtDNA haplotypes among individuals within a log, M indicates a male (i.e., father or son), and F indicates a female (i.e., mother or daughter). Multi-haplotype logs that unambiguously contain multiple families are labeled MF, whereas those which are consistent with the existence of only a single family are labeled OMH. Logs that contain a single mtDNA haplotype are designated SH.

No. of different mtDNA haplotypes	Individual 1	Individual 2	Individual 3	Rotting log classification
1	M	M	M	SH
2	M	M	M	OMH
3	M	M	*M*	MF
1	F	F	F	SH
2	F	F	F	MF
3	F	F	*F*	MF
2	M	F	F	OMH
2	M	F	F	MF
1	M	F	F	SH
3	M	*F*	F	MF
2	M	M	F	OMH
2	M	M	F	MF
1	M	M	F	SH
3	M	*M*	F	MF