Phylogeographic structure in three North American tent caterpillar species (Lepidoptera: Lasiocampidae): *Malacosoma americana*, *M. californica*, and *M. disstria*

Linda A Lait [Corresp., 1] , Paul DN Hebert [1]

[1] Centre for Biodiversity Genomics, University of Guelph, Guelph, Ontario, Canada

Corresponding Author: Linda A Lait
Email address: llait@uoguelph.ca

While phylogeographic structure has been examined in many North American vertebrate species, insects have received much less attention despite their central ecological roles. The moth genus *Malacosoma* (Hübner, 1820), is an important group of forestry pests responsible for large-scale defoliation across much of the Nearctic and Palearctic. The present study uses sequence variation in the mitochondrial cytochrome *c* oxidase 1 (COI) gene to examine the population genetic structure of the three widespread *Malacosoma* species (*M. americana*, *M. californica*, and *M. disstria*). Populations of all three species showed highest diversity in the south, suggesting that modern populations derived from southern refugia with loss of variation as these lineages dispersed northwards. However, despite similar life histories and dispersal abilities, the extent of regional variation varied among the taxa. *M. americana*, a species restricted to eastern North America, showed much less genetic structure than the western *M. californica* or the widespread *M. disstria*. The regional differentiation in the latter reflects the likely derivation of modern lineages from several refugia, as well as taxonomic uncertainty in *M. californica*. In these respects, the three species of *Malacosoma* share phylogeographic patterns similar to those detected in vertebrates which are characterised by greater phylogeographic breaks in the western half of the continent and limited structure in the east.

1 **ARTICLE TYPE:** Original article

2 **Phylogeographic structure in three North American tent caterpillar species (Lepidoptera:**

3 **Lasiocampidae):** *Malacosoma americana*, *M. californica*, **and** *M. disstria*

4 **Linda A Lait[1] and Paul DN Hebert[1]**

5 *[1] Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road E, Guelph, Ontario,*

6 *N1G 2W1, Canada*

7 **Correspondence:** Linda A Lait, Centre for Biodiversity Genomics, llait@uoguelph.ca.

8 **Running title:** Phylogeography of three *Malacosoma*

9   **ABSTRACT**

10   While phylogeographic structure has been examined in many North American vertebrate

11 species, insects have received much less attention despite their central ecological roles. The moth

12 genus *Malacosoma* (Hübner, 1820), is an important group of forestry pests responsible for large-

13 scale defoliation across much of the Nearctic and Palearctic. The present study uses sequence

14 variation in the mitochondrial cytochrome *c* oxidase 1 (COI) gene to examine the population

15 genetic structure of the three widespread *Malacosoma* species (*M. americana*, *M. californica*, and

16 *M. disstria*). Populations of all three species showed highest diversity in the south, suggesting

17 that modern populations derived from southern refugia with loss of variation as these lineages

18 dispersed northwards. However, despite similar life histories and dispersal abilities, the extent of

19 regional variation varied among the taxa. *M. americana,* a species restricted to eastern North

20 America, showed much less genetic structure than the western *M. californica* or the widespread

21 *M. disstria*. The regional differentiation in the latter reflects the likely derivation of modern

22 lineages from several refugia, as well as taxonomic uncertainty in *M. californica*. In these

23 respects, the three species of *Malacosoma* share phylogeographic patterns similar to those

24 detected in vertebrates which are characterised by greater phylogeographic breaks in the western

25 half of the continent and limited structure in the east.

INTRODUCTION

27   The patterns of genetic variation in species and the processes which underlie them are of

28   particular interest to evolutionary biologists. Diverse factors, both historical and contemporary,

29   influence how variation is distributed among populations; these include geological and climatic

30   events, and the presence of physical and behavioural barriers (Avise, 2004). Past glaciations have

31   had a major impact on the extent and patterning of genetic structure in Northern Hemisphere

32   species (Hewitt, 2000). In North America, ice sheets covered much of present-day Canada and

33   the northern United States, temperatures in ice-free areas were cooler than today, and sea levels

34   dropped by up to 140 m (Pielou, 1991; Barendregt & Irving, 1998; Dyke *et al.*, 2002). The

35   distributions of many species were fragmented with their populations persisting in small ice-free

36   refugia (Hewitt, 1996; Hewitt, 2000; Stewart & Lister, 2001). In addition, physical and ecological

37   barriers influenced genetic structure by their impacts on dispersal and gene flow. Recent

38   environmental changes, both anthropogenic and natural, are now causing range shifts and

39   population changes with varied impacts on both inter- and intraspecific genetic variation (Walther

40   *et al.*, 2002; Chen *et al.,* 2011).

41   Comparisons of genetic variation spanning multiple species can identify both general and

42   species-specific patterns, revealing how particular life history characteristics impact population

43   structure. Species-specific traits, such as dispersal ability or niche requirements, may affect how a

44   species responds to environmental, climatic, and geological changes. For example, the majority

45   of northern species experienced major population declines reflecting the loss of habitat during the

46   Pleistocene glaciations, while the interglacials favoured range expansion and population growth

47   (Nilsson, 1983; Pielou, 1991; Hewitt, 2000; Hewitt, 2004). In contrast, cold-adapted species

48   experienced habitat loss, often retreating to high altitude and high latitude locations during

49   interglacials (Stewart & Lister, 2001; Dalén *et al.*, 2005; Galbreath *et al.*, 2009; Stewart *et al.*,

50   2009). Topographic features also have differing effects, with mountains and rivers restricting

51   gene flow in some species while acting as dispersal corridors for others. For example, the Rocky

52   Mountains prevent gene flow between some populations (Crease *et al.*, 1997; Burg *et al.*, 2005),

53   but provide habitat as "sky islands" with the intervening lowlands limiting gene flow in others

54   (Knowles, 2000; DeChaine & Martin, 2005; Galbreath *et al.*, 2009).

55   Past studies of phylogeographic structure in terrestrial organisms have largely examined

56   vertebrates. Given their high diversity and abundance, phylogeographic patterns in insects have

57   been understudied; past work has revealed diverse outcomes ranging from global panmixis

58   (Alvial *et al.*, 2007; Correa *et al.*, 2017) to highly fragmented, structured populations (Dinca,

59   Dapporto & Vila, 2011; Frantine-Silva *et al.*, 2017; Karthika *et al.*, 2017). Phylogeographic

60   studies of Lepidoptera have employed both nuclear and mitochondrial markers, particularly the

61   cytochrome *c* oxidase 1(COI) locus (Vandewoestijne *et al.*, 2004; Craft *et al.,* 2010; Kirichenko

62   *et al.,* 2017). This study represents the first step in a broad investigation of phylogeographic

63   patterns in North American Lepidoptera.

64   *Malacosoma* (Hübner, 1820) is a Holarctic genus found across much of North America,

65   Europe, and Asia, with six species native to Canada and the United States (Franclemont, 1973).

66   While males are strong fliers, females are usually sedentary until they deposit their egg mass

67   (Stehr & Cook, 1968; Franclemont, 1973), suggesting that the dispersal of maternal markers will

68   be low. As their name (tent caterpillar) suggests, they build large tents or moulting mats which

69   can accommodate a large number of larvae (Franclemont, 1973). All six species feed on diverse

70   deciduous trees and shrubs, with host preferences varying by taxon and region (Stehr & Cook,

71   1968; Franclemont, 1973; Parry & Goyer, 2004). The group contains a number of important

72   forestry pests, species that experience cyclical outbreaks which often lead to extensive forest

73   defoliation (Hildahl & Reeks, 1960; Stehr & Cook, 1968; Roland, 1993). Despite these impacts,

74   there have been few studies of population genetic structure in these moths. One study, which

75   assessed allozyme variation in *Malacosoma americana* (Fabricius, 1793) from eastern United

76   States, found limited variation and a lack of regional genetic differentiation (Costa & Ross,

77   1994). A second employed microsatellites and short DNA sequences to compare five populations

78   of *Malacosoma californica pluviale* in coastal British Columbia (Franklin, Myers & Cory, 2014).

79   Although high levels of variation were evident, there was little genetic differentiation between

80   island and mainland populations. Both studies found limited genetic differentiation at a relatively

81   small geographic scale.

82   This study examines the population genetic structure of the three widely distributed North

83   American *Malacosoma* species, the eastern tent caterpillar *M. americana*, the western tent

84   caterpillar *M. californica* (Packard, 1864), and the forest tent caterpillar *M. disstria* (Hübner,

85   1820) (Fig. 1). We employ the 659 bp cytochrome *c* oxidase I gene region at the continental scale

86   to determine whether limited dispersal abilities and contemporary barriers are preventing

87   movement in these species, what role the Pleistocene glaciations may have played in their current

88   structure and distribution, and if these three *Malacosoma* species show concordant

89   phylogeographic patterns or if there are differences in the patterns observed that may be

90   explained by their life history characteristics. By understanding what influences have impacted

91   the population genetic structure in these species we can add to our understanding of

92   phylogeography in North American insects and we can begin to predict how the populations may

93   expand with changing climate conditions.

94   **MATERIALS AND METHODS**

95   **Phylogenetic analyses**

96   Sequences of the 658 bp barcoding region of cytochrome *c* oxidase I (COI) for all

97   *Malacosoma* specimens from the United States and Canada were downloaded from the Barcode

98   of Life Data System (BOLD; see Supplemental Table S1; Ratnasingham & Hebert, 2007).

99   Locations were recorded by state or province. Sequences were aligned in MEGA v6 (Tamura *et*

100   *al.*, 2013). In order to verify that the samples examined belonged to monophyletic species, a

101   Bayesian network was constructed in BEAST v2.3 (Bouckaert *et al.*, 2014) using the Hasegawa,

102   Kishino, and Yano model with gamma-distributed rate variation and allowing for invariable sites

103   (HKY + Γ + I). The analysis was run for 10,000,000 MCMC steps, and sampled every 2,000

104   steps with a 25% burn-in. The lasiocampid *Phyllodesma americana* (GenBank Accession

105   Number JF842281) was used as the outgroup.

106   **Genetic analysis**

107   Intraspecific analyses were performed on the three widespread *Malacosoma* species: *M.*

108   *americana*, *M. californica*,and *M. disstria*. These species were selected based on distribution and

109   available sample size (Table 1, Fig. 1). Haplotypes were assigned to each species with TCS v1.21

110   (Clement*, Posada & Crandall, 2000) and confirmed manually. Genetic diversity was measured

111   with haplotype (H_d) and nucleotide (π) diversity indices, calculated in DNAsp v5.10 (Rozas *et al.*,

112   2003; Librado & Rozas, 2009). In order to test for population genetic structure an analysis of

113   molecular variance (AMOVA; Excoffier, Smouse & Quattro, 1992) and pairwise genetic

114   differences (Φ_{ST}; 100,000 permutations) were calculated in Arlequin v3.5.1.2 (Excoffier &

115   Lischer, 2010). Nearby sampling locations were grouped to increase sample sizes, and a modified

116   false discovery rate was applied to correct for multiple tests (Benjamini & Yekutieli, 2001).

117   To test for the presence of genetic clusters two methods were used: a spatial analysis of

118   molecular variance (SAMOVA) which identifies the maximum between group variance with the

119   use of additional geographic information (K = 2 to 6; 1,000 iterations; Dupanloup, Schneider &

120   Excoffier, 2002), and a clustering analysis as performed in Bayesian Analysis of Population

121   Structure v5.2 (BAPS; Corander & Tang, 2007; Corander *et al.*, 2008) which allows the

122   assignment of individuals to genetic clusters with no *a priori* population information. Results

123 from the *Malacosoma* Bayesian analysis were also analysed at the species level. In order to

124 visualise the pattern of variation and relationship among haplotypes, a principal coordinates

125 analysis (PCoA) was run in GenAlEx v6.5 (Peakall & Smouse, 2006; Peakall & Smouse, 2012),

126 and a statistical parsimony network was constructed in TCS v1.2.1(Clement, Posada & Crandall,

127 2000).

128 **RESULTS**

129 **Phylogenetic analyses**

130   A total of 474 *Malacosoma* sequences from five species were downloaded from BOLD.

131 The Bayesian network identified two main lineages within North America: the first group

132 included specimens of *M. constricta* and *M. disstria* with each species forming a well-defined

133 monophyletic group; the second included *M. americana*, *M. incurva*, and *M. californica*

134 (Supplemental Figure S1). While *M. americana* was monophyletic, specimens assigned to *M.*

135 *californica* were paraphyletic, suggesting a species complex. Of particular note, specimens of *M.*

136 *californica* from Alberta (AB),Saskatchewan (SK),and New Brunswick (NB), as well as those

137 identified as *M. californica pluviale*, did not group with the majority of *M. californica* samples.

138 The taxonomic status of *M. californica* has been widely debated; it is currently viewed as

139 including six largely allopatric subspecies, many of which have previously been viewed as

140 distinct species, although Franclemont questioned the validity of *M. californica pluviale* as a

141 subspecies rather than a species (Franclemont, 1973). By contrast, both *M. americana* and *M.*

142 *disstria* have no described subspecies (Stehr & Cook, 1968; Franclemont, 1973). Further study

143 with additional markers is required to clarify the taxonomic status of lineages within the *M.*

144 *californica* complex, and to determine how many species should be recognized. As such, the

145 intraspecific analyses here focus only on the large monophyletic portion of *M. californica* and do

146 not include *M. californica pluviale* or the NB, AB, or SK samples (see Fig. 1).

147 **Cytochrome *c* oxidase I sequences**

148 Intraspecific analyses examined 79 *M. americana* samples, 207 *M. californica* samples,

149 and 139 *M. disstria* samples (Supplemental Table S1). For each species the 658 bp gene region

150 was highly polymorphic with 37 variable sites defining 33 haplotypes in *M. americana*, 61

151 variable sites defining 64 haplotypes in *M. californica*, and 43 variable sites defining 42

152 haplotypes in *M. disstria* (Table 1). There were 8, 13, and 13 anticipated amino acid substitutions,

153 no frameshift mutations, and no stop codons. Fixed nucleotide differences were present in *M.*

154 *californica* between sCA, AZ, and the other populations of this species, and in *M. disstria*

155 between western (BC, AB, and SK) and eastern groups. There were no fixed differences in *M.*

156 *americana*. Haplotype and nucleotide diversities were high in all species (H_d = 0.918 - 0.926; π =

157 0.0049 - 0.0097; Table 1), with diversity generally higher in southern and central populations (r =

158 -0.284 to -0.508; Supplemental Table S2, Figure S2). When populations with small sample sizes

159 (n < 5) were removed from the correlation analysis, coefficients were much stronger for *M.*

160 *americana* and *M. californica* (r = -0.937 and r = -0.703, respectively), and slightly weaker for

161 *M. disstria* (r = -0.360).

162 **Genetic analyses**

163 ***Malacosoma americana***

164 Of the three species, *M. americana* showed the least population structure (overall Φ_{ST} =

165 0.244, p < 0.0001), the fewest significant pairwise comparisons (7 out of 15; Table 2), and the

166 lowest diversity (Table 1). The greatest pairwise Φ_{ST} values were seen between NB and the other

167 populations. The SAMOVA analysis identified four groups (F_{CT} = 0.263, p = 0.02): MN, NB, and

168 the separation of the remaining samples into northern (ON, MD/NC) and southern (TN,

169   AR/OK/TX) populations, while the Bayesian clustering analysis separated the samples into three

170   clusters (Fig. 1, Supplemental Table S2): a "northern" group found primarily in ON and NB, and

171   two "southern" groups, one found in all populations except NB, and a smaller group primarily in

172   MN and OK. The distribution of haplotypes in the Bayesian groups was significantly non-random

173   ($X^2 = 47.65$, p < 0.0001). The BEAST analysis generally supported the BAPS groups with

174   BAPS1 and BAPS2 forming weakly supported clades while BAPS3 contained the remaining

175   samples (posterior probability = 0.65 and 0.5, respectively; Supplemental Table S1, Figure S3).

176   The principal coordinates analysis showed a general separation of the samples into northern and

177   southern groups along coordinate 1 (34.7%; Fig. 2a), while coordinates 2 and 3 explained 11.5%

178   and 8.8% of the variation, respectively. The statistical parsimony network showed little pattern to

179   the variation, with haplotypes generally being closely related, although more divergent

180   haplotypes were present in ON, OK, and MN, and there was a general clustering of southern

181   samples (OK, TX, AR, and TN) and NB samples despite a lack of fixed differences (Fig.3).

182   ***Malacosoma californica***

183   *M. californica* had strong population structure (overall $\Phi_{ST} = 0.48$, p < 0.0001), and the

184   highest nucleotide diversity (Table 1). All pairwise comparisons were significant except that

185   between cBC and swBC, and those with WA (22 out of 28; Table 2). These are likely a result of

186   the small WA sample size (n=3), and the small geographic distance between cBC and swBC. The

187   greatest differences existed between sCA and all other populations ($\Phi_{ST} = 0.73 - 0.94$), and

188   between AZ and all other populations ($\Phi_{ST} = 0.54 - 0.79$).The SAMOVA analysis identified

189   genetic breaks between three groups: sCA, AZ, and all other populations ($F_{CT} = 0.57$, p = 0.049).

190   Bayesian clustering analysis identified six clusters (Fig. 1, Supplemental Table S2): one in sCA,

191   one in AZ, one in eBC, and three shared between multiple populations. The eBC population had

192   representatives in four clusters. The distribution of samples in the six clusters was highly

193   significant ($X^2 = 486.6$, p <0.0001). When sCA, CA, and AZ were removed from analysis, the

194   distribution was still significantly different than random ($X^2 = 57.02$, p < 0.0001). The BEAST

195   analysis identified four main clades with numerous subclades within them: two well-supported

196   clades representing the AZ and sCA samples (posterior probability = 0.95 and 1, respectively),

197   one clade found in eBC with one CA sample (posterior = 0.75), and one found across BC and WA

198   (posterior = 0.77; Supplemental Figure S4). The CA samples were found outside of these clades.

199   The principal coordinates analysis separated the BC samples into two groups along

200   coordinate 1, while the sCA and AZ samples were separated along coordinate 2 (Fig.2b). A total

201   of 65.1% of the variation was allocated to the first three coordinates (39.0%, 16.4%, and 9.7%,

202   respectively). The statistical parsimony network showed a similar pattern, separating the AZ,

203   sCA, CA, and TX samples from the more northern populations, with two large groups containing

204   all other samples (Fig.4). One group consisted of haplotypes from across the Pacific Northwest,

205   and contained all samples from cBC, VI, swBC, and WA, as well as many eBC samples. There

206   were several common haplotypes separated by one to three mutations. The second cluster was

207   restricted to eBC; it was less diverse and contained two common haplotypes, one represented by

208   44 individuals. Most CA samples had unique haplotypes, while the sCA and AZ haplotypes were

209   divergent with 13 - 15 (sCA) and 5 - 16 (AZ) mutations separating them from the nearest

210   population (Fig.4).

211   *Malacosoma disstria*

212   *M. disstria* had the strongest population structure (overall $\Phi_{ST} = 0.52$, p < 0.0001), and 42

213   of 55 pairwise comparisons were significant following correction for multiple tests (Table 2). The

214   13 non-significant comparisons all involved ON (with NB/NS, with NC/FL/GA, or among the

215   three ON locations). In contrast, the largest pairwise differences were between BC and all other

216   populations ($\Phi_{ST} = 0.49 - 0.98$). Diversity within populations was generally high with $H_d > 0.7$ in

217   10 of the 16 populations. The lowest values were in SK ($H_d = 0$) followed by eBC ($H_d = 0.20$;

218   Supplemental Table S2). A similar pattern was seen with nucleotide diversity with the highest

219   values in the three ON populations ($\pi = 0.0053 - 0.0065$) and TX ($\pi = 0.0137$; Supplemental

220   Figure S2).

221   The SAMOVA analysis identified a genetic break between the BC populations and all

222   other populations (including AB), possibly along the Rocky Mountains ($F_{CT} = 0.52$, $p = 0.015$).

223   Bayesian clustering analysis showed a slightly different pattern with four identified clusters (Fig.

224   1, Supplemental Table S2): three were found in the north (two wholly) while one was primarily

225   found in the south. The BC samples fell exclusively in one cluster, the AB,SK, and MB samples

226   were in a cluster together, and the ON, QC, and NB/NS samples generally formed a third cluster.

227   The allocation of samples to clusters was highly significant ($X^2 = 224.7$, $p < 0.0001$). The

228   BEAST analysis identified four main clades, generally supporting the BAPS groups, although

229   with slightly different membership: two well-supported clades found mostly in the west

230   (posterior probability $= 1$ and 0.75), and two clades present mostly in the east and southeast

231   (posterior $= 0.77$ and 0.39, respectively; Supplemental Figure S5).

232   The principal coordinates analysis separated the BC populations from most other

233   populations along coordinate 1 (41.5%), and identified a general separation of northern and

234   southern populations along coordinate 2 (14.8%). Coordinate 3 explained 12.6% of the variation

235   (Fig.2c). The statistical parsimony network identified moderate variation, with five common

236   haplotypes ($n \geq 10$): three found in a single region (BC, AB/SK, or ON), one shared by ON and

237   NB/NS, and one found across several regions in the east and south (Fig.5). The BC samples

238   formed a separate group (with a single AB sample), while AB, SK, and MB mostly grouped

239   together. The southern samples generally grouped together.

240 **DISCUSSION**

241 **Population genetic structure**

242   The analysis of mitochondrial COI sequences from three North American *Malacosoma*

243 species showed high levels of variation and diversity, with some highly divergent populations in

244 *M. californica*. All three species show evidence of persistence in one or more southern refugia

245 with subsequent recolonisation of northern regions. Diversity patterns exhibited the characteristic

246 "southern richness, northern purity" (Hewitt, 2004) found across much of the previously glaciated

247 Northern Hemisphere. This was particularly evident in *M. americana* where diversity in southern

248 populations was twice that in northern regions (π = 0.0046 - 0.0061 versus 0.0013 - 0.0027), and

249 in *M. californica* where diversity was four-fold higher in Arizona (π = 0.0093) than in

250 Washington (π = 0.002; Supplemental Table S2). In *M. disstria* diversity levels were highest in

251 Texas and in the three Ontario populations, likely representing admixture in the latter. All three

252 species exhibited a negative correlation between latitude and nucleotide diversity (Supplemental

253 Figure S2), with the strongest relationships seen in *M. americana* and *M. disstria*.

254   The two species found in the east, *M. americana* and the eastern portion of *M. disstria,*

255 showed limited population structure consistent with a relatively young evolutionary history

256 and/or high levels of gene flow. This pattern is common in species restricted to eastern North

257 America, and in the eastern portion of the range for more widespread species. For example, many

258 bird species found in the eastern half of North America exhibit limited genetic structure (Zink,

259 Rootes & Dittmann, 1991; Vallianatos, Lougheed & Boag, 2001; Veit *et al.*, 2005), while

260 widespread species show shallow divergence in the east (Klein & Brown, 1994; Graham & Burg,

261 2012; van Els, Cicero & Klicka, 2012), likely reflecting a single evolutionary origin and

262 extensive contemporary gene flow. Other studies have identified limited structure in eastern trees

263   (McLachlan, Clark & Manos, 2005; Shaw & Small, 2005; Gerardi *et al.*, 2010) and mammals

264   (Petersen & Stewart, 2006), with the exception of more distinct southern populations (e.g., Texas

265   or Florida). Both *M. americana* and *M. disstria* exhibit this pattern: limited structure in eastern

266   North America with some differences between northern and southern populations. This may be

267   caused by limited ongoing gene flow between regions, likely a result of the limited dispersal

268   capability of female *Malacosoma*. Interestingly, neither species have the genetic break between

269   Atlantic and Gulf coast clades seen in many fish (Bermingham & Avise, 1986; Avise, 1992),

270   insects (Vogler & Desalle, 1993; Ney & Schul, 2017), reptiles (Lamb & Avise, 1992), birds

271   (Avise, 1992), and marine invertebrates (Herke & Foltz, 2002; Young *et al.*, 2002; see Soltis *et*

272   *al.*, 2006 for additional references and an excellent description of this break), suggesting a single

273   southeastern origin.

274   In addition to limited north-south diversification, *M. disstria* also exhibits an east/west

275   separation (Table 2, Fig. 1), a pattern common in widespread North American species that often

276   reflects multiple evolutionary origins (Sperling, Raske & Otvos, 1999; Gerardi *et al.*, 2010;

277   Medina *et al.*, 2010; Lait & Burg, 2013). Lack of strong support for a western or southwestern

278   refugium may be a result of the paucity of samples in this region; the presence of a western

279   refugium may be suggested by the pattern found in *Populus tremuloides*, the favoured host of *M.*

280   *disstria*, that shows evidence of two genetic clusters, one in the southwest and one in the north

281   and east, with higher diversity in the southwest group (Callahan *et al.*, 2013). Many other

282   continent-wide species also possess a large group with low diversity across northern and eastern

283   areas, and multiple diverse groups west of the Rocky Mountains (Ball & Avise, 1992; Byun,

284   Koop & Reimchen, 1997; Graham & Burg, 2012; van Els, Cicero & Klicka, 2012), a pattern

285   linked to a single refugium for the east and multiple isolated refugia in the west. The high genetic

286   diversity in the three Ontario populations (Supplemental Table S2, Figure S2), as well as the

287   presence of all four Bayesian clusters meeting in these populations (Fig. 1), supports the

288   possibility of secondary admixture in this central region. Additional sampling between the

289   Ontario populations and the western populations, as well as from the southwestern portion of the

290   range, should help to clarify whether the mixing seen here is indicative of recolonisation from

291   multiple refugia, or if it suggests a diverse source population.

292   The western *M. californica* possessed a very different pattern of variation with strong

293   population genetic structure and very distinct populations (AZ and sCA) separated by multiple

294   fixed differences which may represent different subspecies or ecotypes. This pattern is common

295   in many southwestern and western species, and is explained by both the current and historical

296   topography of western North America: four major mountain ranges (Cascade, Coastal, Rocky,

297   and Sierra) run along a north-south axis, and large plains and deserts abound, all of which

298   contribute to a complex habitat mosaic. This heterogeneity, coupled with the resulting complex

299   glacial histories of the region, has resulted in extensive structuring in many birds (Barrowclough

300   *et al.*, 2004; Lait *et al.*, 2012; van Els, Cicero & Klicka, 2012), mammals (Byun, Koop &

301   Reimchen, 1997; Riddle, Hafner & Alexander, 2000; Galbreath *et al.*, 2009), insects (Brown *et

302   al.*, 1997), and plants (Golden & Bain, 2000; Richardson, Brunsfeld & Klopfenstein, 2002;

303   Johansen & Latta, 2003). Given the divergence of the AZ and sCA populations (0.9 - 1.9%

304   divergence from the nearest population, CA), it is likely that they have been isolated for multiple

305   glacial cycles with limited recent gene flow. The AZ population shows relatively high diversity,

306   suggesting multiple refugia or impassable barriers within this small region. This has been seen in

307   a number of animal (Orange, Riddle & Nickle, 1999; Zink *et al.*, 2001; Merrill, Ramberg &

308   Hagedorn, 2005; Graham *et al.*, 2013) and plant (Frohlich *et al.*, 1999) species, and is likely due

309   to the particularly heterogeneous nature of Arizona which contains seven ecoregions (Warshall,

310   1995; Poulos, Taylor & Beaty, 2007; Ober & Connolly, 2015; Powell & Steidl, 2015).

311   **Physical barriers**

312   Several physical barriers impede gene flow in North American species. In eastern North

313   America, the Appalachian Mountains act as a barrier to movement in plants (Griffin & Barrett,

314   2004; Joly & Bruneau, 2004; Godbout *et al.*, 2005), reptiles (Bushar *et al.*, 2014; Krysko *et al.*,

315   2017), and amphibians (Church *et al.*, 2003; Jones *et al.*, 2006), while the Mississippi,

316   Tombigbee, and Appalichola Rivers prevent gene flow in many plant and animal species (see

317   Bermingham & Avise, 1986; Avise, 1992; Soltis *et al.*, 2006 and references therein). The two

318   *Malacosoma* species in the east do not show genetic breaks along any of these traditional

319   barriers. This may be due to recent colonisation from a single origin or ongoing gene flow in

320   these regions. As the moths can fly, their dispersal capabilities should be greater than that of

321   sedentary plant and reptile species, allowing them to cross rivers. The fact that the Appalachians

322   have not prevented gene flow may indicate the importance of forested valleys as dispersal

323   corridors. All three *Malacosoma* species are generalist herbivores, albeit with host preferences,

324   thus they should encounter suitable habitat more often than strict specialists.

325   In western North America the main physical barriers are the Rocky, Coastal, Cascade, and

326   Sierra Nevada Mountains (Crease *et al.*, 1997; Nielson, Lohman & Sullivan, 2001; Johansen &

327   Latta, 2003; Burg *et al.*, 2005; Carstens *et al.*, 2005). The Wyoming Basin and the Great Plains

328   have also been shown to act as dispersal barriers, particularly for species associated with

329   montane, forested, or wetland habitats (DeChaine & Martin, 2005; Wilson *et al.*, 2005).

330   *Malacosoma disstria* and *M. californica* both exhibit genetic breaks in western regions: *M.*

331   *disstria* shows a break across the Rocky Mountains (between BC and AB), while *M. californica*

332   has disjunct populations among many of the southwestern deserts (Table 2, Figs. 1 & 2). The

333   Rocky Mountains may also act as a barrier in this species (or species complex); the samples

334   identified as *M. californica* from AB and SK were very different than those in BC (4.9 - 6.7%),

335   likely representing the described subspecies *M. californica lutescens* as a separate species (the

336   Great Plains tent caterpillar; Franclemont, 1973).

337   **Conclusions**

338   Despite considerable overlap in their distributions and life history traits, the patterns of

339   variation and levels of population structure in the three *Malacosoma* species varied considerably.

340   The population genetic structure suggests a single origin in the east and a complex evolutionary

341   history in the west. *M. americana*, restricted to the eastern half of the continent, shows limited

342   structure with a north-south trend and greater diversity in the south. This is consistent with its

343   expansion from a single southern refugium following the last glaciation with limited ongoing

344   gene flow between distance regions. *M. disstria* shows a similar pattern in the east, supporting a

345   single southern refugium, with one or more additional refugia possible in the west. Additional

346   samples are required to elucidate whether the differentiation in its BC population reflects a

347   founder event with subsequent divergence or additional structuring. *M. californica* shows the

348   greatest degree of structure and differentiation, consistent with multiple evolutionary origins in

349   the west and southwest. This study shows the utility of existing DNA barcodes in identifying

350   patterns of genetic structure in insect species, which can uncover previously unknown

351   evolutionary histories and suggest further avenues to explore.

352   **ACKNOWLEDGEMENTS**

353   We would like to acknowledge all of the contributors to the Barcode of Life Database including

354   the collections, laboratory, and bioinformatics staff at the Centre for Biodiversity Genomics, and

355   researchers from around the globe.

356   **DATA ARCHIVING**

357   All sequences are available on the Barcode of Life Database. Sampling locations and sequence

358   data used in this study (BOLD IDs and available GenBank Accession Numbers) have been

359   uploaded as online Supplemental Table S1.

360   **TITLES AND LEGENDS TO TABLES AND FIGURES**

361   **Table 1.** Sample size (n), number of sampling locations (loc), variable sites (VS), mean %

362   pairwise distances (PD), number of haplotypes (h), overall haplotype (Hd) and nucleotide (π)

363   diversities, and fixation index (Φ_{ST}) for three *Malacosoma* species. The three Φ_{ST} values were all

364   highly significant (p < 0.0001).

365   **Table 2.** Population pairwise Φ_{ST} values for (a) *M. americana* (P_{crit} = 0.015), (b) *M. californica*

366   (P_{crit} = 0.013), and (c) *M. disstria* (P_{crit} = 0.011). Φ_{ST} values are given below the diagonal and p-

367   values above the diagonal (* p< 0.05, ** p < 0.01, *** p < 0.001). Values significant following

368   correction for multiple tests are shaded. Refer to Fig. 1 for locations.

369   **Fig. 1.** Approximate distributions (shaded) and sampling locations for (a) *Malacosoma*

370   *americana*, (b) *M. californica,* and (c) *M. disstria*. The pie charts represent the distribution of

371   BAPS groups, scaled for sample size. The green crosses represent the omitted AB, SK, and NB

372   samples. Sampling locations are as follows: Alberta (AB), Arizona (AZ), British Columbia (BC;

373   central [c], eastern [e], and southwest [sw]), California (CA; southern [s]), Kentucky (KY),

374   Maryland (MD), Minnesota (MN), New Brunswick (NB),North Carolina (NC), Nova Scotia

375   (NS), Ontario (ON; central [c], eastern [e], and southern [s]), Saskatchewan (SK), Tennessee

376   (TN), Texas, (TX), Vancouver Island BC (VI), and Washington (WA).The STH includes Arkansas

377   (AR), Oklahoma (OK), and TX combined, and SE includes Florida (FL), Georgia (GA), and NC

378   combined.

379   **Fig.2.** Principal coordinates analysis for (a) *M. americana*, (b) *M. californica*, and (c) *M. disstria*.

380   The northern populations are depicted by a circle or square, while southern populations are

381   represented by a diamond. Samples are colour-coded by sampling location. Refer to Fig. 1 for

382   locations.

383   **Fig. 3.** Statistical parsimony network showing the relationship among the 33 *M. americana*

384   haplotypes. Each square represents one of the 79 sequences colour-coded by location, inferred

385   haplotypes are depicted by black circles, and each line represents a single nucleotide change.

386   Refer to Fig. 1 for locations.

387   **Fig.4.** Statistical parsimony network showing the relationship among the 64 *M. californica*

388   haplotypes. Each square represents one of the 207 sequences colour-coded by location, inferred

389   haplotypes are depicted by black circles, and each line represents a single nucleotide change.

390   Refer to Fig. 1 for locations.

391   **Fig.5.** Statistical parsimony network showing the relationship among 42 *M. disstria* haplotypes.

392   Each square represents one of the 139 sequences colour-coded by location, inferred haplotypes

393   are depicted by black circles, and each line represents a single nucleotide change. Refer to Fig. 1

394   for locations.

REFERENCES

Alvial I, Veliz D, Vargas H, Esquivel C, Vila I (2007). Lack of genetic structure in *Pantala*

flavescens among Central and South American localities (Odonata: Libellulidae).

Odonatologica **46:** 67-82. DOI: 10.5281/zenodo.572357.

Avise JC (1992). Molecular population structure and the biogeographic history of a regional

fauna - a case history with lessons for conservation biology. *Oikos* **63:** 62-76. DOI:

10.2307/3545516.

Avise JC (2004). *Molecular markers, natural history, and evolution*, 2nd edn. Sinauer &

Associates: Sunderland, MA.

Ball RM, Avise JC (1992). Mitochondrial-DNA phylogeographic differentiation among avian

populations and the evolutionary significance of subspecies. *The Auk* **109:** 626-636.

Barendregt RW, Irving E (1998). Changes in the extent of North American ice sheets during the

late Cenozoic. *Canadian Journal of Earth Sciences* **35:** 504-509. DOI: 10.1139/e97-126.

Barrowclough GF, Groth JG, Mertz LA, Gutiérrez RJ (2004). Phylogeographic structure, gene

flow and species status in blue grouse (*Dendragapus obscurus*). *Molecular Ecology* **13:**

1911-1922. DOI: 10.111/j.1365-294X.2004.02215.x.

Benjamini Y, Yekutieli D (2001). The control of the false discovery rate in multiple testing under

dependency. *The Annals of Statistics* **29:** 1165-1188. DOI: 10.1214/aos/1013699998.

Bermingham E, Avise JC (1986). Molecular zoogeography of freshwater fishes in the

southeastern United States. *Genetics* **113:** 939-965.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A,

Drummond AJ. (2014). BEAST 2: a software platform for Bayesian evolutionary

analysis. *PLOS Computational Biology* **10:** e1003537. DOI:

10.1371/journal.pcbi.1003537.

Brown JM, LeebensMack JH, Thompson JN, Pellmyr O, Harrison RG (1997). Phylogeography

and host association in a pollinating seed parasite *Greya politella* (Lepidoptera:

Prodoxidae). *Molecular Ecology* **6:** 215-224. DOI: 10.1046/j.1365-294X.1997.t01-1-

00171.x.

423   Burg TM, Gaston AJ, Winker K, Friesen VL (2005). Rapid divergence and postglacial

424   colonization in western North American Steller's jays (*Cyanocitta stelleri*). *Molecular*

425   *Ecology* **14:** 3745-3755. DOI: 10.1111/j.1365-294X.2005.02710.x.

426   Bushar LM, Aborde CCB, Gao SS, Gonzalez MV, Hoffman JA, Massaro IK, Savitzky AH,

427   Reinert HK. (2014). Genetic structure of timber rattlesnake (*Crotalus horridus*)

428   populations: physiographic influences and conservation implications. *Copeia* **2014:** 694-

429   706. DOI: 10.1643/CE-14-047.

430   Byun SA, Koop BF, Reimchen TE (1997). North American black bear mtDNA phylogeography:

431   Implications for morphology and the Haida Gwaii glacial refugium controversy.

432   *Evolution* **51:** 1647-1653. DOI: 10.2307/2411216.

433   Callahan CM, Rowe CA, Ryel RJ, Shaw JD, Madritch MD, Mock KE (2013). Continental-scale

434   assessment of genetic diversity and population structure in quaking aspen (*Populus*

435   *tremuloides*). *Journal of Biogeography* **40:** 1780-1791. DOI: 10.1111/jbi.12115.

436   Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J (2005). Investigating the

437   evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing

438   within a comparative phylogeographic framework. *Evolution* **59:** 1639-1652. DOI:

439   10.1554/04-661.1.

440   Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011). Rapid range shifts of species

441   associated with high levels of climate warming. *Science* **333:** 1024-1026. DOI:

442   10.1126/science.1206432.

443   Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003). Evidence for multiple

444   Pleistocene refugia in the postglacial expansion of the eastern tiger salamander,

445   *Ambystoma tigrinum tigrinum*. *Evolution* **57:** 372-383. DOI: 10.1111/j.0014-3820-

446   2003.tb00271.x.

447   Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene

448   genealogies. *Molecular Ecology* **9:** 1657-1659. DOI: 10.1046/j.1365-294x.2000.01020.x.

449   Corander J, Marttinen P, Siren J, Tang J (2008). Enhanced Bayesian modelling in BAPS software

450   for learning genetic structures of populations. *BMC Bioinformatics* **9:** 539. DOI:

451   10.1186/1471-2105-9-539.

452   Corander J, Tang J (2007). Bayesian analysis of population structure based on linked molecular

453   information. *Mathematical Biosciences* **205:** 19-31. DOI: 10.1016/j.mbs.2006.09.015.

454   Correa A, Vinson C, Braga L, Guedes R, de Oliveira L (2017). Ancient origin and recent range

455   expansion of the maize weevil *Sitophilus zeamais*, and its genealogical relationship to the

456   rice weevil *S. oryzae*. *Bulletin of Entomological Research* **107:** 9-20. DOI:

457   10.1017/S0007485316000687.

458   Costa JT, Ross KG (1994). Hierarchical genetic structure and gene flow in macrogeographic

459   populations of the eastern tent caterpillar (*Malacosoma americanum*). *Evolution* **48:**

460   1158-1167. DOI: 10.1111/j.1558-5646.1994.tb05302.x.

461   Craft KJ, Pauls SU, Darrow K, Miller SE, Hebert PDN, Helgen LE, Novotny V, Weiblen GD.

462   (2010). Population genetics of ecological communities with DNA barcodes: an example

463   from New Guinea Lepidoptera. *Proceedings of the National Academy of Sciences USA*

464   **107:** 5041-5046. DOI: 10.1073/pnas.0913084107.

465   Crease TJ, Lee S-K, Yu S-L, Spitze K, Lehman N, Lynch M (1997). Allozyme and mtDNA

466   variation in populations of the *Daphnia pulex* complex from both sides of the Rocky

467   Mountains. *Heredity* **79:** 242-251. DOI: 10.1038/hdy.1997.151.

468   Dalén L, Fuglei E, Hersteinsson P, Kapel C, Roth J, Samelius G, Tannerfeld M, Angerbjörn A.

469   (2005). Population history and genetic structure of a circumpolar species: the arctic fox.

470   *Biological Journal of the Linnean Society* **84:** 79-89. DOI: 10.1111/j.1095-

471   8312.2005.00415.x.

472   DeChaine EG, Martin AP (2005). Historical biogeography of two alpine butterflies in the Rocky

473   Mountains: broad-scale concordance and local-scale discordance. *Journal of*

474   *Biogeography* **32:** 1943-1956. DOI: 10.1111/j.1365-2699.2005.01356.x.

475   Dinca V, Dapporto L, Vila R (2011). A combined genetic-morphometric analysis unravels the

476   complex biogeographical history of *Polyommatus icarus* and *Polyommatus celina*

477   common blue butterflies. *Molecular Ecology* **20:** 3921-3935. DOI: 10.1111/j.1365-

478   294X.2011.05223.x.

479 Dupanloup I, Schneider S, Excoffier L (2002). A simulated annealing approach to define the

480   genetic structure of populations. *Molecular Ecology* **11:** 2571-2581. DOI: 10.1046/j.1365-

481   294X.2002.01650.x.
482 Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ. (2002). The

483   Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary*

484   *Science Reviews* **21:** 9-31. DOI: 10.1016/S0277-3791(01)00095-6.
485 Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform

486   population genetics analyses under Linux and Windows. *Molecular Ecology Resources*

487   **10:** 564-567. DOI: 10.1111/j.1755-0998.2010.02847.x.
488 Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric

489   distances among DNA haplotypes: application to human mitochondrial DNA restriction

490   data. *Genetics* **131:** 479-491.
491 Franclemont JG (1973). *Mimallonoidea: Mimallonidae and Bombycoidea: Apatelodidae,*

492   *Bombycidae, Lasiocampidae. The Moths of North America North of Mexico, Fascicle*

493   *20.1*. The Wedge Entomological Research Foundation.
494 Franklin MT, Myers JH, Cory JS (2014). Genetic similarity of island populations of tent

495   caterpillars during successive outbreaks. *PLOS One* **9:** e96679. DOI:

496   10.1371/journal.pone.0096679.
497 Frantine-Silva W, Giangarelli DC, Penha RES, Suzuki KM, Dec E, Gaglianone MC, Alves-dos-

498   Santos I, Sofia SH. (2017). Phylogeography and historical demography of the orchid bee

499   *Euglossa iopoecila*: signs of vicariant events associated to Quaternary climatic changes.

500   *Conservation Genetics* **18:** 539-552. DOI: 10.1007/s10592-016-0905-7.
501 Frohlich DR, Torres-Jerez I, Bedford ID, Markham PG, Brown JK (1999). A phylogeographical

502   analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers.

503   *Molecular Ecology* **8:** 1683-1691.
504 Galbreath KE, Hafner DJ, Zamudio KR, Agnew K (2009). Isolation and introgression in the

505   intermountain west: Contrasting gene genealogies reveal the complex biogeographic

506   history of the American pika (*Ochotona princeps*). *Journal of Biogeography* **37:** 344-362.

507   DOI: 10.1111/j.1365-2699.2009.02201.x.

508   Gerardi S, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2010). From glacial refugia to modern

509   populations: new assemblages of organelle genomes generated by differential cytoplasmic

510   gene flow in transcontinental black spruce. *Molecular Ecology* **19:** 5265-5280. DO:

511   10.1111/j.1365-294X.2010.04881.x.
512   Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005). A mitochondrial DNA

513   minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range

514   North American conifer. *Molecular Ecology* **14:** 3497-3512. DOI: 10.1111/j.1365-

515   294X.2005.02674.x.
516   Golden JL, Bain JF (2000). Phylogeographic patterns and high levels of chloroplast DNA

517   diversity in four *Packera* (Asteraceae) species in southwestern Alberta. *Evolution* **54:**

518   1566-1579. DOI: 10.1111/j.0014-3820.2000.tb00702.x.
519   Graham BA, Burg TM (2012). Molecular markers provide insight into contemporary and historic

520   gene flow for a non-migratory species. *Journal of Avian Biology* **43:** 198-214. DOI:

521   10.1111/j.1600-048X.2012.05604.x.
522   Graham MR, Jaeger JR, Prendini L, Riddle BR (2013). Phylogeography of the Arizona hairy

523   scorpion (Hadrurus arizonensis) supports a model of biotic assembly in the Mojave Desert

524   and adds a new Pleistocene refugium. *Journal of Biogeography* **40:** 1298-1312. DOI:

525   10.1111/jbi.12079.
526   Griffin SR, Barrett SCH (2004). Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in

527   eastern North America: inferences from phylogeography. *American Journal of Botany* **91:**

528   465-473. DOI: 10.3732/ajb.91.3.465.
529   Herke SW, Foltz DW (2002). Phylogeography of two squid (*Loligo pealei* and *L. plei*) in the Gulf

530   of Mexico and northwest Atlantic Ocean. *Marine Biology* **140:** 103-115. DOI:

531   10.1007/s002270100680.
532   Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and

533   speciation. *Biological Journal of the Linnean Society* **58:** 247-276. DOI: 10.1111/j.1095-

534   8312.1996.tb01434.x.
535   Hewitt GM (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907-913. DOI:

536   10.1038/35016000.

537   Hewitt GM (2004). Genetic consequences of climatic oscillations in the Quaternary.

538   *Philosophical Transactions of the Royal Society B: Biological Sciences* **359:** 183-195.

539   DOI: 10.1098/rstb.2003.1388.
540   Hildahl V, Reeks WA (1960). Outbreaks of the forest tent caterpillar, *Malacosoma disstria* Hbn.,

541   and their effects on stands of trembling aspen in Manitoba and Saskatchewan. *The*

542   *Canadian Entomologist* **92:** 199-209. DOI: 10.4039/Ent92199-3.
543   Johansen AD, Latta RG (2003). Mitochondrial haplotype distribution, seed dispersal and patterns

544   of postglacial expansion of ponderosa pine. *Molecular Ecology* **12:** 293-298. DOI:

545   10.1046/j.1365-294X.2003.01723.x.
546   Joly S, Bruneau A (2004). Evolution of triploidy in *Apios americana* (Leguminosae) revealed by

547   genealogical analysis of the histone H3-D gene. *Evolution* **58:** 284-295. DOI:

548   10.1111/j.0014-3820.2004.tb01645.x.
549   Jones MT, Voss SR, Ptacek MB, Weisrock DW, Tonkyn DW (2006). River drainages and

550   phylogeography: an evolutionary significant lineage of shovel-nosed salamander

551   (*Desmognathus marmoratus*) in the southern Appalachians. *Molecular Phylogenetics and*

552   *Evolution* **38:** 280-287. DOI: 10.1016/j.ympev.2005.05.007.
553   Karthika P, Vadivalagan C, Krishnaveni N, Murugan K, Nicoletti M, Canale A, Benelli G. (2017).

554   Contrasting genetic diversity and intra-population polymorphism of the invasive pest

555   *Henosepilachna vigintioctopunctata* (Coleoptera, Coccinellidae): a DNA barcoding

556   approach. *Journal of Asia-Pacific Entomology* **20:** 23-29. DOI:

557   10.1016/j.aspen.2016.11.011.
558   Kirichenko N, Triberti P, Ohshima I, Haran J, Byun B-K, Li H, Augustin S, Roques A, Lopez-

559   Vaamonde C. (2017) From east to west across the Palearctic: phylogeography of the

560   invasive lime leaf miner *Phyllonorycter issikii* (Lepidoptera: Gracillariidae) and discovery

561   of a putative new cryptic species in East Asia. *PLOS One* **12:***e0171104.* DOI:

562   10.1371/journal.pone.0171104.

563   Klein NK, Brown WM (1994). Intraspecific molecular phylogeny in the yellow warbler

564   (*Dendroica petechia*), and implications for avian biogeography in the West Indies.

565   *Evolution* **48:** 1914-1932. DOI: 10.1111/j.1558-5646.1994.tb02223.x.

566   Knowles L (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*)

567   from the sky islands of western North America. *Evolution* **54:** 1337-1348. DOI:

568   10.1111/j.0014-3820.2000.tb00566.x.

569   Krysko KL, Nunez LP, Newman CE, Bowen BW (2017). Phylogenetics of kingsnakes,

570   *Lampropeltis getula* complex (Serpentes: Colubridae), in eastern North America. *Journal*

571   *of Heredity* **108:** 226-238. DOI: 10.1093/jhered/esw086.

572   Lait LA, Burg TM (2013). When east meets west: population structure of a high-latitude resident

573   species, the boreal chickadee (*Poecile hudsonicus*). *Heredity* **111:** 321-329. DOI:

574   10.1038/hdy.2013.54.

575   Lait LA, Friesen VL, Gaston AJ, Burg TM (2012). The post-Pleistocene population genetic

576   structure of a western North American passerine: the chestnut-backed chickadee (*Poecile*

577   *rufescens*). *Journal of Avian Biology* **43:** 541-552. DOI: 10.1111/j.1600-

578   048X.2012.05761.x.

579   Lamb T, Avise JC (1992). Molecular and population genetic aspects of mitochondrial DNA

580   variability in the diamondback terrapin, *Malaclemys terrapin*. *Journal of Heredity* **83:**

581   262-269. DOI: 10.1093/oxfordjournals.jhered.a111211.

582   Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA

583   polymorphism data. *Bioinformatics* **25:** 1451-1452. DOI: 10.1093/bioinformatics/btp187.

584   McLachlan JS, Clark JS, Manos PS (2005). Molecular indicators of tree migration capacity under

585   rapid climate change. *Ecology* **86:** 2088-2098. DOI: 10.1890/04-1036.

586   Medina RF, Rondon SI, Reyna SM, Dickey AM (2010). Population structure of *Phthorimaea*

587   *operculella* (Lepidoptera: Gelechiidae) in the United States. *Environmental Entomology*

588   **39:** 1037-1042. DOI: 10.1603/en09286.

589   Merrill SA, Ramberg FB, Hagedorn HH (2005). Phylogeography and population structure of

590   *Aedes aegypti* in Arizona. *The American Journal of Tropical Medicine and Hygiene* **72:**

591   304-310.

592   Ney G, Schul J (2017). Population structure within the one-dimensional range of a coastal plain

593   katydid. *PLOS One* **12**. DOI: 10.1371/journal.pone.0179361.
594   Nielson M, Lohman K, Sullivan J (2001). Phylogeography of the tailed frog (*Ascaphus truei*):

595   implications for the biogeography of the Pacific Northwest. *Evolution* **55:** 147-160. DOI:

596   10.1111/j.0014-3820.2001.tb01280.x.
597   Nilsson T (1983). *The Pleistocene: Geology and Life in the Quaternary Ice Age*. D. Reidel

598   Publishing Company: Dordrecht, Holland.
599   Ober KA, Connolly CT (2015). Geometric morphometric and phylogenetic analyses of Arizona

600   Sky Island populations of *Scaphinotus petersi* Roeschke (Coleoptera: Carabidae).

601   *Zoological Journal of the Linnean Society* **175:** 107-118. DOI: 10.1111/zoj.12269.
602   Orange DI, Riddle BR, Nickle DC (1999). Phylogeography of a wide-ranging desert lizard,

603   *Gambelia wislizenii* (Crotaphytidae). *Copeia* **1999:** 267-273. DOI: 10.2307/1447471.
604   Parry D, Goyer RA (2004). Variation in the suitability of host tree species for geographically

605   discrete populations of forest tent caterpillar. *Environmental Entomology* **33:** 1477-1487.

606   DOI: 10.1603/0046-225X-33.5.1477.
607   Peakall ROD, Smouse PE (2006). GenAlEx 6: Genetic Analysis in Excel. Population genetic

608   software for teaching and research. *Molecular Ecology Notes* **6:** 288-295. DOI:

609   10.1111/j.1471-8286.2005.01155.x.
610   Peakall ROD, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic

611   software for teaching and research - an update. *Bioinformatics* **28:** 2537-2539. DOI:

612   10.1093/bioinformatics/bts460.
613   Petersen SD, Stewart DT (2006). Phylogeography and conservation genetics of southern flying

614   squirrels(*Glaucomys volans*) from Nova Scotia. *Journal of Mammalogy* **87:** 153-160.

615   DOI: 10.1644/05-MAMM-A-062R1.1.
616   Pielou EC (1991). *After the Ice Age: The Return of Life to Glaciated North America*. University

617   of Chicago Press: Chicago, IL.
618   Poulos HM, Taylor AH, Beaty RM (2007). Environmental controls on dominance and diversity of

619   woody plant species in a Madrean, sky island ecosystem, Arizona, USA. *Plant Ecology*

620   **193:** 15-30. DOI: 10.1007/s11258-006-9245-x.

621 Powell BF, Steidl RJ (2015). Influence of vegetation on montane riparian bird communities in the

622   sky islands of Arizona, USA. *Southwestern Naturalist* **60:** 65-71. DOI: 10.1894/mcg-09.1.
623 Ratnasingham S, Hebert PDN (2007). BOLD: The Barcode of Life Data System

624   (www.barcodinglife.org). *Molecular Ecology Notes* **7:** 355-364. DOI: 10.1111/j.1471-

625   8286.2006.01678.x.
626 Richardson BA, Brunsfeld SJ, Klopfenstein NB (2002). DNA from bird-dispersed seed and wind-

627   disseminated pollen provides insights into postglacial colonization and population genetic

628   structure of whitebark pine (*Pinus albicaulis*). *Molecular Ecology* **11:** 215-227.
629 Riddle BR, Hafner DJ, Alexander LF (2000). Comparative phylogeography of Baileys' pocket

630   mouse (*Chaetodipus*) and the *Peromyscus eremicus* species group: historical vicariance of

631   the Baja California peninsular desert. *Molecular Phylogenetics and Evolution* **17:** 161-

632   172. DOI: 10.1006/mpev.2000.0842.
633 Roland J (1993). Large-scale forest fragmentation increases the duration of tent caterpillar

634   outbreak. *Oecologia* **93:** 25-30. DOI: 10.1007/BF00321186.
635 Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism

636   analyses by the coalescent and other methods. *Bioinformatics* **19:** 2496-2497. DOI:

637   10.1093/bioinformatics/btg359.
638 Shaw J, Small RL (2005). Chloroplast DNA phylogeny and phylogeography of the North

639   American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American*

640   *Journal of Botany* **92:** 2011-2030. DOI: 10.3732/ajb.92.12.2011.
641 Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006). Comparative phylogeography

642   of unglaciated eastern North America. *Molecular Ecology* **15:** 4261-4293. DOI:

643   10.1111/j.1365-294X.2006.03061.x.
644 Sperling FAH, Raske AG, Otvos IS (1999). Mitochondrial DNA sequence variation among

645   populations and host races of *Lambdina fiscellaria* (Gn.) (Lepidoptera : Geometridae).

646   *Insect Molecular Biology* **8:** 97-106. DOI: 10.1046/j.1365-2583.1999.810097.x.
647 Stehr FW, Cook EF (1968). *A revision of the genus Malacosoma Hübner in North America*

648   *(Lepidoptera: Lasiocampidae): systematics, biology, immatures, and parasites*, Vol 276.

649   Smithsonian Institution Press: Washington D.C.

650 Stewart JR, Lister AM (2001). Cryptic northern refugia and the origins of the modern biota.

651 *Trends in Ecology & Evolution* **16:** 608-613. DOI: 10.1016/S0169-5347(01)02338-2.
652 Stewart JR, Lister AM, Barnes I, Dálen L (2009). Refugia revisited: individualistic responses of

653 species in space and time. *Proceedings of the Royal Society B: Biological Sciences* **277:**

654 661-671. DOI: 10.1098/rspb.2009.1272.
655 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary

656 Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30:** 2725-2729. DOI:

657 10.1093/molbev/mst197.
658 Vallianatos M, Lougheed SC, Boag PT (2001). Phylogeography and genetic characteristics of a

659 putative secondary-contact zone of the loggerhead shrike in central and eastern North

660 America. *Canadian Journal of Zoology* **79:** 2221-2227. DOI: 10.1139/z01-157.
661 Vandewoestijne S, Baguette M, Brakefield PM, Saccheri IJ (2004). Phylogeography of *Aglais*

662 *urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and

663 control region. *Molecular Phylogenetics and Evolution* **31:** 630-646. DOI:

664 10.1016/j.ympev.2003.09.007.
665 van Els P, Cicero C, Klicka J (2012). High latitudes and high genetic diversity: phylogeography

666 of a widespread boreal bird, the gray jay (*Perisoreus canadensis*). *Molecular*

667 *Phylogenetics and Evolution* **63:** 456-465. DOI: 10.1016/j.ympev.2012.01.019.
668 Veit ML, Robertson RJ, Hamel PB, Friesen VL (2005). Population genetic structure and dispersal

669 across a fragmented landscape in cerulean warblers (*Dendroica cerulea*). *Conservation*

670 *Genetics* **6:** 159-174.
671 Vogler AP, Desalle R (1993). Phylogeographic patterns in coastal North American tiger beetles

672 (*Cicindela dorsalis* Say) inferred from mitochondrial DNA sequences. *Evolution* **47:**

673 1192-1202. DOI: 10.1111/j.1558-5646.1993.tb02146.x.
674 Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J-M, Hoegh-

675 Guldberg O, Bairlein F (2002). Ecological responses to recent climate change. *Nature*

676 **416:** 389-395.
677 Warshall P (1995). *The Madrean sky island archipelago: a planetary overview*, Vol 264.
678 Wilson GM, Den Bussche RA, McBee K, Johnson LA, Jones CA (2005). Intraspecific

679 phylogeography of red squirrels (*Tamiasciurus hudsonicus*) in the central Rocky

680   Mountain region of North America. *Genetica* **125:** 141-154. DOI: 10.1007/s10709-005-

681   5154-5.

682 Young AM, Torres C, Mack JE, Cunningham CW (2002). Morphological and genetic evidence

683   for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit crab

684   *Pagurus longicarpus*. *Marine Biology* **140:** 1059-1066.

685 Zink RM, Kessen AE, Line TV, Blackwell-Rago RC (2001). Comparative phylogeography of

686   some aridland bird species. *The Condor* **103:** 1-10. DOI: 10.1650/0010-

687   5422(2001)103[0001:CPOSAB]2.0.CO;2.

688 Zink RM, Rootes WL, Dittmann DL (1991). Mitochondrial DNA variation, population structure,

689   and evolution of the common grackle (*Quiscalus quiscula*). *The Condor* **93:** 318-329.

690   DOI: 10.2307/1368947.

Sample details for three *Malacosoma* species.

Sample size (n), number of sampling locations (loc), variable sites (VS), mean % pairwise distances (PD), number of haplotypes (h), overall haplotype (Hd) and nucleotide (π) diversities, and fixation index (Φ_{ST}) for three *Malacosoma* species. The three F_{ST} values were all highly significant ($p < 0.0001$).

Species	n	loc	VS	PD ± SE	h	Hd	π	Φ_{ST}
M. americana	79	12	37	0.490 ± 0.303	33	0.918	0.0049	0.244
M. californica	207	9	61	0.968 ± 0.668	64	0.925	0.0097	0.477
M. disstria	139	19	43	0.628 ± 0.383	42	0.926	0.0063	0.524

Population pairwise Φ_{ST} values for three *Malacosoma* species.

Population pairwise Φ_{ST} values for (a) *M. americana* (P_{crit} = 0.015), (b) *M. californica* (P_{crit} = 0.013), and (c) *M. disstria* (P_{crit} = 0.011). F_{ST} values are given below the diagonal and p-values above the diagonal (* p< 0.05, ** p < 0.01, *** p < 0.001). Values significant following correction for multiple tests are shaded. Refer to Fig. 1 for locations.

(a)

	NB	ON/QC	MN	MD/NC	TN	STH[1]
NB	-	**	**	**	**	***
ON/QC	0.195	-	**	0.702	*	***
MN	0.637	0.395	-	0.298	0.257	*
MD/NC	0.481	0.000	0.326	-	0.484	0.321
TN	0.564	0.175	0.202	0.000	-	0.376
STH[1]	0.444	0.213	0.271	0.015	0.007	-

(b)

	CBC	EBC	SWBC	VI	WA	CA	sCA	AZ
CBC	-	***	0.185	***	*	***	**	***
EBC	0.249	-	**	***	*	**	***	***
SWBC	0.069	0.233	-	**	0.143	**	**	***
VI	0.403	0.313	0.247	-	0.117	***	***	***
WA	0.352	0.258	0.229	0.138	-	*	*	**
CA	0.448	0.241	0.422	0.641	0.500	-	**	***
sCA	0.892	0.725	0.907	0.918	0.935	0.799	-	**
AZ	0.682	0.618	0.661	0.789	0.662	0.544	0.732	-

(c)

	cBC	eBC	AB	SK	cON	sON	eON	NB/NS	STH[1]	TN	SE[2]
cBC	-	0.305	***	***	**	***	***	***	***	***	**
eBC	0.044	-	***	***	***	***	***	***	***	***	***
AB	0.677	0.719	-	0.284	***	***	***	***	***	***	**
SK	0.948	0.978	0.026	-	***	***	**	***	***	***	***
cON	0.541	0.626	0.357	0.494	-	0.319	0.897	*	**	***	0.057
sON	0.617	0.660	0.255	0.362	0.016	-	0.264	0.109	***	***	*
eON	0.486	0.530	0.294	0.367	0.000	0.012	-	*	**	**	*
NB/NS	0.769	0.808	0.284	0.468	0.201	0.037	0.134	-	***	***	**
STH[1]	0.612	0.683	0.354	0.496	0.259	0.212	0.201	0.269	-	***	0.241
TN	0.873	0.912	0.473	0.835	0.440	0.300	0.313	0.407	0.232	-	*
SE[2]	0.824	0.890	0.349	0.781	0.272	0.155	0.188	0.246	0.035	0.317	-

3 [1]STH = AR, OK, TX   [2]SE = NC, GA, FL   * p< 0.05, ** p < 0.01, *** p < 0.001

Figure 1

Distributions, sampling locations, and Bayesian cluster membership for three *Malacosoma* species.

Approximate distributions (shaded) and sampling locations for (a) *Malacosoma americana*, (b) *M. californica*, and (c) *M. disstria*. The pie charts represent the distribution of BAPS groups, scaled for sample size. The green crosses represent the omitted AB, SK, and NB samples. Sampling locations are as follows: Alberta (AB), Arizona (AZ), British Columbia (BC; central [c], eastern [e], and southwest [sw]), California (CA; southern [s]), Kentucky (KY), Maryland (MD), Minnesota (MN), New Brunswick (NB),North Carolina (NC), Nova Scotia (NS), Ontario (ON; central [c], eastern [e], and southern [s]), Saskatchewan (SK), Tennessee (TN), Texas, (TX), Vancouver Island BC (VI), and Washington (WA).The STH includes Arkansas (AR), Oklahoma (OK), and TX combined, and SE includes Florida (FL), Georgia (GA), and NC combined.

(a)

MN
ON
QC
NB
IL
KY
TN
MD/NC
STH

BAPS1 ▢ (yellow)
BAPS2 ▢ (red)
BAPS3 ▢ (blue)

(b)

cBC
eBC
VI
WA
swBC
CA
sCA
AZ
TX

BAPS1 ▢ (yellow)   BAPS4 ▢ (cyan)
BAPS2 ▢ (red)   BAPS5 ▢ (pink)
BAPS3 ▢ (blue)   BAPS6 ▢ (green)

(c)

cBC
eBC
SK
MB
cON
QC
sON
NB
NS
KY
eON
TN
STH
SE

BAPS1 ▢ (yellow)
BAPS2 ▢ (red)
BAPS3 ▢ (blue)
BAPS4 ▢ (cyan)
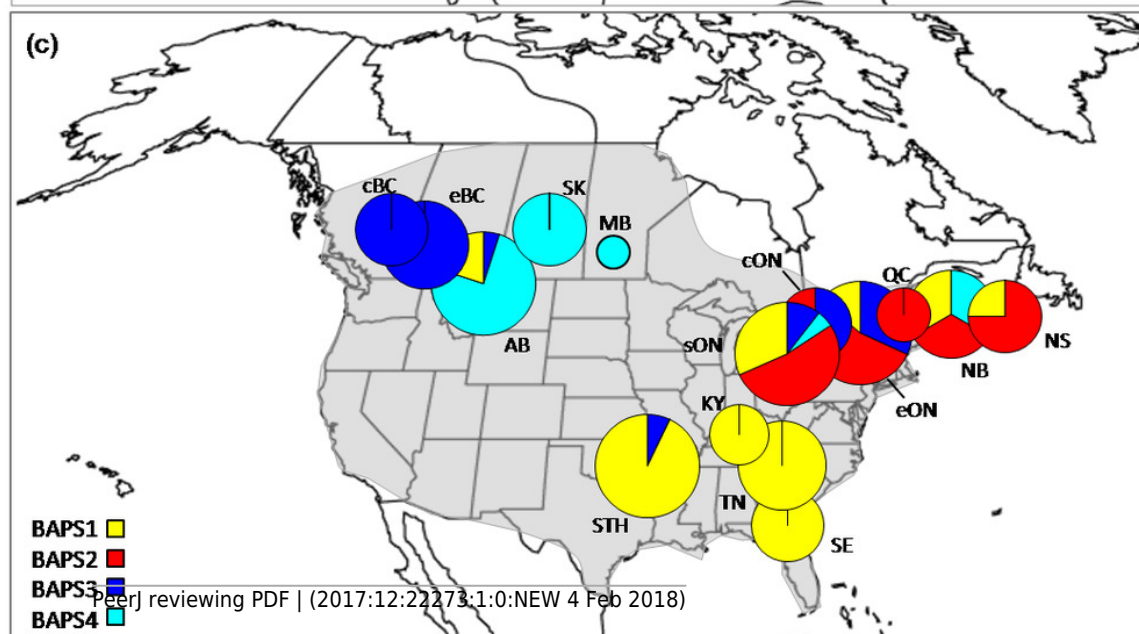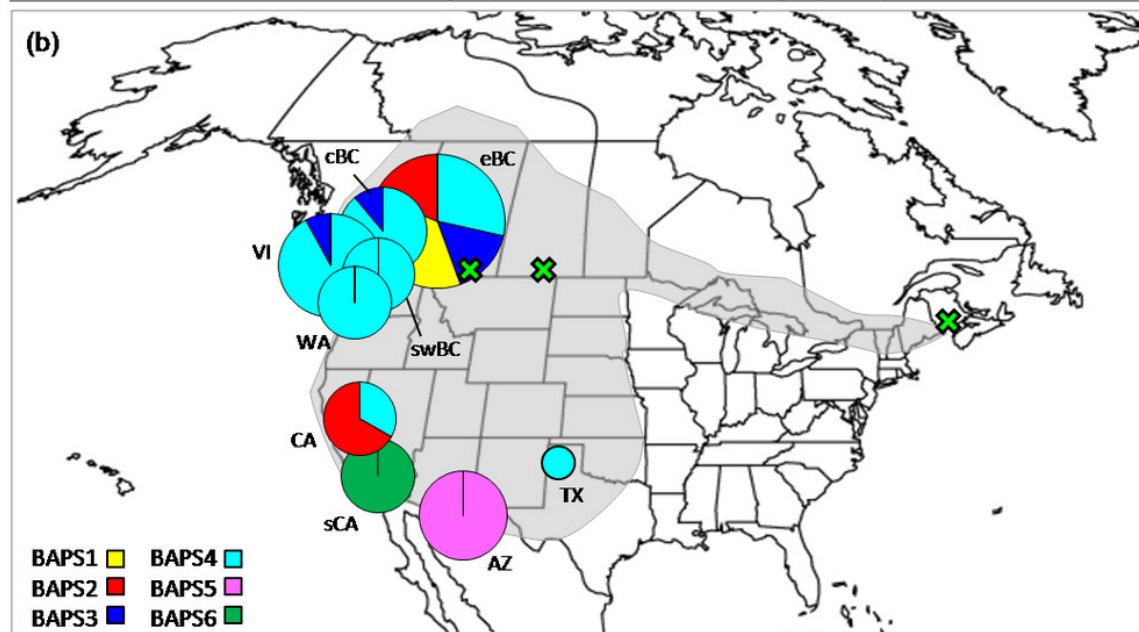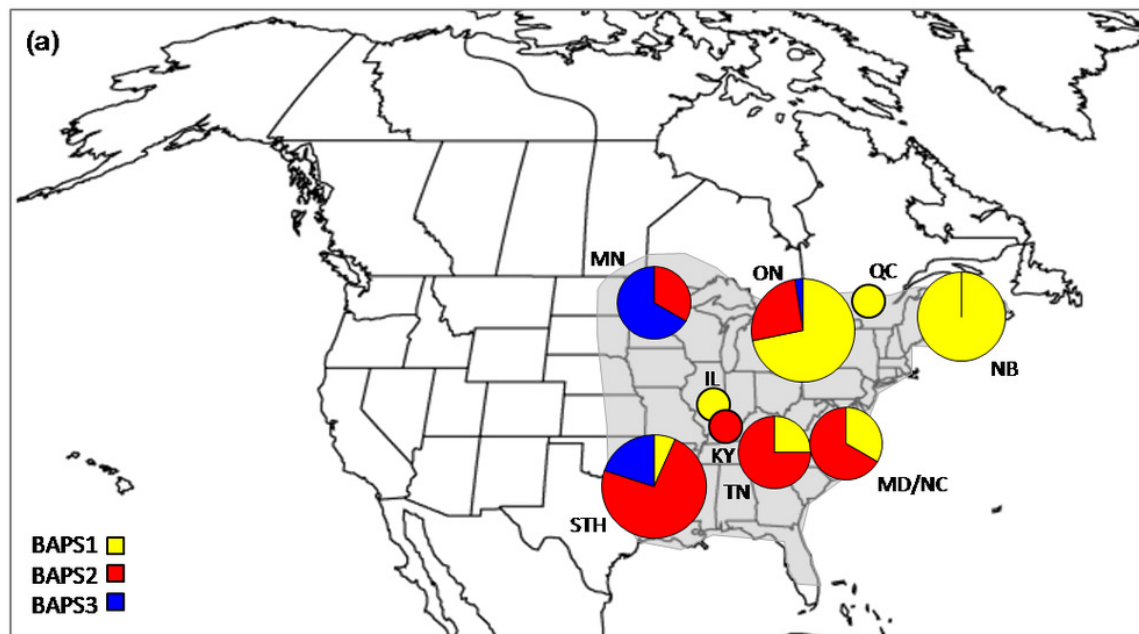
Figure 2

Principal coordinates analysis for three *Malacosoma* species.

Principal coordinates analysis for (a) *M. americana*, (b) *M. californica*, and (c) *M. disstria*. The northern populations are depicted by a circle or square, while southern populations are represented by a diamond. Samples are colour-coded by sampling location. Refer to Fig. 1 for locations.

PeerJ
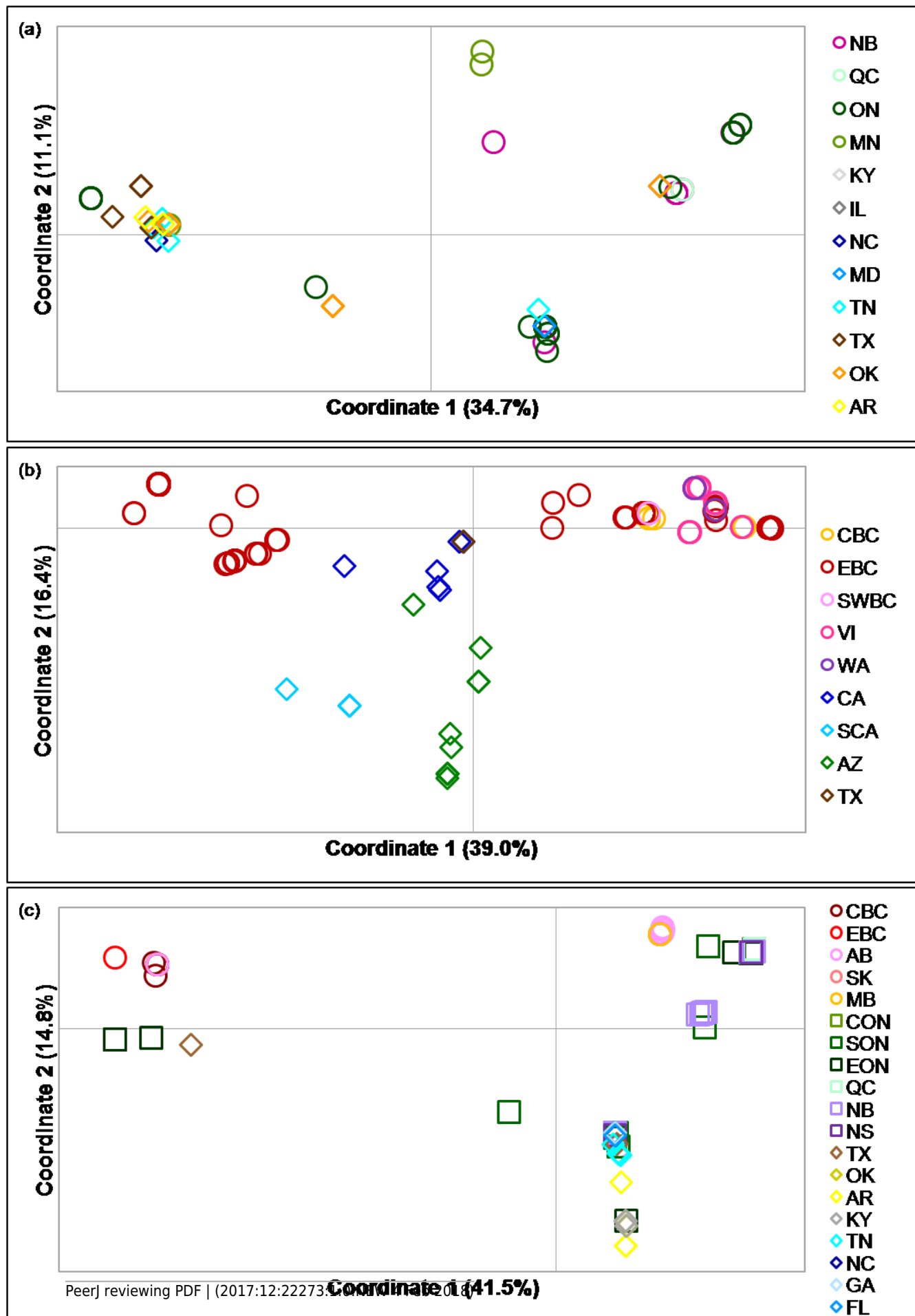
Figure 3

Statistical parsimony network for *M. americana.*

Statistical parsimony network showing the relationship among the 33 *M. americana*
haplotypes. Each square represents one of the 79 sequences colour-coded by location,
inferred haplotypes are depicted by black circles, and each line represents a single
nucleotide change. Refer to Fig. 1 for locations.
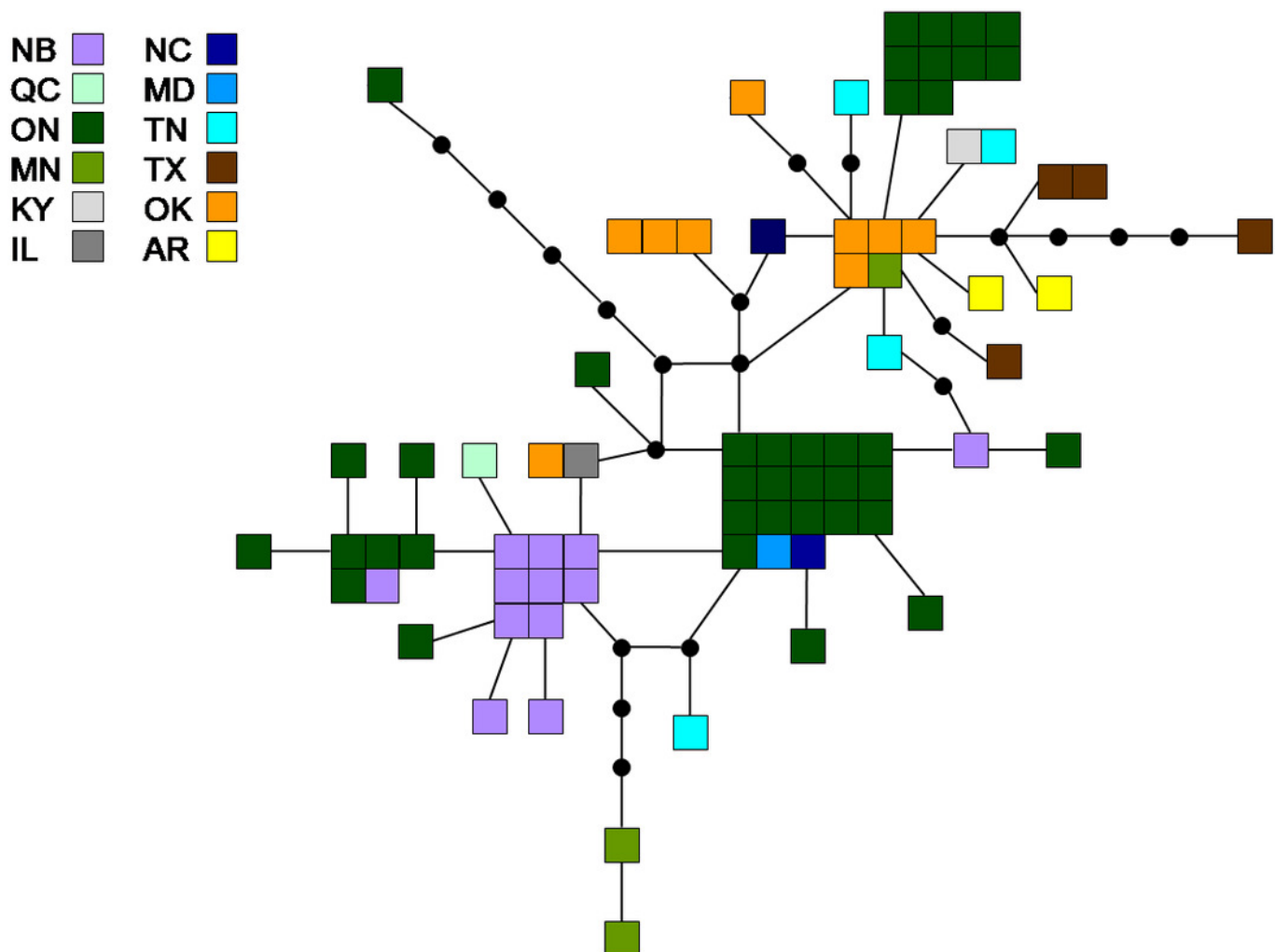
Figure 4

Statistical parsimony network for *M. californica.*

Statistical parsimony network showing the relationship among the 64 *M. californica* haplotypes. Each square represents one of the 207 sequences colour-coded by location, inferred haplotypes are depicted by black circles, and each line represents a single nucleotide change. Refer to Fig. 1 for locations.
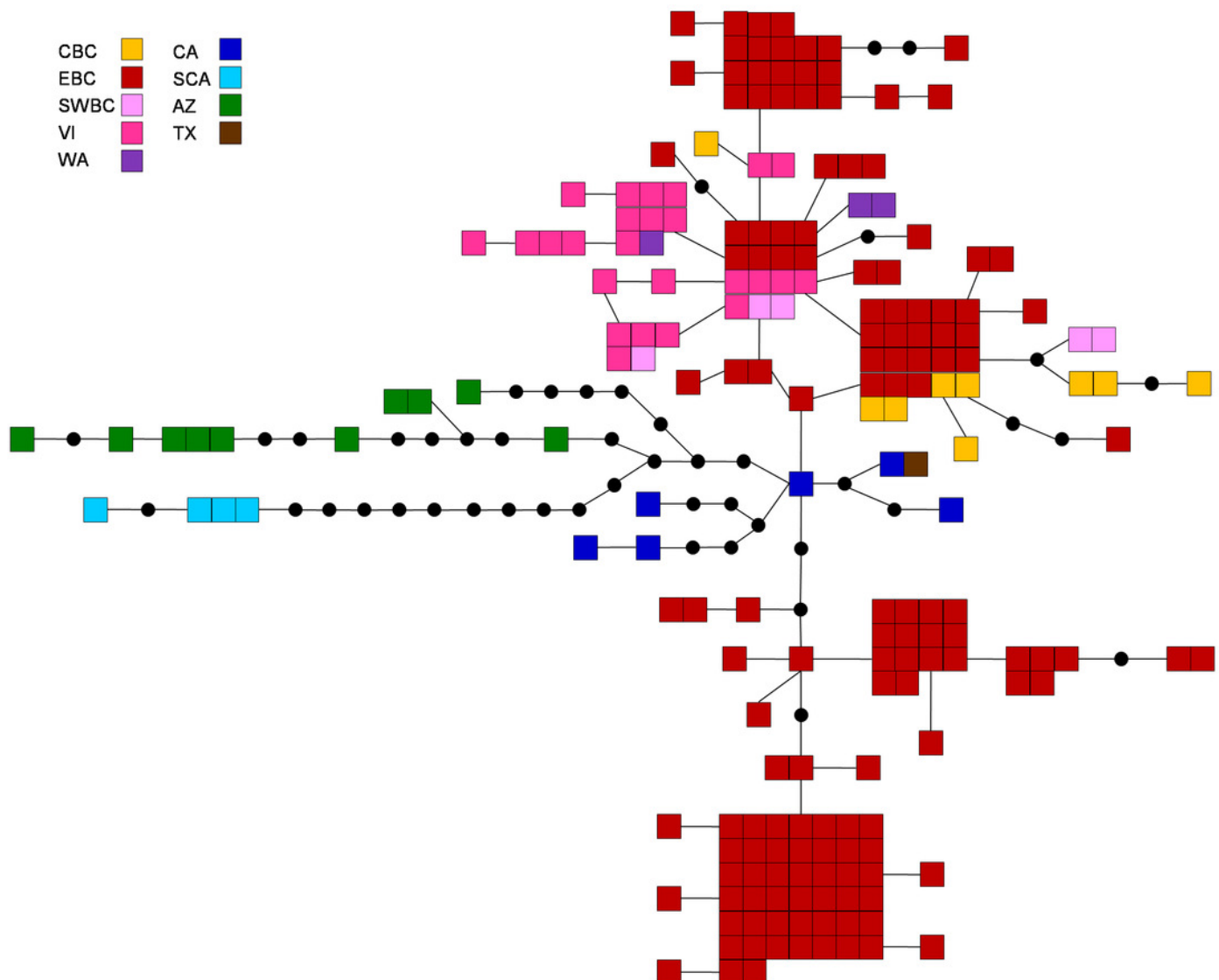
Figure 5

Statistical parsimony network for *M. disstria.*

Statistical parsimony network showing the relationship among 42 *M. disstria* haplotypes. Each square represents one of the 139 sequences colour-coded by location, inferred haplotypes are depicted by black circles, and each line represents a single nucleotide change. Refer to Fig. 1 for locations.