Phylogeographic structure in three North American tent caterpillar species (Lepidoptera: Lasiocampidae): *Malacosoma americana*, *M. californica*, and *M. disstria* (#22273)

First submission

Editor guidance

Please submit by **22 Dec 2017** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

7 Figure file(s)
4 Table file(s)

Structure and Criteria

2

Structure your review

The review form is divided into 5 sections.
Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

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. The present study begins to address this gap by examining population structure in members of the moth genus *Malacosoma*, an important group of forestry pests. The study begins by examining taxonomic boundaries among 474 specimens from five North American species based on sequence variation in the mitochondrial cytochrome *c* oxidase 1 (COI) gene. This work revealed that three of the species formed monophyletic lineages while *M. californica* and *M. incurva* were paraphyletic. The diversity of *M. californica* suggests a species complex, a conclusion supported by prior taxonomic work which assigned its members to six subspecies. Subsequent analysis examined the genetic structure of the three widespread 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 their 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. 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 than in the east.

ARTICLE TYPE: Original article

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

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

[1] *Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road E, Guelph, Ontario, N1G 2W1, Canada*

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

Running title: Phylogeography of three *Malacosoma*

ABSTRACT

While phylogeographic structure has been examined in many North American vertebrate species, insects have received much less attention. The present study begins to address this gap by examining population structure in members of the moth genus *Malacosoma*, an important group of forestry pests. The study begins by examining taxonomic boundaries among 474 specimens from five North American species based on sequence variation in the mitochondrial cytochrome *c* oxidase 1 (COI) gene. This work revealed that three of the species formed monophyletic lineages while *M. californica* and *M. incurva* were paraphyletic. The diversity of *M. californica* suggests a species complex, a conclusion supported by prior taxonomic work which assigned its members to six subspecies. Subsequent analysis examined the genetic structure of the three widespread 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 their 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. 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 than in the east.

INTRODUCTION

The patterns of genetic variation in species and the processes which underlie them are of particular interest to evolutionary biologists. Diverse factors, both historical and contemporary, influence how variation is distributed among populations; these include geological and climatic events, and the presence of physical and behavioural barriers (Avise, 2004). Past glaciations have had a major impact on the extent and patterning of genetic structure in Northern Hemisphere species (Hewitt, 2000). In North America, ice sheets covered much of present-day Canada and the northern United States, temperatures in ice-free areas were cooler than today, and sea levels dropped by up to 140 m (Pielou, 1991; Barendregt & Irving, 1998; Dyke *et al.*, 2002). The distributions of many species were fragmented with their populations persisting in small ice-free refugia (Hewitt, 1996; Hewitt, 2000; Stewart & Lister, 2001). In addition, physical and ecological barriers influenced genetic structure by their impacts on dispersal and gene flow. Recent environmental changes, both anthropogenic and natural, are now causing range shifts and population changes with varied impacts on genetic variation.

Comparisons of genetic variation spanning multiple species can identify both general and species-specific patterns, revealing how particular life history characteristics impact population structure. Species-specific traits, such as dispersal ability or niche requirements, may affect how a species responds to environmental, climatic, and geological changes. For example, the majority of northern species experienced major population declines reflecting the loss of habitat during the Pleistocene glaciations, while the interglacials favoured range expansion and population growth (Nilsson, 1983; Pielou, 1991; Hewitt, 2000; Hewitt, 2004). In contrast, cold-adapted species experienced habitat loss, often retreating to high altitude and high latitude locations during interglacials (Stewart & Lister, 2001; Dalén *et al.*, 2005; Galbreath *et al.*, 2009; Stewart *et*

al., 2009). Topographic features also have differing effects, with mountains and rivers restricting gene flow in some species while acting as dispersal corridors for others. For example, the Rocky Mountains prevent gene flow between some populations (Crease *et al.*, 1997; Burg *et al.*, 2005), but provide habitat as "sky islands" for others with the intervening lowlands limiting gene flow (Knowles, 2000; DeChaine & Martin, 2005; Galbreath *et al.*, 2009).

Past studies of phylogeographic structure in terrestrial organisms have largely examined vertebrates. Given their high diversity and abundance, phylogeographic patterns in insects have been understudied; past work has revealed diverse outcomes ranging from global panmixis (Alvial *et al.*, 2007; Correa *et al.*, 2017) to highly fragmented, structured populations (Dinca, Dapporto & Vila, 2011; Frantine-Silva *et al.*, 2017; Karthika *et al.*, 2017). Phylogeographic studies of Lepidoptera have employed both nuclear and mitochondrial markers, particularly the cytochrome *c* oxidase 1 (COI) locus (Vandewoestijne *et al.*, 2004; Craft *et al.,* 2010; Kirichenko *et al.,* 2017). This study represents the first step in a broad investigation of phylogeographic patterns in North American Lepidoptera. Here we examine three species in the genus *Malacosoma*, and describe the analytical framework that will be employed in later investigations on larger assemblages.

Malacosoma (Hübner, 1820) contains important forestry pests, species that experience cyclical outbreaks which often lead to extensive forest defoliation (Hildahl & Reeks, 1960; Stehr & Cook, 1968; Roland, 1993). Despite these impacts, there have been few studies of population structure in these moths. One study, which assessed allozyme variation in *Malacosoma americana* (Fabricius, 1793) from eastern United States, found limited variability and a lack of regional genetic differentiation (Costa & Ross, 1994). A second employed microsatellites and short DNA sequences to compare five populations of *Malacosoma californica pluvial* coastal

British Columbia (Franklin, Myers & Cory, 2014). Although high levels of variation were evident, there was little genetic differentiation between island and mainland populations.

This study employs sequence variation in a 658 bp segment of the mitochondrial COI gene to examine genetic structure in three widely distributed North American *Malacosoma* species: the eastern tent caterpillar *M. americana*, the western tent caterpillar *M. californica* (Packard, 1864), and the forest tent caterpillar *M. disstria* (Hübner, 1820; ~~Fig. 1~~). The taxonomic status of *M. californica* has been widely debated. It is currently viewed as including six largely allopatric subspecies, many of which have previously been viewed as distinct species. By contrast, both *M. americana* and *M. disstria* have no described subspecies (Stehr & Cook, 1968; Franclemont, 1973). All three species feed on diverse deciduous trees and shrubs, with host preferences varying by taxon and region, and a narrower host range in *M. americana* (Stehr & Cook, 1968; Franclemont, 1973; Parry & Goyer, 2004). While males are strong fliers, females are usually sedentary until they deposit their egg mass (Stehr & Cook, 1968; Franclemont, 1973), suggesting that the dispersal of mitochondrial markers will be low. While previous studies found ~~limited~~ genetic differentiation, they examined nuclear genes or very short mitochondrial sequences (<154 bp) at a relatively small geographic scale. By contrast, the present study assesses the genetic structure of a 658 bp mitochondrial gene region on a continental scale to ~~determine~~ if these *Malacosoma* species show concordant phylogeographic patterns.

MATERIALS AND METHODS

Species Relationships

A total of 474 COI sequences from specimens of five *Malacosoma* species collected in the United States and Canada were downloaded from the Barcode of Life Data System (BOLD;

see Supplemental Table S1; Ratnasingham & Hebert, 2007). These sequences were aligned in MEGA v6 (Tamura *et al.*, 2013) before a Bayesian network was constructed in BEAST v2.3 (Bouckaert *et al.*, 2014). The network was run with the HKY + Γ + I model for 10,000,000 MCMC steps, sampled every 2,000 steps, and had a 25% burn-in. Another lasiocampid, *Phyllodesma americana* (GenBank Accession Number JF842281), was used as the outgroup.

Genetic analysis

Haplotypes were assigned to each species manually and confirmed with TCS v1.21 (Clement, Posada & Crandall, 2000). Haplotype and nucleotide diversity indices were calculated in DNAsp v5.10 (Rozas *et al.*, 2003; Librado & Rozas, 2009), and an analysis of molecular variance (AMOVA; Excoffier, Smouse & Quattro, 1992) was performed in Arlequin v3.5.1.2 (Excoffier & Lischer, 2010). A spatial analysis of molecular variance (SAMOVA) was employed to detect genetic clusters (Dupanloup, Schneider & Excoffier, 2002). The SAMOVA was run for K = 2 to 6 with 1,000 iterations. Pairwise genetic differences (Φ_{ST}; 100,000 permutations) were calculated in Arlequin using K2P distances; a modified false discovery rate was applied to correct for multiple tests (Benjamini & Yekutieli, 2001). Nearby sampling locations were grouped to increase sample sizes. Isolation by distance was examined with a Mantel test as performed in Genepop v4.2 (10,000 permutations; Raymond & Rousset, 1995; Rousset, 2008); geographic distances were calculated with the Geographic Distance Matrix Calculator v1.2.3 (Ersts, 2015). Monmonier's algorithm, a 'splitting' analysis which uses Voronoi tessellation and Delauney triangulation, was employed to test for possible barriers to dispersal. The algorithm was implemented in Barrier v2.2 and used pairwise Φ_{ST} values as genetic distances (Manni, Guérard & Heyer, 2004).

A statistical parsimony network was constructed in TCS v1.2.1 to visualise the relationship among haplotypes. Cluster analyses were performed in Bayesian Analysis of Population Structure v5.2 (BAPS; Corander *et al.*, 2008) using the clustering with linked loci option (Corander & Tang, 2007). BAPS allows the assignment of individuals to genetic clusters with no *a priori* population information. A principal coordinates analysis (PCoA) was run in GenAlEx v6.5 (Peakall & Smouse, 2006; Peakall & Smouse, 2012) to identify genetic structure with no population constraints.

RESULTS

Species affinities and boundaries

The Bayesian network identified two main lineages within North American *Malacosoma*: the first group included specimens of *M. constricta* and *M. disstria* with each species forming a well-defined monophyletic ~~species;~~ the second included *M. americana*, *M. incurva*, and *M. californica* ~~(Fig. 2)~~. While *M. americana* was monophyletic, specimens assigned to *M. californica* were paraphyletic, suggesting a species complex. Of particular note, specimens of *M. californica* from Alberta, Saskatchewan, and New Brunswick showed closer similarity to those identified as *M. californica pluviale* than to specimens in the west (likely *M. californica californica*). Based on their geographic origin, these populations likely represent *M. californica lutescens*, the Great Plains tent caterpillar, or *M. californica pluviale* with the remaining samples corresponding to *M. californica californica* and one or more of the three southern subspecies (*M. californica ambisimile*, *M. californica fragilis*, and *M. californica recenseo*). To further evaluate relationships, a Bayesian tree was run as above in BEAST v2.3 on the 474 North American samples plus 45 samples from four European *Malacosoma* species (*M. neustria*, *M. alpicola*, *M.*

franconica, *M. castrensis*), and four *M. incurva* samples from Mexico (see Supplemental Table S1). This analyses also identified two groups: one including the European taxa with *M. constricta* and *M. disstria*, all forming well-defined species, and a second group that generally mirrored the results from North American analysis. Interestingly, the four Mexican samples of *M. incurva* grouped with the *M. californica* AB and SK samples (Supplemental Fig. S1). As further study with additional markers is required to establish the taxonomic status of lineages within the *M. californica* complex, the current analyses does not include *M. californica pluviale* or the AB, SK, and NB samples (see Fig. 1).

Cytochrome *c* oxidase I sequences

Subsequent analyses examined 79 COI sequences for *M. americana*, 207 for *M. californica*, and 139 for *M. disstria* (Supplemental Table S1). The 658 bp gene region was highly polymorphic with 37 variable sites defining 33 haplotypes in *M. americana*, 61 variable sites defining 64 haplotypes in *M. californica*, and 43 variable sites defining 42 haplotypes in *M. disstria* (Table 1). There were 8, 13, and 13 anticipated amino acid substitutions, no frameshift mutations, and no stop codons. Fixed nucleotide differences were present in *M. californica* between sCA, AZ, and the other populations of this species, and in *M. disstria* between western (BC, AB, and SK) and eastern groups. There were no fixed differences in *M. americana*. Haplotype and nucleotide diversities were high in all species (H_d = 0.918 - 0.926; π = 0.0049 - 0.0097; Table 1), with diversity generally higher in southern populations (Supplemental Table S2).

Genetic analyses

175 ***Malacosoma americana***

176 Of the three species, *M. americana* showed the least population structure (overall Φ_{ST} =
177 0.24, $p < 0.0001$), the fewest significant pairwise comparisons (7 out of 15; Table 2), and the
178 lowest diversity (Table 1). The greatest pairwise Φ_{ST} values were seen between the MN and ON
179 populations, and between NB and the other populations. Monmonier's analysis supported the
180 separation of MN and NB from the other populations (Fig. 1), while the SAMOVA analysis
181 identified four groups ($F_{CT} = 0.263$, $p = 0.02$): MN, NB, and the separation of the remaining
182 samples into northern (ON, MD/NC) and southern (TN, AR/OK/TX) populations. There was
183 evidence of significant isolation by distance ($R^2 = 0.3$, $p = 0.006$)

184 The statistical parsimony network showed a general clustering of southern samples (OK,
185 TX, AR, and TN) and NB samples (Fig. 3). The haplotypes were generally closely related,
186 although ON, OK, and MN had more divergent haplotypes. The principal coordinates analysis
187 allocated 55% of the variation to the first three coordinates (34.7%, 11.5%, and 8.8%,
188 respectively). The samples separated into northern and southern groups along coordinate 1 (Fig.
189 4a). Bayesian clustering analysis separated the samples into three clusters (Fig. 1, Supplemental
190 Table S2): a "northern" group found primarily in ON and NB, and two "southern" groups, a
191 larger group found in all populations except NB, and a smaller group primarily in MN and OK.
192 The distribution of haplotypes was significantly non-random ($X^2 = 47.65$, $p < 0.0001$).

193

194 ***Malacosoma californica***

195 *M. californica* had the strongest population structure (overall $\Phi_{ST} = 0.48$, $p < 0.0001$),
196 and the highest nucleotide diversity (Table 1). All pairwise comparisons were significant except
197 those with WA (17 out of 21; Table 2), and only a single comparison lacked significance before

198 correction (WA with swBC, p = 0.15). The greatest differences existed between sCA and all
199 other populations (Φ_{ST} = 0.73 - 0.94), and between AZ and all other populations (Φ_{ST} = 0.54 -
200 0.79). Both Monmonier's analysis and SAMOVA analysis identified genetic breaks between
201 three groups: sCA, AZ, and all other populations (F_{CT} = 0.57, p = 0.049; Fig. 1). The Mantel test
202 showed moderate and significant isolation by distance (R^2 = 0.15, p = 0.016), with the three
203 highest genetic distances (sCA with cBC, swBC, and WA) at large, but not the maximum,
204 geographic distances.

205 The statistical parsimony network showed a dumbbell-like pattern, with two large groups
206 containing most of the samples except those from CA, sCA, and AZ (Fig. 5). One group
207 consisted of haplotypes from across the Pacific Northwest, and contained all samples from cBC,
208 VI, swBC, and WA, as well as many eBC samples. There were several common haplotypes
209 separated by one to three mutations. The second cluster was restricted to eBC; it was less diverse
210 and contained two common haplotypes, one represented by 44 individuals. Most CA samples
211 had unique, fairly divergent haplotypes, while the sCA and AZ haplotypes were divergent with
212 13 - 15 (sCA) and 5 - 16 (AZ) mutations separating them from the nearest population (Fig. 5).

213 The principal coordinates analysis allocated 65.1% of the variation to the first three
214 coordinates (39.0%, 16.4%, and 9.7%, respectively). The BC samples separated into two groups
215 along coordinate 1, while the sCA and AZ samples were separated along coordinate 2 (Fig. 4b).
216 Bayesian clustering analysis identified six clusters (Fig. 1, Supplemental Table S2): one in sCA,
217 one in AZ, one in eBC, and three shared between multiple populations. The eBC population had
218 representatives in four clusters. The distribution of samples in the six clusters was highly
219 significant (X^2 = 486.6, p <0.0001). When sCA, CA, and AZ were removed from analysis, the
220 distribution was still significantly different than random (X^2 = 57.02, p < 0.0001).

221

222 ***Malacosoma disstria***

223 *M. disstria* also had significant population structure (overall $\Phi_{ST} = 0.42$, $p < 0.0001$).
224 Forty-two of 55 pairwise comparisons were significant following correction for multiple tests
225 (Table 2). The 13 comparisons lacking significance involved ON with NB/NS, ON with
226 NC/FL/GA, or were among the three ON locations. In contrast, the largest pairwise differences
227 were between BC and all other populations ($\Phi_{ST} = 0.49$ - 0.98). Diversity within populations was
228 generally high with $H_d > 0.7$ in 10 of the 16 populations. The lowest values were in SK ($H_d = 0$)
229 followed by eBC ($H_d = 0.20$; Supplemental Table S2). A similar pattern was seen with
230 nucleotide diversity with the highest values in the three ON populations ($\pi = 0.0053$ - 0.0065)
231 and TX ($\pi = 0.0137$). Monmonier's analysis identified a genetic break between the BC
232 populations and all other populations (including AB), possibly along the Rocky Mountains. This
233 was supported by the SAMOVA analysis ($F_{CT} = 0.52$, $p = 0.015$; Fig. 1). A second barrier was
234 identified on the Great Plains (east of the SK and MB populations; $F_{CT} = 0.41$, $p = 0.002$), but
235 this was not supported by SAMOVA. The Mantel test showed significant isolation by distance (p
236 $= 0.0001$), although the pattern was very weak ($R^2 < 0.0001$) due to very high genetic distances
237 between two sites in BC (eBC and cBC) and one in SK, likely a result of the monomorphic
238 nature of the SK population. When the comparison was run without SK the pattern was much
239 stronger ($R^2 = 0.29$, $p = 0.0005$).

240 The statistical parsimony network identified moderate variation, with five common
241 haplotypes ($n \geq 10$): three restricted to a single region (BC, AB/SK, and ON), one shared by ON
242 and NB/NS, and a third found across several regions (Fig. 6). The BC samples formed a separate
243 group (with a single AB sample), while AB, SK, and MB generally grouped together. In general,

northern and southern samples were separated, although there were exceptions. The principal coordinates analysis revealed similar structure with BC separating along coordinate 1 (41.5%) and a general separation of northern and southern populations along coordinate 2 (14.8%). Coordinate 3 explained 12.6% of the variation (Fig. 4c). Bayesian clustering analysis identified four clusters (Supplemental Table S2): three were restricted to the north while one was primarily found in the south. The BC samples clustered with some ON samples, AB and SK were in a cluster together (with four NB samples), and the remaining ON and NB/NS samples formed a third cluster. The allocation of samples to clusters was highly significant ($X^2 = 224.7$, $p < 0.0001$).

DISCUSSION

Variation among species

The analysis of mitochondrial COI sequences from three North American *Malacosoma* species showed high levels of variation and diversity, with some highly divergent populations. Despite considerable overlap in their distributions, particularly between *M. disstria* and the other two, patterns of variation and levels of population structure varied considerably.

The eastern species, *M. americana,* showed limited population structure consistent with a relatively young evolutionary history and/or high levels of gene flow. This pattern is common in species restricted to eastern North America, and in the eastern portion of the range for more widespread species. For example, many bird species found in the eastern half of North America exhibit limited genetic structure (Zink, Rootes & Dittmann, 1991; Vallianatos, Lougheed & Boag, 2001; Veit *et al.*, 2005), while widespread species show shallow divergence in the east (Klein & Brown, 1994; Graham & Burg, 2012; van Els, Cicero & Klicka, 2012), likely ~~reflecting~~

a single evolutionary origin and extensive contemporary gene flow. Other studies have identified limited structure in eastern trees (McLachlan, Clark & Manos, 2005; Shaw & Small, 2005; Gerardi *et al.*, 2010) and mammals (Petersen & Stewart, 2006), with the exception of more distinct southern populations (e.g., Texas or Florida). Both *M. americana* and *M. disstria* exhibit this pattern: limited structure in eastern North America with some differences between northern and southern populations. This is likely caused by colonisation from a single source population with limited ongoing gene flow. Give the limited dispersal capability of female *Malacosoma*, one would expect less gene flow than in highly-mobile bird species.

The western *M. californica* possessed a very different pattern with strong population structure and very distinct populations (AZ and sCA) which may represent different subspecies or ecotypes. Both the current and historical topography of this region can help explain these patterns. Western North America contains four major mountain ranges (Cascade, Coastal, Rocky, and Sierra) which run along a north-south axis, large plains, and deserts, all of which contribute to a complex habitat mosaic. This heterogeneity, coupled with the resulting complex glacial histories of the region, has resulted in extensive structuring in birds (Barrowclough *et al.*, 2004; Lait *et al.*, 2012; van Els, Cicero & Klicka, 2012), mammals (Byun, Koop & Reimchen, 1997; Riddle, Hafner & Alexander, 2000; Galbreath *et al.*, 2009), insects (Brown *et al.*, 1997), and plants (Richardson, Brunsfeld & Klopfenstein, 2002; Johansen & Latta, 2003). Interestingly, the western populations of *M. disstria* also show increased structure.

In addition to regional diversification, *M. disstria* exhibits a strong east/west separation, a pattern common in widespread North American species that often reflects multiple evolutionary origins (Sperling, Raske & Otvos, 1999; Gerardi *et al.*, 2010; Medina *et al.*, 2010; Lait & Burg, 2013). It is particularly frequent in species with low dispersal capability or strong natal

philopatry. The sustained separation of multiple genetic lineages, often with intervening zones of admixture, may reflect the presence of a conspecific population impeding the establishment of a new population, or by large areas of unsuitable habitat. Additional sampling from the central region could help to clarify if there is a distinct break, or whether secondary mixing is occurring.

Glacial refugia and recolonisation

All three species show evidence of persistence in one or more southern refugia with subsequent recolonisation of northern regions. Diversity patterns exhibited the characteristic "southern richness, northern purity" (Hewitt, 2004) found across much of the previously glaciated Northern Hemisphere. This was particularly evident in *M. americana* where diversity in southern populations was twice that in northern regions (π = 0.0046 - 0.0061 versus 0.0013 - 0.0027; Supplemental Table S2), and in *M. californica* where diversity was four-fold higher in Arizona (π = 0.0093) than in Washington (π = 0.002). In *M. disstria* diversity levels decreased to the northwest, likely representing a founder event in the BC populations with limited ongoing gene flow across the Rocky Mountains. All three species showed strong isolation by distance, indicating that recolonisation likely occurred in a stepping-stone fashion.

The fact that *M. americana* lacked fixed differences among populations, while *M. disstria* had only a few between the western and eastern populations, suggests a single refugial origin for both species, likely in the south-eastern United States. This pattern has been reported in both plant (Gerardi *et al.*, 2010) and animal (Zink, Rootes & Dittmann, 1991; Veit *et al.*, 2005; Petersen & Stewart, 2006) species across eastern North America. Interestingly, neither of these *Malacosoma* species have the genetic break between Atlantic and Gulf coast clades seen in many fish (Bermingham & Avise, 1986; Avise, 1992), insects (Vogler & Desalle, 1993; Ney & Schul,

2017), reptiles (Lamb & Avise, 1992), birds (Avise, 1992), and marine invertebrates (Herke & Foltz, 2002; Young *et al.*, 2002; see Soltis *et al.*, 2006 for additional references and an excellent description of this break).

As the BC populations are significantly different than those in the east, it remains possible that current populations of *M. disstria* derive from at least two refugia: one in the southeast and another in the west. There may have been multiple western refugia that were missed due to the lack of samples from the southwestern portion of the range; additional sampling is required to test this. This prospect is consistent with the pattern found in *Populus tremuloides*, its favoured host, that shows evidence of two genetic clusters, one in the southwest and one in the north and east, with higher diversity in the southwest group (Callahan *et al.*, 2013). Many other continent-wide species also possess a large group with low diversity across northern and eastern areas, and multiple diverse groups west of the Rocky Mountains (Ball & Avise, 1992; Byun, Koop & Reimchen, 1997; Graham & Burg, 2012; van Els, Cicero & Klicka, 2012), a pattern linked to a single refugium for the east and multiple isolated refugia in the west.

M. californica showed multiple fixed differences among sCA, AZ, and the other populations. This pattern, which is common in many southwestern and western species (Golden & Bain, 2000; Riddle, Hafner & Alexander, 2000; Galbreath *et al.*, 2009; Hamilton, Formanowicz & Bond, 2011), suggests isolation in multiple glacial refugia. Given the divergence of these two populations (0.9 - 1.9% divergence from the nearest population, CA), it is likely that they have been isolated for multiple glacial cycles with limited recent gene flow. The AZ population shows relatively high diversity, suggesting multiple refugia or impassable barriers within this small region. This is seen in a number of animal (Orange, Riddle & Nickle, 1999; Zink *et al.*, 2001; Merrill, Ramberg & Hagedorn, 2005; Graham *et al.*, 2013) and plant

(Frohlich *et al.*, 1999) species, and is likely due to the heterogeneous nature of Arizona which contains seven ecoregions (Warshall, 1995; Poulos, Taylor & Beaty, 2007; Ober & Connolly, 2015; Powell & Steidl, 2015).

Physical barriers

Several physical barriers impede gene flow in North American species. In eastern North America, the Appalachian Mountains act as a barrier to plants (Griffin & Barrett, 2004; Joly & Bruneau, 2004; Godbout *et al.*, 2005), reptiles (Bushar *et al.*, 2014; Krysko *et al.*, 2017), and amphibians (Church *et al.*, 2003; Jones *et al.*, 2006), while the Mississippi, Tombigbee, and Appalichola Rivers prevent gene flow in many plant and animal species (see Bermingham & Avise, 1986; Avise, 1992; Soltis *et al.*, 2006 and references therein). The two *Malacosoma* species in the east do not show genetic breaks along any of these traditional barriers. This may be due to recent colonisation from a single origin or ongoing gene flow in these regions. As the moths can fly, their dispersal capabilities should be greater than that of sedentary plant and reptile species, allowing them to cross rivers. The fact that the Appalachians have not prevented gene flow may indicate the importance of forested valleys as dispersal corridors. All three *Malacosoma* species are generalist herbivores, as they should encounter suitable habitat more often than specialists.

In western North America, the main physical barriers are the Rocky, Coastal, Cascade, and Sierra Nevada Mountains. These act as a barrier to gene flow in many species (Crease *et al.*, 1997; Nielson, Lohman & Sullivan, 2001; Johansen & Latta, 2003; Burg *et al.*, 2005; Carstens *et al.*, 2005). The Wyoming Basin and the Great Plains have also been shown to act as dispersal barriers, particularly for species associated with montane, forested, or wetland habitats

(DeChaine & Martin, 2005; Wilson *et al.*, 2005). *Malacosoma disstria* and *M. californica* both exhibit genetic breaks in western regions: *M. disstria* shows a clear break across the Rocky Mountains (between BC and AB), while *M. californica* has disjunct populations among many of the southwestern deserts. The Rocky Mountains may also act as a barrier in this species (or species complex); the samples identified as *M. californica* from AB and SK were very different than those in BC, likely representing a different species.

Conclusions

The population genetic structure of the three *Malacosoma* species suggest a single origin in the east and a complex evolutionary history in the west. *M. americana*, restricted to the eastern half of the continent, shows limited structure with a north-south trend and greater diversity in the south. This is consistent with its expansion from a single southern refugium following the last glaciation. *M. disstria* shows a similar pattern in the east, supporting a single southern refugium, with a second possible ~~origin in~~ the west. Additional samples are required to elucidate whether the differentiation in its BC population reflects a founder event or additional structuring. *M. californica* shows the greatest structure and differentiation, consistent with multiple evolutionary origins in the west and southwest. Further research into the taxonomic relationship within and among these species is required to determine where the *M. californica pluviale* and *M. californica* AB/SK samples fit within this species, or whether they actually represent cryptic species. This study shows the utility of existing DNA barcodes in identifying patterns of genetic structure in insect species, which can uncover previously unknown evolutionary histories and suggest further avenues to explore.

ACKNOWLEDGEMENTS

We would like to acknowledge all of the contributors to the Barcode of Life Database including the collections, laboratory, and bioinformatics staff at the Centre for Biodiversity Genomics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

All sequences are available on the Barcode of Life Database. Sampling locations and sequence data used in this study (BOLD IDs and available GenBank Accession Numbers) have been uploaded as online Supplemental Table S1.

TITLES AND LEGENDS TO TABLES AND FIGURES

Table 1. Sample size (n), number of 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 Φ_{ST} values were all significant ($p < 0.0001$).

Table 2. Population pairwise Φ_{ST} values for (a) *M. americana* ($P_{crit} = 0.015$), (b) *M. californica* ($P_{crit} = 0.014$), and (c) *M. disstria* ($P_{crit} = 0.011$). Φ_{ST} values are given below the diagonal and p-values above the diagonal. Values significant following correction for multiple tests are shaded. Refer to Fig. 1 for locations.

405 **Fig. 1.** Approximate distributions and sampling locations for (a) *Malacosoma americana*, (b) *M.*
406 *californica*, and (c) *M. disstria*. The dashed lines indicate genetic breaks identified by
407 Monmonier's algorithm and SAMOVA. The pie charts represent the distribution of BAPS
408 groups, scaled for sample size. The green crosses represent the omitted AB, SK, and NB *M.*
409 *californica* samples. Sampling locations are as follows: Alberta (AB), Arizona (AZ), Arkansas
410 (AR), British Columbia (BC; central [c], eastern [e], and southwest [sw]), California (CA;
411 southern [s]), Kentucky (KY), Maryland (MD), Minnesota (MN), New Brunswick (NB), North
412 Carolina (NC), Nova Scotia (NS), Oklahoma (OK), Ontario (ON; central [c], eastern [e], and
413 southern [s]), Saskatchewan (SK), Tennessee (TN), Texas, (TX), Vancouver Island BC (VI), and
414 Washington (WA). In (a) STH includes AR, OK, and TX. In (c) STH includes AR, KY, and OK,
415 and SE includes FL, GA, and NC.
416
417 **Fig. 2.** Bayesian analysis based on 474 COI sequences from five North American species of
418 *Malacosoma*. The triangles represent multiple specimens from the same species, and the length
419 of the triangle is representative of the sequence variation. Posterior values > 0.8 are given. The
420 shaded boxes represent the samples used in this study.
421
422 **Fig. 3.** Statistical parsimony network showing the relationship among the 33 *M. americana*
423 haplotypes. The 79 samples are colour-coded by location, and inferred haplotypes are depicted
424 by black circles. Refer to Fig. 1 for locations.
425
426 **Fig. 4.** Principle coordinates analysis for (a) *M. americana*, (b) *M. californica*, and (c) *M.*
427 *disstria*. Samples are colour-coded by sampling location. Refer to Fig. 1 for locations.

428

429 **Fig. 5.** Statistical parsimony network showing the relationship among the 64 *M. californica*
430 haplotypes. The 207 samples are colour-coded by location, and inferred haplotypes are depicted
431 by black circles. Refer to Fig. 1 for locations.

432

433 **Fig. 6.** Statistical parsimony network showing the relationship among 42 *M. disstria* haplotypes.
434 The 139 samples are colour-coded by location, and inferred haplotypes are depicted by black
435 circles. Refer to Fig. 1 for locations.

436

437 **SUPPLEMENTARY MATERIAL**

438 **Supplemental Table S1.** Sample information for nine *Malacosoma* species. The BOLD sample
439 ID, GenBank Accession Number, and sampling location are given. GPS coordinates for *M.*
440 *americana*, *M. californica*, and *M. disstria* are given where available.

441

442 **Supplemental Table S2.** Sample size (n) and distribution of haplotypes (h), haplotype diversity
443 (Hd), and nucleotide diversity (π) among the (a) *M. americana*, (b) *M. californica*, and (c) *M.*
444 *disstria* sampling locations. Shared haplotypes have been allocated a letter. The allocation of
445 samples into Bayesian clusters (BAPS) is also given.

446

447 **Supplementary Fig. 1.** Bayesian analysis of nine *Malacosoma* species from North America and
448 Europe. *Phyllodesma americana* is used as outgroup. Posterior values are given.

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 *et al.* (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.

Burg TM, Gaston AJ, Winker K, Friesen VL (2005). Rapid divergence and postglacial colonization in western North American Steller's jays (*Cyanocitta stelleri*). *Molecular Ecology* **14:** 3745-3755. DOI: 10.1111/j.1365-294X.2005.02710.x.

Bushar LM, Aborde CCB, Gao SS, Gonzalez MV, Hoffman JA, Massaro IK *et al.* (2014). Genetic structure of timber rattlesnake (*Crotalus horridus*) populations: physiographic influences and conservation implications. *Copeia* **2014:** 694-706. DOI: 10.1643/CE-14-047.

Byun SA, Koop BF, Reimchen TE (1997). North American black bear mtDNA phylogeography: Implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution* **51:** 1647-1653. DOI: 10.2307/2411216.

Callahan CM, Rowe CA, Ryel RJ, Shaw JD, Madritch MD, Mock KE (2013). Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*). *Journal of Biogeography* **40:** 1780-1791. DOI: 10.1111/jbi.12115.

Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J (2005). Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing

493 within a comparative phylogeographic framework. *Evolution* **59:** 1639-1652. DOI:
494 10.1554/04-661.1.
495 Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003). Evidence for multiple
496 Pleistocene refugia in the postglacial expansion of the eastern tiger salamander,
497 *Ambystoma tigrinum tigrinum*. *Evolution* **57:** 372-383. DOI: 10.1111/j.0014-3820-
498 2003.tb00271.x.
499 Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene
500 genealogies. *Molecular Ecology* **9:** 1657-1659. DOI: 10.1046/j.1365-294x.2000.01020.x.
501 Corander J, Marttinen P, Siren J, Tang J (2008). Enhanced Bayesian modelling in BAPS
502 software for learning genetic structures of populations. *BMC Bioinformatics* **9:** 539. DOI:
503 10.1186/1471-2105-9-539.
504 Corander J, Tang J (2007). Bayesian analysis of population structure based on linked molecular
505 information. *Mathematical Biosciences* **205:** 19-31. DOI: 10.1016/j.mbs.2006.09.015.
506 Correa A, Vinson C, Braga L, Guedes R, de Oliveira L (2017). Ancient origin and recent range
507 expansion of the maize weevil *Sitophilus zeamais*, and its genealogical relationship to the
508 rice weevil *S. oryzae*. *Bulletin of Entomological Research* **107:** 9-20. DOI:
509 10.1017/S0007485316000687.
510 Costa JT, Ross KG (1994). Hierarchical genetic structure and gene flow in macrogeographic
511 populations of the eastern tent caterpillar (*Malacosoma americanum*). *Evolution* **48:**
512 1158-1167. DOI: 10.1111/j.1558-5646.1994.tb05302.x.
513 Craft KJ, Pauls SU, Darrow K, Miller SE, Hebert PDN, Helgen LE *et al.* (2010). Population
514 genetics of ecological communities with DNA barcodes: an example from New Guinea

515 Lepidoptera. *Proceedings of the National Academy of Sciences USA* **107:** 5041-5046.
516 DOI: 10.1073/pnas.0913084107.
517 Crease TJ, Lee S-K, Yu S-L, Spitze K, Lehman N, Lynch M (1997). Allozyme and mtDNA
518 variation in populations of the *Daphnia pulex* complex from both sides of the Rocky
519 Mountains. *Heredity* **79:** 242-251. DOI: 10.1038/hdy.1997.151.
520 Dalén L, Fuglei E, Hersteinsson P, Kapel C, Roth J, Samelius G *et al.* (2005). Population history
521 and genetic structure of a circumpolar species: the arctic fox. *Biological Journal of the*
522 *Linnean Society* **84:** 79-89. DOI: 10.1111/j.1095-8312.2005.00415.x.
523 DeChaine EG, Martin AP (2005). Historical biogeography of two alpine butterflies in the Rocky
524 Mountains: broad-scale concordance and local-scale discordance. *Journal of*
525 *Biogeography* **32:** 1943-1956. DOI: 10.1111/j.1365-2699.2005.01356.x.
526 Dinca V, Dapporto L, Vila R (2011). A combined genetic-morphometric analysis unravels the
527 complex biogeographical history of *Polyommatus icarus* and *Polyommatus celina*
528 common blue butterflies. *Molecular Ecology* **20:** 3921-3935. DOI: 10.1111/j.1365-
529 294X.2011.05223.x.
530 Dupanloup I, Schneider S, Excoffier L (2002). A simulated annealing approach to define the
531 genetic structure of populations. *Molecular Ecology* **11:** 2571-2581. DOI:
532 10.1046/j.1365-294X.2002.01650.x.
533 Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J *et al.* (2002). The Laurentide
534 and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews*
535 **21:** 9-31. DOI: 10.1016/S0277-3791(01)00095-6.
536 Ersts PJ (2015). American Museum of Natural History: Center for Biodiversity and
537 Conservation.

538 Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform
539 population genetics analyses under Linux and Windows. *Molecular Ecology Resources*
540 **10:** 564-567. DOI: 10.1111/j.1755-0998.2010.02847.x.
541 Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from
542 metric distances among DNA haplotypes: application to human mitochondrial DNA
543 restriction data. *Genetics* **131:** 479-491.
544 Franclemont JG (1973). *Mimallonoidea: Mimallonidae and Bombycoidea: Apatelodidae,*
545 *Bombycidae, Lasiocampidae. The Moths of North America North of Mexico, Fascicle*
546 *20.1*. The Wedge Entomological Research Foundation.
547 Franklin MT, Myers JH, Cory JS (2014). Genetic similarity of island populations of tent
548 caterpillars during successive outbreaks. *PLOS One* **9:** e96679. DOI:
549 10.1371/journal.pone.0096679.
550 Frantine-Silva W, Giangarelli DC, Penha RES, Suzuki KM, Dec E, Gaglianone MC *et al.*
551 (2017). Phylogeography and historical demography of the orchid bee *Euglossa iopoecila*:
552 signs of vicariant events associated to Quaternary climatic changes. *Conservation*
553 *Genetics* **18:** 539-552. DOI: 10.1007/s10592-016-0905-7.
554 Frohlich DR, Torres-Jerez I, Bedford ID, Markham PG, Brown JK (1999). A phylogeographical
555 analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers.
556 *Molecular Ecology* **8:** 1683-1691.
557 Galbreath KE, Hafner DJ, Zamudio KR, Agnew K (2009). Isolation and introgression in the
558 intermountain west: Contrasting gene genealogies reveal the complex biogeographic
559 history of the American pika (*Ochotona princeps*). *Journal of Biogeography* **37:** 344-
560 362. DOI: 10.1111/j.1365-2699.2009.02201.x.

561 Gerardi S, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2010). From glacial refugia to modern
562 populations: new assemblages of organelle genomes generated by differential
563 cytoplasmic gene flow in transcontinental black spruce. *Molecular Ecology* **19:** 5265-
564 5280. DO: 10.1111/j.1365-294X.2010.04881.x.
565 Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005). A mitochondrial DNA
566 minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range
567 North American conifer. *Molecular Ecology* **14:** 3497-3512. DOI: 10.1111/j.1365-
568 294X.2005.02674.x.
569 Golden JL, Bain JF (2000). Phylogeographic patterns and high levels of chloroplast DNA
570 diversity in four *Packera* (Asteraceae) species in southwestern Alberta. *Evolution* **54:**
571 1566-1579. DOI: 10.1111/j.0014-3820.2000.tb00702.x.
572 Graham BA, Burg TM (2012). Molecular markers provide insight into contemporary and historic
573 gene flow for a non-migratory species. *Journal of Avian Biology* **43:** 198-214. DOI:
574 10.1111/j.1600-048X.2012.05604.x.
575 Graham MR, Jaeger JR, Prendini L, Riddle BR (2013). Phylogeography of the Arizona hairy
576 scorpion (Hadrurus arizonensis) supports a model of biotic assembly in the Mojave
577 Desert and adds a new Pleistocene refugium. *Journal of Biogeography* **40:** 1298-1312.
578 DOI: 10.1111/jbi.12079.
579 Griffin SR, Barrett SCH (2004). Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in
580 eastern North America: inferences from phylogeography. *American Journal of Botany*
581 **91:** 465-473. DOI: 10.3732/ajb.91.3.465.

582 Hamilton CA, Formanowicz DR, Bond JE (2011). Species delimitation and phylogeography of
583 *Aphonopelma hentzi* (Araneae, Mygalomorphae, Theraphosidae): cryptic diversity in
584 North American tarantulas. *PLOS One* **6:** e26207. DOI: 10.1371/journal.pone.0026207.
585 Herke SW, Foltz DW (2002). Phylogeography of two squid (*Loligo pealei* and *L. plei*) in the
586 Gulf of Mexico and northwest Atlantic Ocean. *Marine Biology* **140:** 103-115. DOI:
587 10.1007/s002270100680.
588 Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and
589 speciation. *Biological Journal of the Linnean Society* **58:** 247-276. DOI: 10.1111/j.1095-
590 8312.1996.tb01434.x.
591 Hewitt GM (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907-913. DOI:
592 10.1038/35016000.
593 Hewitt GM (2004). Genetic consequences of climatic oscillations in the Quaternary.
594 *Philosophical Transactions of the Royal Society B: Biological Sciences* **359:** 183-195.
595 DOI: 10.1098/rstb.2003.1388.
596 Hildahl V, Reeks WA (1960). Outbreaks of the forest tent caterpillar, *Malacosoma disstria* Hbn.,
597 and their effects on stands of trembling aspen in Manitoba and Saskatchewan. *The*
598 *Canadian Entomologist* **92:** 199-209. DOI: 10.4039/Ent92199-3.
599 Johansen AD, Latta RG (2003). Mitochondrial haplotype distribution, seed dispersal and patterns
600 of postglacial expansion of ponderosa pine. *Molecular Ecology* **12:** 293-298. DOI:
601 10.1046/j.1365-294X.2003.01723.x.
602 Joly S, Bruneau A (2004). Evolution of triploidy in *Apios americana* (Leguminosae) revealed by
603 genealogical analysis of the histone H3-D gene. *Evolution* **58:** 284-295. DOI:
604 10.1111/j.0014-3820.2004.tb01645.x.

605 Jones MT, Voss SR, Ptacek MB, Weisrock DW, Tonkyn DW (2006). River drainages and
606 phylogeography: an evolutionary significant lineage of shovel-nosed salamander
607 (*Desmognathus marmoratus*) in the southern Appalachians. *Molecular Phylogenetics and*
608 *Evolution* **38:** 280-287. DOI: 10.1016/j.ympev.2005.05.007.
609 Karthika P, Vadivalagan C, Krishnaveni N, Murugan K, Nicoletti M, Canale A *et al.* (2017).
610 Contrasting genetic diversity and intra-population polymorphism of the invasive pest
611 *Henosepilachna vigintioctopunctata* (Coleoptera, Coccinellidae): a DNA barcoding
612 approach. *Journal of Asia-Pacific Entomology* **20:** 23-29. DOI:
613 10.1016/j.aspen.2016.11.011.
614 Kirichenko N, Triberti P, Ohshima I, Haran J, Byun B-K, Li H *et al.* (2017) From east to west
615 across the Palearctic: phylogeography of the invasive lime leaf miner *Phyllonorycter*
616 *issikii* (Lepidoptera: Gracillariidae) and discovery of a putative new cryptic species in
617 East Asia. *PLOS One* **12:** *e0171104.* DOI: 10.1371/journal.pone.0171104.
618 Klein NK, Brown WM (1994). Intraspecific molecular phylogeny in the yellow warbler
619 (*Dendroica petechia*), and implications for avian biogeography in the West Indies.
620 *Evolution* **48:** 1914-1932. DOI: 10.1111/j.1558-5646.1994.tb02223.x.
621 Knowles L (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*)
622 from the sky islands of western North America. *Evolution* **54:** 1337-1348. DOI:
623 10.1111/j.0014-3820.2000.tb00566.x.
624 Krysko KL, Nunez LP, Newman CE, Bowen BW (2017). Phylogenetics of kingsnakes,
625 *Lampropeltis getula* complex (Serpentes: Colubridae), in eastern North America. *Journal*
626 *of Heredity* **108:** 226-238. DOI: 10.1093/jhered/esw086.

627 Lait LA, Burg TM (2013). When east meets west: population structure of a high-latitude resident
628 species, the boreal chickadee (*Poecile hudsonicus*). *Heredity* **111:** 321-329. DOI:
629 10.1038/hdy.2013.54.
630 Lait LA, Friesen VL, Gaston AJ, Burg TM (2012). The post-Pleistocene population genetic
631 structure of a western North American passerine: the chestnut-backed chickadee (*Poecile*
632 *rufescens*). *Journal of Avian Biology* **43:** 541-552. DOI: 10.1111/j.1600-
633 048X.2012.05761.x.
634 Lamb T, Avise JC (1992). Molecular and population genetic aspects of mitochondrial DNA
635 variability in the diamondback terrapin, *Malaclemys terrapin*. *Journal of Heredity* **83:**
636 262-269. DOI: 10.1093/oxfordjournals.jhered.a111211.
637 Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA
638 polymorphism data. *Bioinformatics* **25:** 1451-1452. DOI: 10.1093/bioinformatics/btp187.
639 Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic)
640 variation: how barriers can be detected by "Monmonier's algorithm". *Human Biology* **76:**
641 173-190.
642 McLachlan JS, Clark JS, Manos PS (2005). Molecular indicators of tree migration capacity
643 under rapid climate change. *Ecology* **86:** 2088-2098. DOI: 10.1890/04-1036.
644 Medina RF, Rondon SI, Reyna SM, Dickey AM (2010). Population structure of *Phthorimaea*
645 *operculella* (Lepidoptera: Gelechiidae) in the United States. *Environmental Entomology*
646 **39:** 1037-1042. DOI: 10.1603/en09286.
647 Merrill SA, Ramberg FB, Hagedorn HH (2005). Phylogeography and population structure of
648 *Aedes aegypti* in Arizona. *The American Journal of Tropical Medicine and Hygiene* **72:**
649 304-310.

650 Ney G, Schul J (2017). Population structure within the one-dimensional range of a coastal plain
651 katydid. *PLOS One* **12**. DOI: 10.1371/journal.pone.0179361.
652 Nielson M, Lohman K, Sullivan J (2001). Phylogeography of the tailed frog (*Ascaphus truei*):
653 implications for the biogeography of the Pacific Northwest. *Evolution* **55:** 147-160. DOI:
654 10.1111/j.0014-3820.2001.tb01280.x.
655 Nilsson T (1983). *The Pleistocene: Geology and Life in the Quaternary Ice Age*. D. Reidel
656 Publishing Company: Dordrecht, Holland.
657 Ober KA, Connolly CT (2015). Geometric morphometric and phylogenetic analyses of Arizona
658 Sky Island populations of *Scaphinotus petersi* Roeschke (Coleoptera: Carabidae).
659 *Zoological Journal of the Linnean Society* **175:** 107-118. DOI: 10.1111/zoj.12269.
660 Orange DI, Riddle BR, Nickle DC (1999). Phylogeography of a wide-ranging desert lizard,
661 *Gambelia wislizenii* (Crotaphytidae). *Copeia* **1999:** 267-273. DOI: 10.2307/1447471.
662 Parry D, Goyer RA (2004). Variation in the suitability of host tree species for geographically
663 discrete populations of forest tent caterpillar. *Environmental Entomology* **33:** 1477-1487.
664 DOI: 10.1603/0046-225X-33.5.1477.
665 Peakall ROD, Smouse PE (2006). GenAlEx 6: Genetic Analysis in Excel. Population genetic
666 software for teaching and research. *Molecular Ecology Notes* **6:** 288-295. DOI:
667 10.1111/j.1471-8286.2005.01155.x.
668 Peakall ROD, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic
669 software for teaching and research - an update. *Bioinformatics* **28:** 2537-2539. DOI:
670 10.1093/bioinformatics/bts460.

671 Petersen SD, Stewart DT (2006). Phylogeography and conservation genetics of southern flying
672 squirrels (*Glaucomys volans*) from Nova Scotia. *Journal of Mammalogy* **87:** 153-160.
673 DOI: 10.1644/05-MAMM-A-062R1.1.
674 Pielou EC (1991). *After the Ice Age: The Return of Life to Glaciated North America*. University
675 of Chicago Press: Chicago, IL.
676 Poulos HM, Taylor AH, Beaty RM (2007). Environmental controls on dominance and diversity
677 of woody plant species in a Madrean, sky island ecosystem, Arizona, USA. *Plant*
678 *Ecology* **193:** 15-30. DOI: 10.1007/s11258-006-9245-x.
679 Powell BF, Steidl RJ (2015). Influence of vegetation on montane riparian bird communities in
680 the sky islands of Arizona, USA. *Southwestern Naturalist* **60:** 65-71. DOI: 10.1894/mcg-
681 09.1.
682 Ratnasingham S, Hebert PDN (2007). BOLD: The Barcode of Life Data System
683 (www.barcodinglife.org). *Molecular Ecology Notes* **7:** 355-364. DOI: 10.1111/j.1471-
684 8286.2006.01678.x.
685 Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact
686 tests and ecumenicism. *Journal of Heredity* **86:** 248-249. DOI:
687 10.1093/oxfordjournals.jhered.a111573.
688 Richardson BA, Brunsfeld SJ, Klopfenstein NB (2002). DNA from bird-dispersed seed and
689 wind-disseminated pollen provides insights into postglacial colonization and population
690 genetic structure of whitebark pine (*Pinus albicaulis*). *Molecular Ecology* **11:** 215-227.
691 Riddle BR, Hafner DJ, Alexander LF (2000). Comparative phylogeography of Baileys' pocket
692 mouse (*Chaetodipus*) and the *Peromyscus eremicus* species group: historical vicariance

693 of the Baja California peninsular desert. *Molecular Phylogenetics and Evolution* **17:** 161-
694 172. DOI: 10.1006/mpev.2000.0842.
695 Roland J (1993). Large-scale forest fragmentation increases the duration of tent caterpillar
696 outbreak. *Oecologia* **93:** 25-30. DOI: 10.1007/BF00321186.
697 Rousset F (2008). Genepop'007: A complete reimplementation of the Genepop software for
698 Windows and Linux. *Molecular Ecology Resources* **8:** 103-106. DOI: 10.1111/j.1471-
699 8286.2007.01931.x.
700 Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism
701 analyses by the coalescent and other methods. *Bioinformatics* **19:** 2496-2497. DOI:
702 10.1093/bioinformatics/btg359.
703 Shaw J, Small RL (2005). Chloroplast DNA phylogeny and phylogeography of the North
704 American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American*
705 *Journal of Botany* **92:** 2011-2030. DOI: 10.3732/ajb.92.12.2011.
706 Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006). Comparative
707 phylogeography of unglaciated eastern North America. *Molecular Ecology* **15:** 4261-
708 4293. DOI: 10.1111/j.1365-294X.2006.03061.x.
709 Sperling FAH, Raske AG, Otvos IS (1999). Mitochondrial DNA sequence variation among
710 populations and host races of *Lambdina fiscellaria* (Gn.) (Lepidoptera : Geometridae).
711 *Insect Molecular Biology* **8:** 97-106. DOI: 10.1046/j.1365-2583.1999.810097.x.
712 Stehr FW, Cook EF (1968). *A revision of the genus Malacosoma Hübner in North America*
713 *(Lepidoptera: Lasiocampidae): systematics, biology, immatures, and parasites*, Vol 276.
714 Smithsonian Institution Press: Washington D.C.

715 Stewart JR, Lister AM (2001). Cryptic northern refugia and the origins of the modern biota.
716 *Trends in Ecology & Evolution* **16:** 608-613. DOI: 10.1016/S0169-5347(01)02338-2.
717 Stewart JR, Lister AM, Barnes I, Dálen L (2009). Refugia revisited: individualistic responses of
718 species in space and time. *Proceedings of the Royal Society B: Biological Sciences* **277:**
719 661-671. DOI: 10.1098/rspb.2009.1272.
720 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular
721 Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30:** 2725-
722 2729. DOI: 10.1093/molbev/mst197.
723 Vallianatos M, Lougheed SC, Boag PT (2001). Phylogeography and genetic characteristics of a
724 putative secondary-contact zone of the loggerhead shrike in central and eastern North
725 America. *Canadian Journal of Zoology* **79:** 2221-2227. DOI: 10.1139/z01-157.
726 Vandewoestijne S, Baguette M, Brakefield PM, Saccheri IJ (2004). Phylogeography of *Aglais*
727 *urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and
728 control region. *Molecular Phylogenetics and Evolution* **31:** 630-646. DOI:
729 10.1016/j.ympev.2003.09.007.
730 van Els P, Cicero C, Klicka J (2012). High latitudes and high genetic diversity: phylogeography
731 of a widespread boreal bird, the gray jay (*Perisoreus canadensis*). *Molecular*
732 *Phylogenetics and Evolution* **63:** 456-465. DOI: 10.1016/j.ympev.2012.01.019.
733 Veit ML, Robertson RJ, Hamel PB, Friesen VL (2005). Population genetic structure and
734 dispersal across a fragmented landscape in cerulean warblers (*Dendroica cerulea*).
735 *Conservation Genetics* **6:** 159-174.

736 Vogler AP, Desalle R (1993). Phylogeographic patterns in coastal North American tiger beetles
737 (*Cicindela dorsalis* Say) inferred from mitochondrial DNA sequences. *Evolution* **47:**
738 1192-1202. DOI: 10.1111/j.1558-5646.1993.tb02146.x.
739 Warshall P (1995). *The Madrean sky island archipelago: a planetary overview*, Vol 264.
740 Wilson GM, Den Bussche RA, McBee K, Johnson LA, Jones CA (2005). Intraspecific
741 phylogeography of red squirrels (*Tamiasciurus hudsonicus*) in the central Rocky
742 Mountain region of North America. *Genetica* **125:** 141-154. DOI: 10.1007/s10709-005-
743 5154-5.
744 Young AM, Torres C, Mack JE, Cunningham CW (2002). Morphological and genetic evidence
745 for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit
746 crab *Pagurus longicarpus*. *Marine Biology* **140:** 1059-1066.
747 Zink RM, Kessen AE, Line TV, Blackwell-Rago RC (2001). Comparative phylogeography of
748 some aridland bird species. *The Condor* **103:** 1-10. DOI: 10.1650/0010-
749 5422(2001)103[0001:CPOSAB]2.0.CO;2.
750 Zink RM, Rootes WL, Dittmann DL (1991). Mitochondrial DNA variation, population structure,
751 and evolution of the common grackle (*Quiscalus quiscula*). *The Condor* **93:** 318-329.
752 DOI: 10.2307/1368947.

Table 1(on next page)

Sample details for three *Malacosoma* species.

Sample size (n), number of 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}) are given. The three Φ_{ST} values were all 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	0.478
M. disstria	139	19	43	0.628 ± 0.383	42	0.926	0.0063	0.419

1

2

Table 2 (on next page)

Population pairwise Φ_{ST} values for (a) *M. americana* (P_{crit} = 0.015), (b) *M. californica* (P_{crit} = 0.014), and (c) *M. disstria* (P_{crit} = 0.011).

Φ_{ST} values are given below the diagonal and p-values above the diagonal. Values significant following correction for multiple tests are shaded. Refer to Fig. 1 for locations.

(a)	NB	ON	STH	MN	MD/NC	TN
NB	*	**0.001**	**0.000**	**0.002**	**0.007**	**0.001**
ON	0.195	*	**0.000**	**0.001**	0.702	0.027
STH	0.444	0.213	*	0.019	0.321	0.376
MN	0.637	0.395	0.271	*	0.298	0.257
MD/NC	0.481	0.000	0.015	0.326	*	0.484
TN	0.564	0.175	0.007	0.202	0.000	*

STH = AR, OK, TX

(b)	CBC	swBC	eBC	AZ	sCA	CA	WA
CBC	*	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	0.014
swBC	0.329	*	**0.000**	**0.000**	**0.000**	**0.000**	0.154
eBC	0.249	0.303	*	**0.000**	**0.000**	**0.003**	0.018
AZ	0.682	0.785	0.618	*	**0.001**	**0.000**	**0.004**
sCA	0.892	0.905	0.725	0.732	*	**0.004**	0.029
CA	0.448	0.612	0.241	0.544	0.799	*	**0.012**
WA	0.352	0.099	0.258	0.662	0.935	0.500	*

(c)	AB	eBC	cBC	NB/NS	eON	sON	cON	SK	STH	SE	TN
AB	*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.284	**0.000**	**0.001**	**0.000**
eBC	0.719	*	0.300	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
cBC	0.677	0.044	*	**0.000**	**0.000**	**0.000**	**0.003**	**0.000**	**0.000**	**0.002**	**0.000**
NB/NS	0.284	0.808	0.769	*	0.021	0.110	0.026	**0.000**	**0.000**	**0.005**	**0.000**
eON	0.294	0.530	0.486	0.134	*	0.265	0.896	**0.002**	**0.000**	0.040	**0.001**
sON	0.255	0.660	0.617	0.037	0.012	*	0.317	**0.000**	**0.000**	0.049	**0.000**
cON	0.357	0.626	0.541	0.201	0.000	0.016	*	**0.000**	**0.000**	0.055	**0.000**
SK	0.026	0.978	0.948	0.468	0.367	0.362	0.494	*	**0.000**	**0.001**	**0.000**
STH	0.391	0.701	0.644	0.315	0.245	0.260	0.316	0.523	*	0.073	**0.000**
SE	0.349	0.890	0.824	0.246	0.188	0.155	0.272	0.781	0.111	*	0.015
TN	0.473	0.912	0.873	0.407	0.313	0.300	0.440	0.835	0.286	0.317	*

STH = AR, KY, OK; SE = NC, GA, FL

Figure 1

Approximate distributions and sampling locations for (a) *Malacosoma americana*, (b) *M. californica*, and (c) *M. disstria*.

The dashed lines indicate genetic breaks identified by Monmonier's algorithm and SAMOVA. The pie charts represent the distribution of BAPS groups, scaled for sample size. The green crosses represent the omitted AB, SK, and NB *M. californica* samples. Sampling locations are as follows: Alberta (AB), Arizona (AZ), Arkansas (AR), 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), Oklahoma (OK), Ontario (ON; central [c], eastern [e], and southern [s]), Saskatchewan (SK), Tennessee (TN), Texas, (TX), Vancouver Island BC (VI), and Washington (WA). In (a) STH includes AR, OK, and TX. In (c) STH includes AR, KY, and OK, and SE includes FL, GA, and NC.

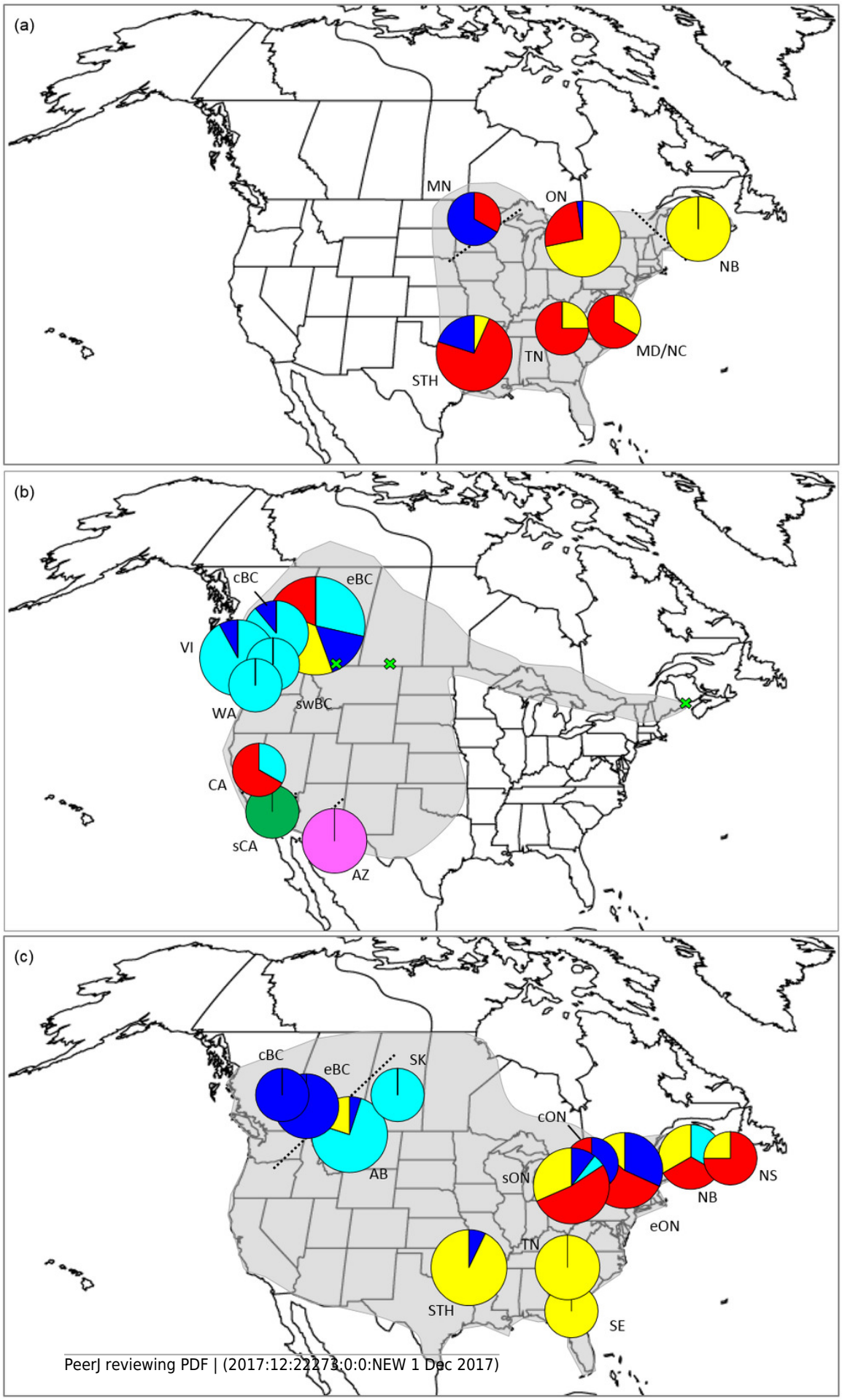

(a)
MN
ON
NB
TN
MD/NC
STH
(b)
cBC
eBC
VI
WA
swBC
CA
sCA
AZ
(c)
cBC
eBC
SK
AB
cON
sON
NB
NS
eON
TN
STH
SE

Figure 2

Bayesian analysis based on 474 COI sequences from five North American species of *Malacosoma*.

The triangles represent multiple specimens from the same species, and the length of the triangle is representative of the sequence variation. Posterior values > 0.8 are given. The shaded boxes represent the samples used in this study.

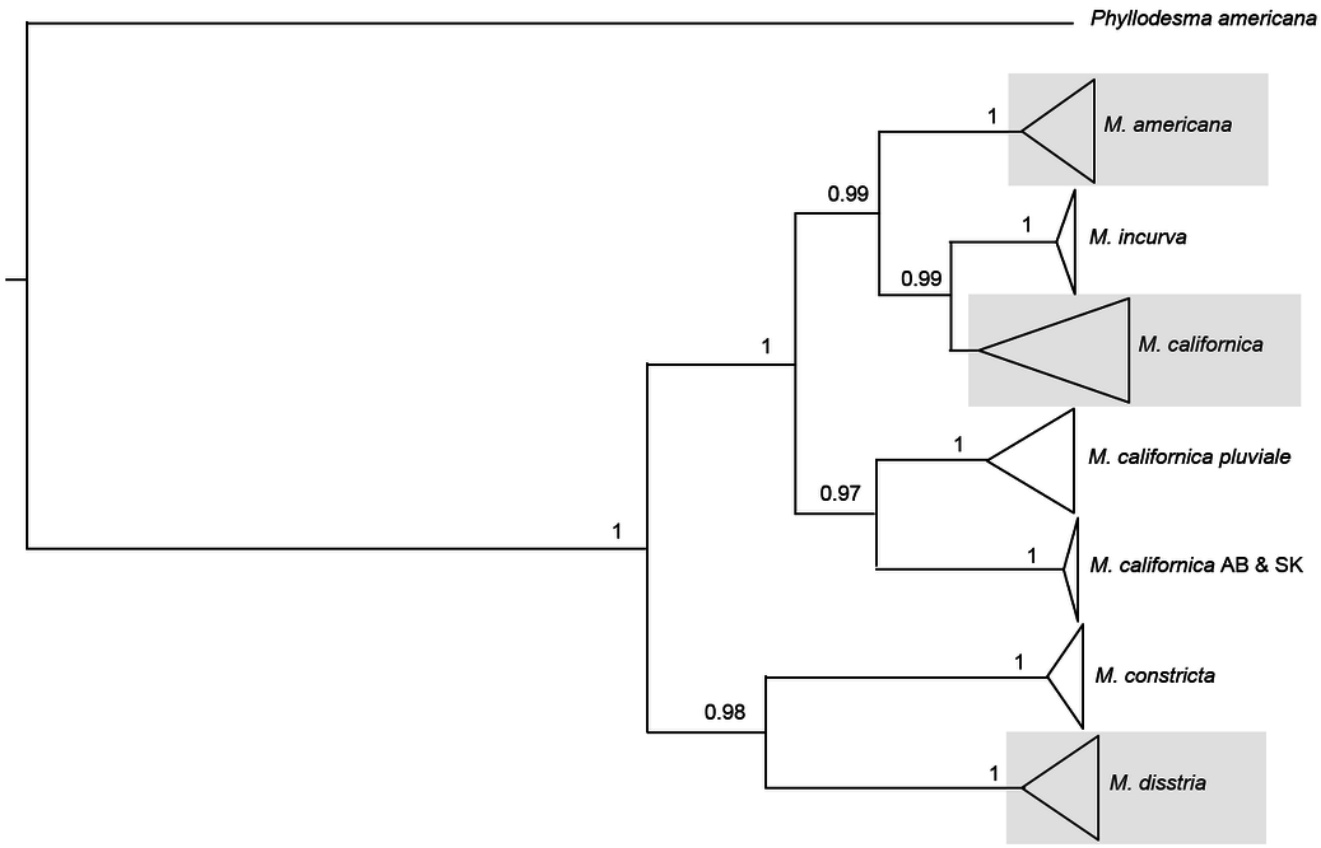


Figure 3

Statistical parsimony network showing the relationship among the 33 *M. americana* haplotypes.

The 79 samples are colour-coded by location, and inferred haplotypes are depicted by black circles. Refer to Fig. 1 for locations.

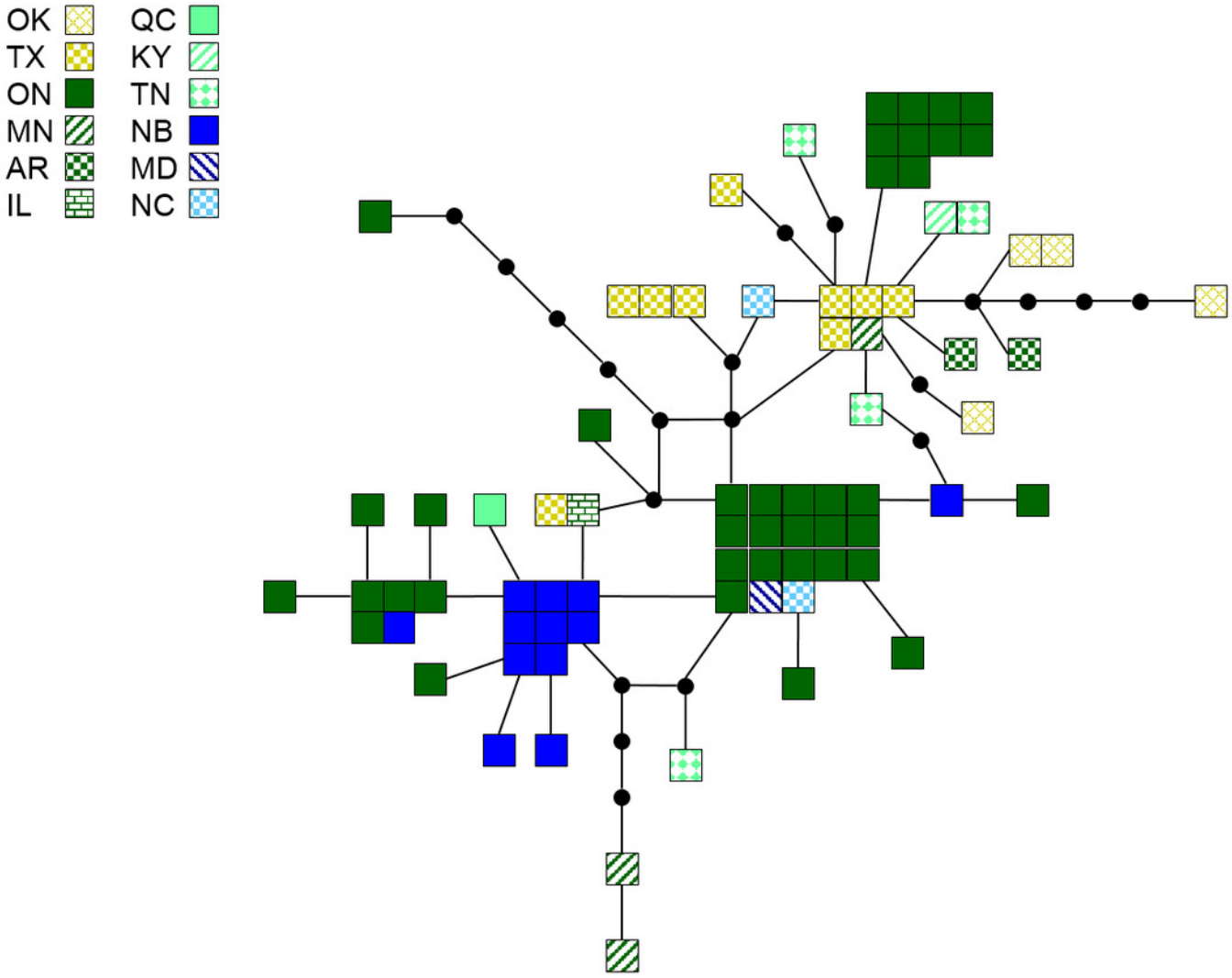
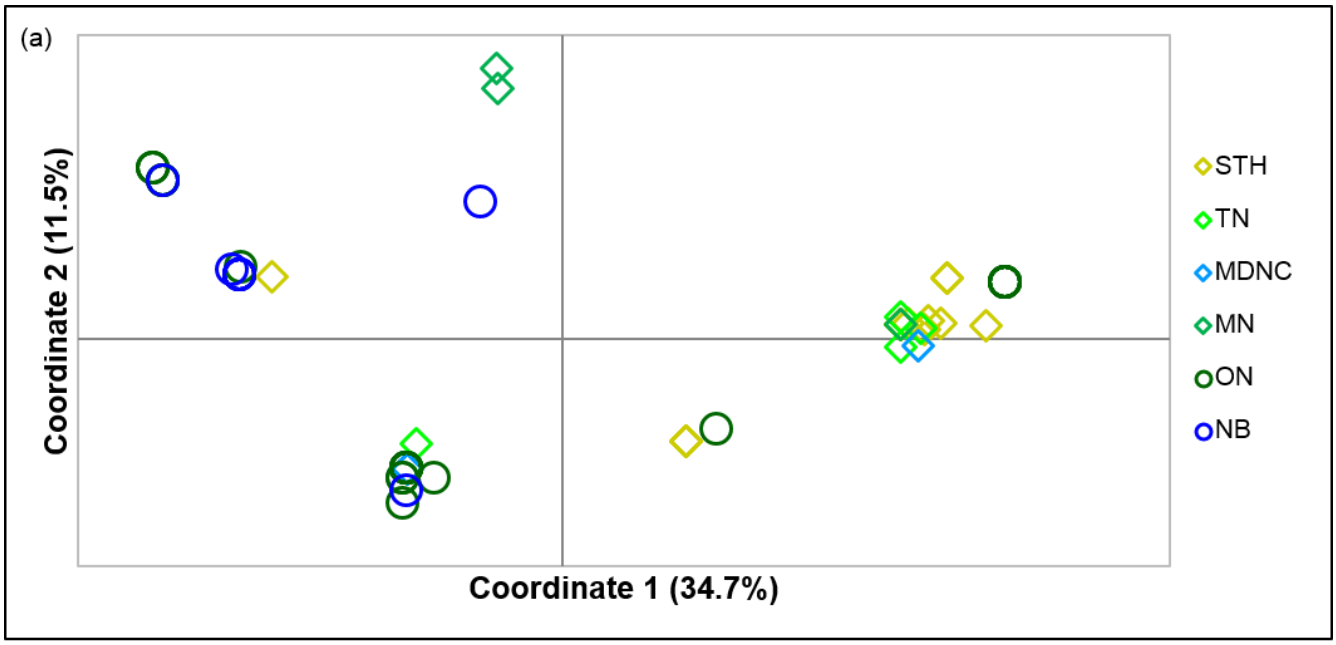
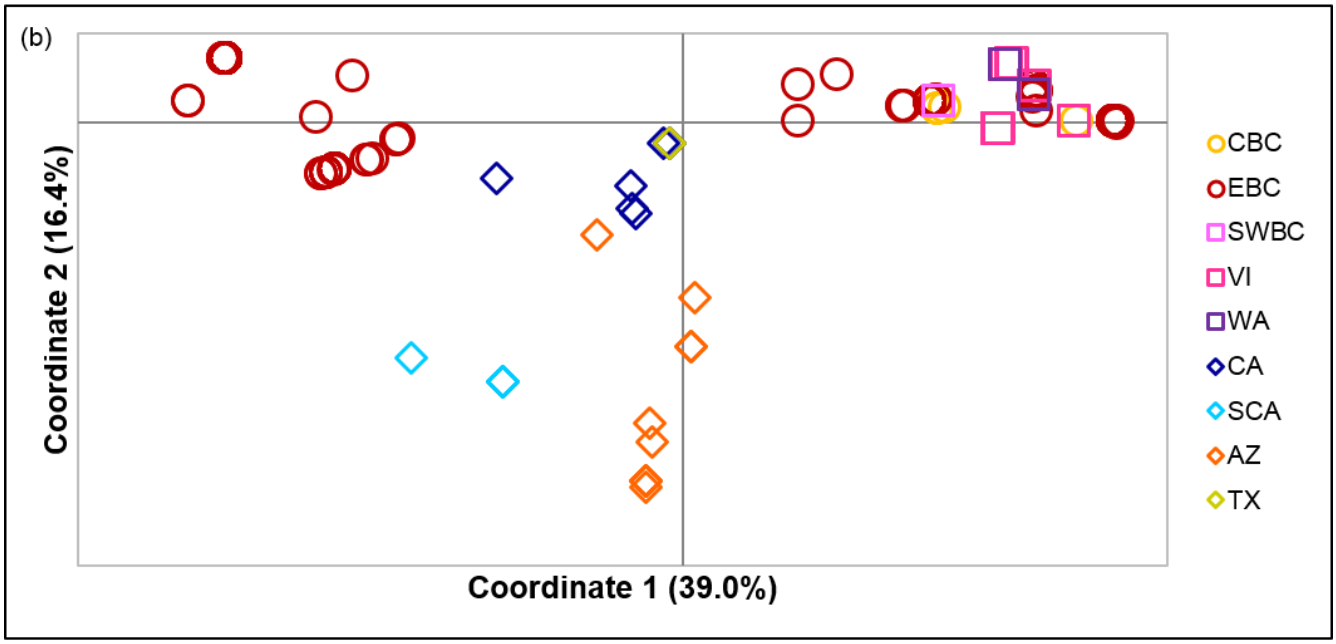


Figure 4

Principle coordinates analysis for (a) *M. americana*, (b) *M. californica*, and (c) *M. disstria*.

Samples are colour-coded by sampling location. Refer to Fig. 1 for locations.


(a)
Coordinate 2 (11.5%)
Coordinate 1 (34.7%)
STH
TN
MDNC
MN
ON
NB


(b)
Coordinate 2 (16.4%)
Coordinate 1 (39.0%)
CBC
EBC
SWBC
VI
WA
CA
SCA
AZ
TX

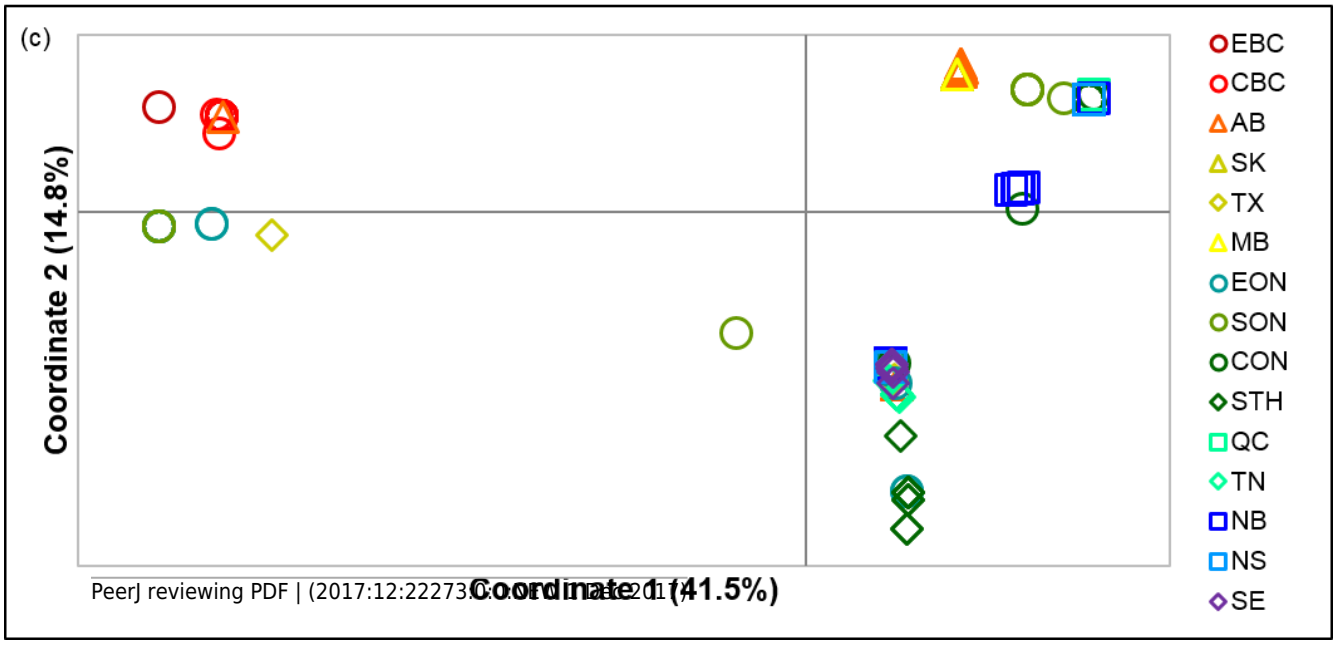

(c)
Coordinate 2 (14.8%)
Coordinate 1 (41.5%)
EBC
CBC
AB
SK
TX
MB
EON
SON
CON
STH
QC
TN
NB
NS
SE

Figure 5

Statistical parsimony network showing the relationship among the 64 *M. californica* haplotypes.

The 207 samples are colour-coded by location, and inferred haplotypes are depicted by black circles. Refer to Fig. 1 for locations.

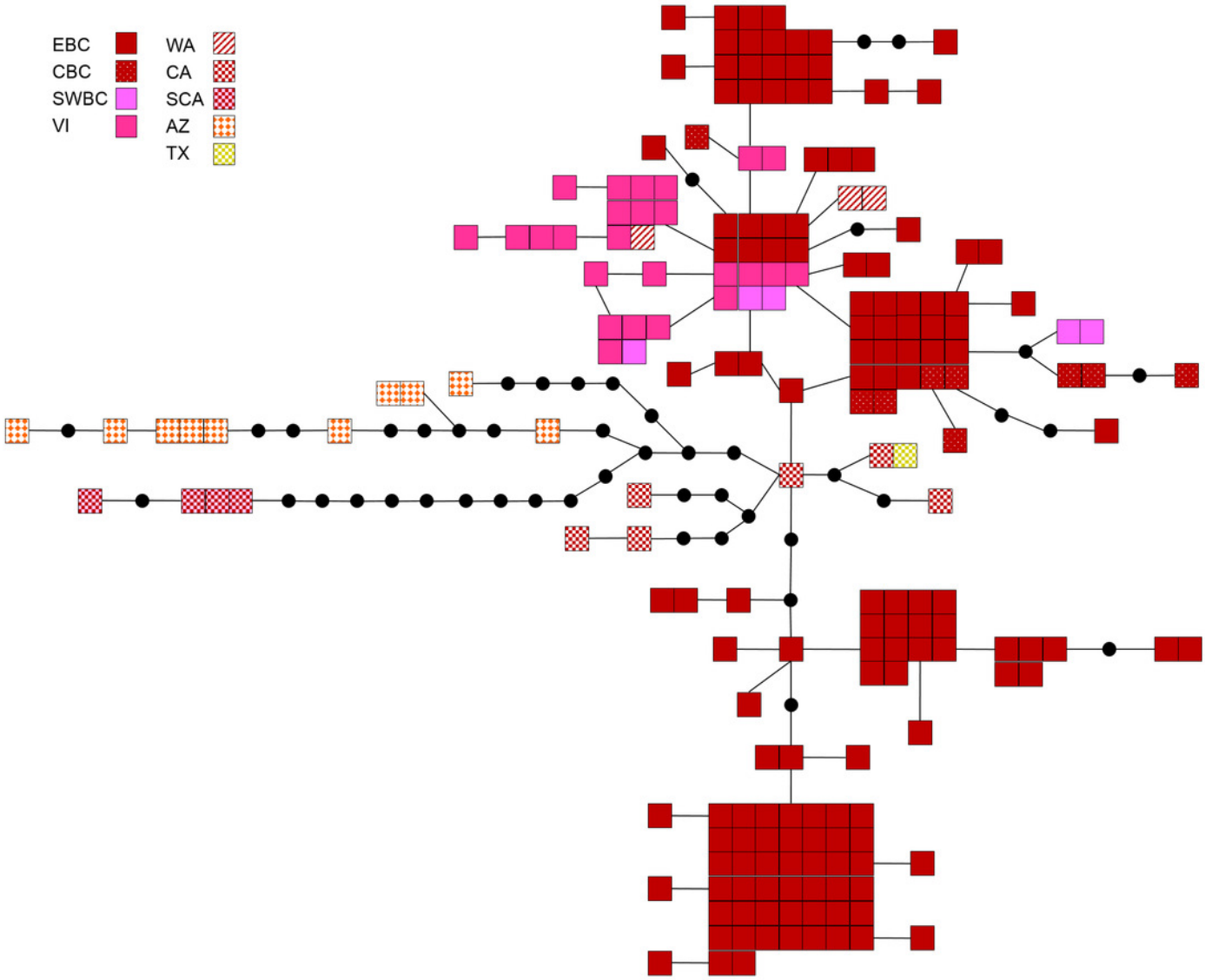


Figure 6

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

The 139 samples are colour-coded by location, and inferred haplotypes are depicted by black circles. Refer to Fig. 1 for locations.