


ARTICLE

Genome-wide markers show continental structuring and mitonuclear discordance in the forest tent caterpillar (*Malacosoma disstria* Hübner) (Lepidoptera: Lasiocampidae)

Amanda D. Roe¹ , Zachary G. MacDonald^{2,3,4} , Kyle L. Snape^{1,2}, and Felix A.H. Sperling²

¹Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste. Marie, Ontario, P6A 2E5, Canada, ²Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2R3, Canada, ³La Kretz Center for California Conservation Science, University of California Los Angeles, Los Angeles, California, 90095, United States of America, and ⁴Institute of the Environmental and Sustainability, University of California Los Angeles, Los Angeles, California, 90095, United States of America

Corresponding author: Amanda D. Roe; Email: amanda.roe@NRCan-RNCan.gc.ca

(Received 20 December 2022; accepted 6 May 2023)

Abstract

The forest tent caterpillar, *Malacosoma disstria* Hübner (Lepidoptera: Lasiocampidae), is an irruptive forest pest found throughout North America. Widespread species such as *M. disstria* are exposed to historical and contemporary processes that are not uniform and can generate regionally distinct genomic variation. Previous analyses used a short mitochondrial fragment to infer broad-scale phylogeographic patterns in *M. disstria*, whereas nuclear markers have been previously applied only in a smaller geographic region. In this study, we quantified *M. disstria* population variation with genome-wide single nucleotide polymorphisms and cytochrome *c* oxidase from mitochondrial DNA. Using highly variable genome-wide markers, we resolved clear genomic differences among populations of *M. disstria* east of the Rocky Mountains that were not detected using mitochondrial variation alone. We also did not detect host-associated divergence in either our genomic or mitochondrial data. Our results highlight the utility of genome-wide markers to resolve intraspecific population structure within a widespread species and support the need for further biogeographic sampling of this forest insect pest.

Introduction

Insect populations are structured by historical and contemporary eco-evolutionary processes interacting at a range of spatial scales. Delimiting and describing spatial genetic structure are fundamental steps towards understanding and disentangling these processes. Mitochondrial DNA has been used extensively to resolve relationships within and among closely related species and was foundational to the field of phylogeography (Avice *et al.* 1987; Avice 2000). However, due to its reduced effective population size and unique pattern of maternal inheritance, this single genetic region can be shaped by neutral or selective forces that do not influence other (*i.e.*, nuclear) genetic markers (Funk and Omland 2003; Ballard and Whitlock 2004; Rubinoff *et al.* 2006). Hence, the evolutionary history of the mitochondrial genome may not reflect the overall history of a species and may provide limited or contradictory information relative to the nuclear genome for inter- and intraspecific genomic variation, leading to observations of mitonuclear discordance (Roe and

Subject editor: Jon Sweeney

© Crown Copyright - Natural Resources Canada, 2023. Published by Cambridge University Press on behalf of The Entomological Society of Canada. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



Sperling 2007; Galtier *et al.* 2009; Dupuis *et al.* 2012; Toews and Brelsford 2012; Teske *et al.* 2018; Campbell *et al.* 2022). Detecting discordance between the mitochondrial and nuclear genomes requires a complementary approach, with both mitochondrial and nuclear markers sampled from the same individuals. Intraspecific variation, especially if it is recently derived, often necessitates large numbers of independent and variable loci to accurately infer population and demographic histories and to resolve spatial genomic variation (Hey and Nielsen 2004; Carling and Brumfield 2007; Gompert *et al.* 2014; Garrick *et al.* 2015; Satler *et al.* 2021). Recent technological advances provide access to single nucleotide polymorphisms (SNPs) spread throughout the genome (Baird *et al.* 2008; Peterson *et al.* 2012; McCormack *et al.* 2013; Garrick *et al.* 2015) and have been used successfully to resolve spatial genetic variation in many widespread, nonmodel organisms (*e.g.*, Bagley *et al.* 2017; Lumley *et al.* 2020; MacDonald *et al.* 2020; Cairns *et al.* 2021; Polic *et al.* 2022).

The forest tent caterpillar, *Malacosoma disstria* Hübner (Lepidoptera: Lasiocampidae), is widely distributed throughout North America, where it exhibits cyclical outbreaks and is an important defoliator of deciduous trees (Cooke *et al.* 2011; Schowalter 2017). The species feeds on a number of different host plants, including aspen (*Populus* spp.) (Salicaceae), maple (*Acer* spp., except *A. rubrum*) (Sapindaceae), oak (*Quercus* spp.) (Fagaceae), and birch (*Betula* spp.) (Betulaceae) in northern parts of its range, along with five additional host genera in the southern United States of America (Fitzgerald 1995, Table 5.1). Previous genetic surveys using mitochondrial DNA detected complex phylogeographic patterns within and among *M. disstria* populations, but underlying mechanisms were not identified (Lait and Hebert 2018). Lait and Hebert (2018) detected a distinct population west of the Rocky Mountains and a variable but unstructured population east of this barrier. Whereas the Rocky Mountains were hypothesised to be a significant barrier to gene flow, spatial genetic variation east of this barrier could not be explained by geographic factors alone (Lait and Hebert 2018). Furthermore, larval host plant was not explored as a potential factor structuring genetic diversity, despite *M. disstria* populations exhibiting significant performance differences on different host plants (*e.g.*, Nicol *et al.* 1997; Parry and Goyer 2004). Recent work by MacDonald *et al.* (2022) used thousands of nuclear genomic markers to identify spatial, environmental, and host-associated divergence among regional populations of *M. disstria* in eastern Canada. This work explored the fine-scale genomic structuring among *M. disstria* collected on distinct host plants across a latitudinal gradient. MacDonald *et al.* (2022) showed that host association, as well as temperature and isolation by distance, collectively shaped genomic variation among eastern *M. disstria* populations. Because this work focused only on populations in eastern Canada, the authors were unable to assess whether wider spatial variation may exist among *M. disstria* populations, nor were they able link their work to the existing mitochondrial DNA data.

Here, we describe the population structure of *M. disstria* in northern boreal and mixed-hardwood forests of Canada east of the Rocky Mountains. We used reduced-representation sequencing to generate thousands of SNPs throughout the genome. Concurrently, we generated mitochondrial sequence data (cytochrome *c* oxidase 1 (*COI*)) from the same individuals and combined these data with previously published data (Lait and Hebert 2018). Together, these nuclear and mitochondrial sequence data were used to: (1) describe the genomic population structure of *M. disstria* throughout the northern boreal and mixed-hardwood forests of Canada; (2) determine whether mitochondrial complexity is related to host association; (3) assess concordance between genome-wide SNPs and mitochondrial DNA, and (4) conclude whether *M. disstria* in northern boreal and mixed-hardwood forests represent a single variable population or distinct regional groups with discrete genetic clustering. Our genome-wide markers show that *M. disstria* exhibit regional population structure east of the Rocky Mountains, but this structure was not related to host association. We also demonstrated the importance of using genome-wide markers when describing intraspecific variation in this widespread species and highlighted the discordance observed between genomic and mitochondrial variation.

Methods

Study organism

Malacosoma disstria has a univoltine life cycle and overwinters as pharate larvae in egg masses laid on host plants during the previous summer (Gray and Ostaff 2012). Females deposit a single egg mass onto a selected host, and larvae emerge in the spring as a family group to feed on new leaves. Synchrony of larval emergence and leaf flush in their host plant is critical for survival (Parry *et al.* 1998; Despland and Noseworthy 2006; McClure and Despland 2010). Siblings form cohesive groups, foraging and resting together within the host plant canopy (Despland and Hamzeh 2004). Larvae remain on their natal host until their later instars and then may disperse to seek additional food sources (Batzer *et al.* 1995). Adult *M. disstria* are capital breeders (*i.e.*, all or most adult energy is obtained as a larvae) and are relatively short lived. Dispersal by flight is male-biased (Struble 1970), whereas females have a short obligatory flight before oviposition (Miller 2006). The dispersal abilities of *M. disstria* are limited, with male moths capable of flying up to 3 km (Evenden *et al.* 2015), but wind-assisted dispersal events over large distances (of more than 100 km) are possible (Brown 1965).

Sample collection and DNA extraction

We sampled *M. disstria* individuals from Alberta ($n = 18$) and Saskatchewan ($n = 6$) in central Canada and from Ontario ($n = 139$) and Quebec ($n = 10$) in eastern Canada. Egg bands or early instar larvae were collected from the four principal host plants in the region: trembling aspen, *Populus tremuloides* Michaux ($n = 131$), sugar maple, *Acer saccharum* Marshall ($n = 32$), white birch, *Betula papyrifera* Marshall ($n = 4$), and northern red oak, *Quercus rubra* Linnaeus ($n = 131$). These hosts were not evenly distributed throughout the range, with only trembling aspen located in central Canada and boreal sites in eastern Canada (Table 1). We sampled egg masses and early instar larvae (at least the fourth instar) because host identity represents the ovipositional choice of females. Sampling of individuals across central and eastern Canada was completed in the summer of 2018, and egg bands were shipped to the Great Lakes Forestry Centre, Sault Ste. Marie, Ontario. Upon arrival, insects were allowed to emerge and were reared at 27° C, 55% relative humidity, and 16:8-hour light:dark photoperiod in a Z-labtech Flex environmental chamber (Z-Sciences Corp., Westmount, Quebec) in the Insect Production and Quarantine Facility. We maintained individual families in separate rearing containers and fed larvae locally collected foliage from their natal hosts. We sampled third to fifth instars from each family (in case of colony collapse due to disease) and reared the remaining larvae to adults when possible. We did in fact lose a number of families to disease, so the larvae were the only samples we had left for extraction from many locations. Adults were frozen at -20° C, and larval specimens were preserved in 100% ethanol before storage at -20° C. Voucher specimens are maintained in a frozen tissue repository at the Great Lakes Forestry Centre.

At each location, we tried to collect egg bands or larvae from 10 to 20 separate trees (Table 1). However, collecting intensity and shipping varied between collaborators, and our final sample size therefore varied among sites, with fewer sampling sites in central Canada than in eastern Canada. Also, some larvae hatched in transit and intermixed with unrelated larvae from other egg bands. This could have resulted in the accidental inclusion of siblings within our analyses, so we performed a kinship analysis on our genomic data to screen for related individuals (see the “[Data filtering and SNP identification](#)” section below).

We dissected thorax tissue from adults ($n = 63$) and the head capsule and upper thorax from larvae ($n = 110$; Table 1). For each larva, we removed the digestive tract to reduce plant or microbial contamination. We extracted genomic DNA from each specimen using a Qiagen DNeasy Blood and Tissue Extraction kit (QIAGEN, Hilden, Germany), following the manufacturer’s protocols, with the addition of bovine pancreatic ribonuclease A (RNaseA,

Table 1. Summary of *Malacosoma disstria* collection details. Sample size represents the number of individuals used in each set of analyses.

| Province | Locality | Host plant | Latitude | Longitude | Collected | Specimens | Sample size (n) | |
|----------|------------------|-----------------------|----------|-----------|--------------|-----------|-----------------|-------|
| | | | | | | | ddRAD* | mtDNA |
| AB | Lac la Biche | <i>P. tremuloides</i> | 54.42 | -111.57 | 2 May 2018 | 18 | 14 | 14 |
| SK | Saskatoon | <i>P. tremuloides</i> | n/a | n/a | 18 May 2018 | 6 | - | 4 |
| QC | Lac Duparquet | <i>P. tremuloides</i> | 48.50 | -79.20 | 18 Jun3 2018 | 4 | 1 | 1 |
| QC | Montebello | <i>A. saccharum</i> | 45.70 | -74.80 | 18 Jun3 2018 | 6 | 4 | 4 |
| ON | Bancroft | <i>A. saccharum</i> | 44.89 | -77.75 | 16 May 2018 | 3 | 2 | 2 |
| ON | Constance Lake | <i>P. tremuloides</i> | 49.81 | -84.19 | 18 May 2018 | 10 | - | 1 |
| ON | Elliot Lake | <i>P. tremuloides</i> | 46.34 | -82.54 | 9 May 2018 | 9 | 5 | 5 |
| ON | Foots Bay | <i>P. tremuloides</i> | 45.17 | -79.76 | 18 May 2018 | 7 | 6 | 6 |
| ON | Hearst | <i>P. tremuloides</i> | 49.73 | -83.91 | 9 May 2018 | 7 | 5 | 5 |
| ON | Kapuskasung | <i>P. tremuloides</i> | 49.64 | -82.30 | 8 May 2018 | 8 | 4 | 5 |
| ON | Kenora | <i>P. tremuloides</i> | 49.76 | -94.48 | 11 May 2018 | 6 | 2 | 2 |
| ON | Killarney | <i>B. papyrifera</i> | 46.01 | -81.40 | 18 May 2018 | 4 | 4 | 4 |
| ON | Lanark | <i>A. saccharum</i> | 44.82 | -76.45 | 5 Jun 2018 | 4 | 4 | 4 |
| ON | Lanark | <i>P. tremuloides</i> | 44.82 | -76.45 | 5 Jun 2018 | 3 | 3 | 3 |
| ON | Latchford | <i>P. tremuloides</i> | 47.33 | -79.81 | 18 May 2018 | 7 | 5 | 4 |
| ON | Little Current | <i>P. tremuloides</i> | 45.88 | -81.90 | 18 May 2018 | 4 | 2 | 2 |
| ON | Marten River | <i>P. tremuloides</i> | 46.67 | -79.73 | 18 May 2018 | 5 | 5 | 5 |
| ON | Nairn | <i>P. tremuloides</i> | 46.30 | -81.68 | 8 May 2018 | 10 | 6 | 6 |
| ON | Oak Shores | <i>Q. rubra</i> | 44.59 | -78.43 | 29 May 2018 | 2 | 2 | 2 |
| ON | Ottawa | <i>A. saccharum</i> | n/a | n/a | 18 May 2018 | 1 | - | 1 |
| ON | Ottawa | <i>P. tremuloides</i> | n/a | n/a | 18 May 2018 | 1 | - | 1 |
| ON | Parry Sound | <i>A. saccharum</i> | 45.38 | -80.04 | 9 May 2018 | 3 | 2 | 2 |
| ON | Rabbit Lake | <i>P. tremuloides</i> | 46.93 | -79.73 | 11 May 2018 | 9 | 6 | 7 |
| ON | Sault Ste Marie | <i>Q. rubra</i> | 46.51 | -84.30 | 7 June 2018 | 4 | 4 | 4 |
| ON | Sault Ste Marie | <i>A. saccharum</i> | 46.51 | -84.30 | 7 June 2018 | 7 | 7 | 7 |
| ON | Sault Ste Marie | <i>P. tremuloides</i> | 46.51 | -84.30 | 7 June 2018 | 18 | 16 | 15 |
| ON | St Joseph Island | <i>A. saccharum</i> | 46.19 | -84.04 | 3 Jun 2018 | 7 | 6 | 6 |
| ON | St Joseph Island | <i>P. tremuloides</i> | 46.19 | -84.04 | 3 June 2018 | 7 | 6 | 7 |
| ON | Wharnclyffe | <i>A. saccharum</i> | 46.54 | -83.54 | 18 May 2018 | 1 | 1 | 1 |

*Final sample size following quality control filtering.

ddRADseq, double-digest restriction site-associated DNA library preparation and sequencing; mt, mitochondrial; AB, Alberta; ON, Ontario; SK, Saskatchewan; QC, Quebec; n/a, location data not available on GenBank.

4 μ L at 100 mg/mL; Sigma-Aldrich Canada Co., Oakville, Ontario), followed by ethanol precipitation and resuspension in Millipore water. Final extracts were stored at -20° C. The quality and quantity of each sample were measured using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Mississauga, Ontario) and Qubit fluorometer (Thermo Fisher Scientific), respectively.

Genome-wide SNPs and analysis

Double-digest restriction site sequencing. We submitted 173 specimens to the Molecular Biology Service Unit (University of Alberta, Edmonton, Alberta, Canada) for double-digest restriction site-associated DNA library preparation and sequencing. Libraries were prepared using 200 ng of genomic DNA, using *PstI* and *MspI* restriction enzymes as in MacDonald *et al.* (2020), generally following Peterson *et al.* (2012) and Poland and Rife (2012). We sequenced our pooled library of individuals using single-end, 75-bp sequencing on an Illumina NextSeq 500 (Illumina, San Diego, California, United States of America) platform in one high-output flowcell.

Data filtering and SNP identification. Bioinformatic processing of raw sequence reads followed MacDonald *et al.* (2022). Briefly, we used Stacks 2.0 (Rochette *et al.* 2019) to demultiplex Illumina single-end, 75-bp reads, removing those with quality scores of less than 20 within a sliding window equal to 15% of the read length. Illumina index sequences and the first 5-bp from the 5' end of each read (corresponding to the *PstI* restriction site) were removed using Cutadapt 1.9.1 (Martin *et al.* 2013). The resulting 62-bp reads were then aligned to a *M. disstria* genome assembly (National Center for Biotechnology Information Accession PRJNA824522; 1090 scaffolds, each over 10 kb) using Burrows–Wheeler Aligner 0.7.17 (BWA-MEM; Li and Durbin 2009; Li 2013). Following alignment, we called SNPs among sequenced individuals using Stacks 2.0, stipulating a single population. Filtering of the resulting data was completed using VCFtools, version 0.1.14 (Danecek *et al.* 2011). We removed: (1) individuals with more than 25% missing data; (2) loci with a read depth less than five; (3) loci with minor allele frequencies of at least 0.05; (4) loci with percentages of more than 5% missing data; and (5) one locus for every pair of loci that were within at least 10 kb of each other.

We screened our data set for full-sibling relationships that could bias assessments of population structure (O'Connell *et al.* 2019). We used the R package SNPRelate, version 1.26.0, (Zheng *et al.* 2012) in the R statistical computing environment (R Core Team 2021) to estimate pairwise kinship coefficients among all sequenced individuals. For diploid organisms, the coefficient value expected for full siblings is 0.25 (Manichaikul *et al.* 2010). A natural break in coefficient values occurred at 0.22. Coefficient values above 0.22 were observed only for pairs of individuals collected from the same location and were therefore inferred to indicate full-sibling relationships. For each pair, we removed the individual with the greater percentage of missing data. Using this final set of individuals, we reverted to original binary alignment map files, recalled SNPs, and repeated filtering using the same parameters listed above.

Population structure. We quantified population structure among our *M. disstria* samples using two different approaches: principal component analysis and the software programme Structure (Pritchard *et al.* 2000). Principal component analyses were conducted in R using the adegenet package, version 2.1.2 (Jombart 2008), and visualised using ggplot, version 2.3.2.1 (Wickham 2016; <https://ggplot2.tidyverse.org>). For Structure analyses, 10 independent runs were completed for each value of $K = 1:10$, each stipulating the admixture model, correlated allele frequencies, a burn-in period of 100 000, and 1 000 000 Markov chain–Monte Carlo (MCMC) repetitions. Location priors (set to collection localities) were used to inform the MCMC algorithm without biasing the model. Runs were visualised in CLUMPAK (Kopelman *et al.* 2015; <http://clumpak.tau.ac.il/index.html>). We identified the optimal value of K using the Pritchard (Pritchard *et al.* 2000) and Evanno methods (Evanno *et al.* 2005). Structure admixture plots were visualised using ggplot2.

Mitochondrial phylogeography

We amplified a 658-bp *COI* fragment for 130 specimens using the LepF and LepR primers based on a modified protocol from Hebert *et al.* (2003). The majority of specimens were also used for double-digest restriction site-associated DNA library preparation and sequencing. However, due to stringent filtering for our downstream SNPs analyses, some individuals were removed, leaving specimens with only *COI* data ($n = 8$). Each polymerase chain reaction (PCR) contained 9.76 μL of dH₂O, 2 μL of 10 \times buffer, 2 μL of MgCl₂, 0.4 μL of deoxynucleoside triphosphate, 0.4 μL of LepF primer, 0.4 μL of LepR primer, 0.04 μL of TopTaq DNA polymerase (QIAGEN, Hilden, Germany), and 5 μL of template DNA. The PCR thermal cycle profile consisted of one cycle at 94° C for 2 minutes, 35 cycles of 94° C for 2 minutes, 45° C for 30 seconds, 72° C for 2 minutes, and a final cycle of 72° C for 5 minutes, followed by a 4° C hold. Afterward, all PCR samples were purified with an exonuclease-shrimp alkaline phosphatase protocol following the manufacturer's instructions (Exo-SAP-IT, Thermo Fisher Scientific, Waltham, Massachusetts, United States of America). Mitochondrial fragments were submitted for Sanger sequencing to the Molecular Biology Service Unit (University of Alberta) on an Applied Biosystems 3730 (Mississauga, Ontario). All sequences were aligned and quality checked using Geneious Prime, version 2019.2.3 (<https://www.geneious.com/>). We combined our data with an additional 139 previously published sequences (Lait and Hebert 2018) downloaded from the Barcode of Life Data System (15 January 2020; <http://www.boldsystems.org>). These included additional material from British Columbia (western) and southern United States of America (southern). We aligned this combined data set using MAFFT online, version 7 (Katoh *et al.* 2019; <https://mafft.cbrc.jp/alignment/server/>; default settings), and trimmed sequences to 658 bp using Mesquite, version 3.6 (Maddison and Maddison 2019; <https://www.mesquiteproject.org/>), to match the previously published data. We constructed Templeton, Crandall, and Sing haplotype networks (also known as statistical parsimony; Templeton *et al.* 1995; Clement *et al.* 2000) in PopART, version 1.7 (Bandelt *et al.* 1999; Leigh and Bryant 2015).

Results

Genome-wide SNPs and analysis

Across all 173 individuals, a total of 186 588 563 75-bp reads were sequenced and passed Illumina quality filters. After filtering and removing adapters and restriction sites, 97.3% of the 62-bp reads were successfully aligned to the *M. disstria* reference genome. After calling and filtering a preliminary set of SNPs, we identified and removed nine individuals with more than 25% missing data and 42 individuals that shared a full-sibling relationship with at least one other individual. We then recalled SNPs and filtered the data based on read depth, minor allele frequency, missing data, and physical proximity. A total of 2828 SNPs with a mean read depth of 26.95 across 122 individuals comprised our final genomic data set.

We observed two distinct clusters of *M. disstria* individuals ($n = 122$) based on genome-wide SNPs (Fig. 1). Our principal component analysis suggested distinct clusters of central ($n = 14$) and eastern ($n = 108$) individuals that separated along principle component 1, which explained 3.55% of total genomic variation (Fig. 1A). Using Structure results, we inferred an optimal number of populations as $K = 3$ (Pritchard method) or $K = 2$ (Evanno method; Fig. 1B). Both $K = 2$ and $K = 3$ identified central samples as a distinct group. We did not, however, see separation of individuals based on host association, with larva from *P. tremuloides* appearing in multiple clusters (Fig. 1).

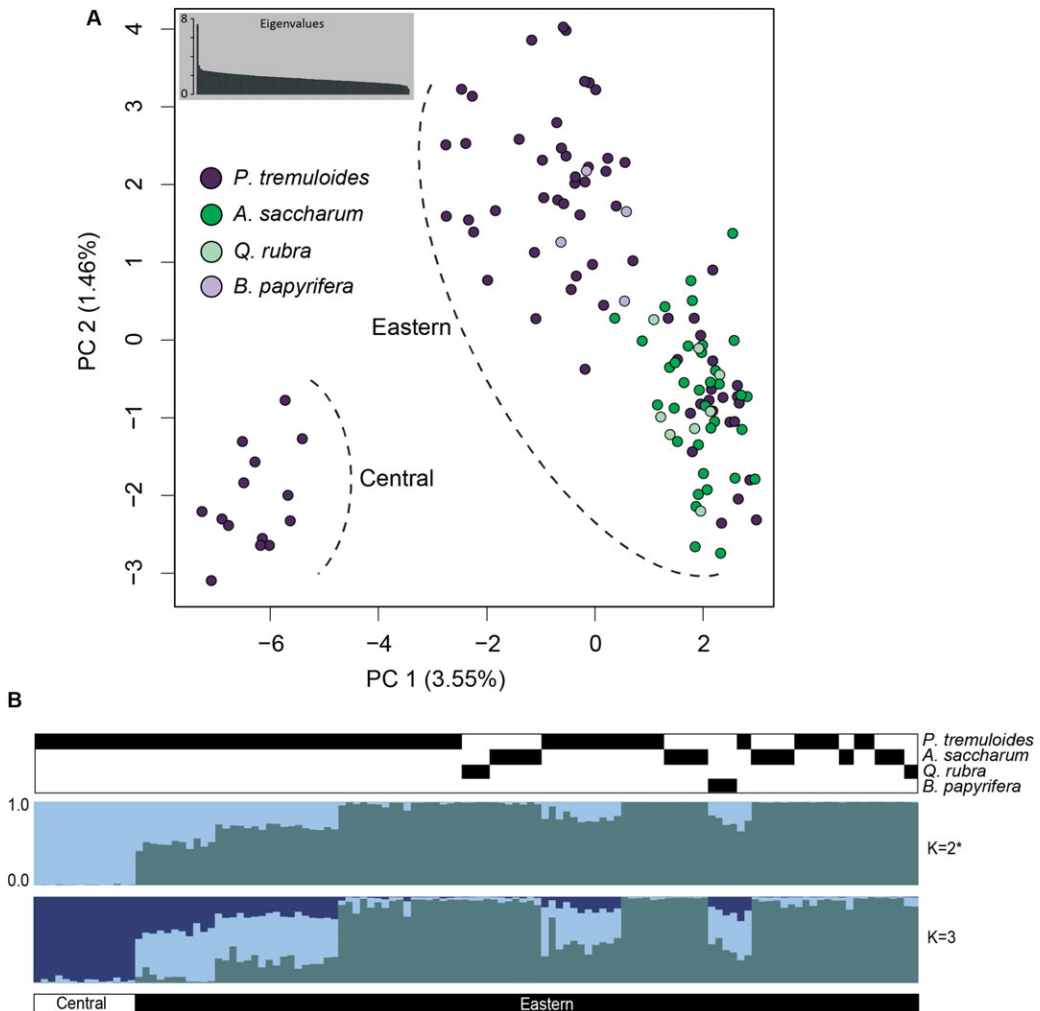


Figure 1. Population genetic structure of *Malacosoma disstria* using **A**, Principal component (PC) analysis and **B**, model-based clustering using *Structure*. In the principal component analysis, every point represents a single individual ($n = 122$) that has been coded for larval host plant. We compared $K = 1-10$ for our *Structure* analyses and identified $K = 2^*$ as optimal using the Evanno method, and $K = 3$ with the Pritchard method (see the “Results” section for details). In the admixture plot from *Structure* individuals are ordered by increasing longitude and we indicate the associated larval host in the bar plot above.

Mitochondrial phylogeography

We sequenced 658 bp of the *COI* gene, corresponding to the DNA barcode region, for 130 *M. disstria* specimens (National Center for Biotechnology Information Genbank Accession numbers MT791498-MT791627). We combined these with an additional 139 sequences previously published by Lait and Hebert (2018; Supplementary material, Table S1). We found complex geographic structuring within the combined data set. Western populations formed a distinct group (Fig. 2), but the mitochondrial haplotypes from central, eastern, and southern populations were variable. These formed an unstructured population that did not align with specific geographic regions (Fig. 2). Furthermore, we observed a number of geographically distant individuals that shared identical haplotypes (Fig. 2). We also did not observe any relationship between mitochondrial genetic variation and host association, with larvae on

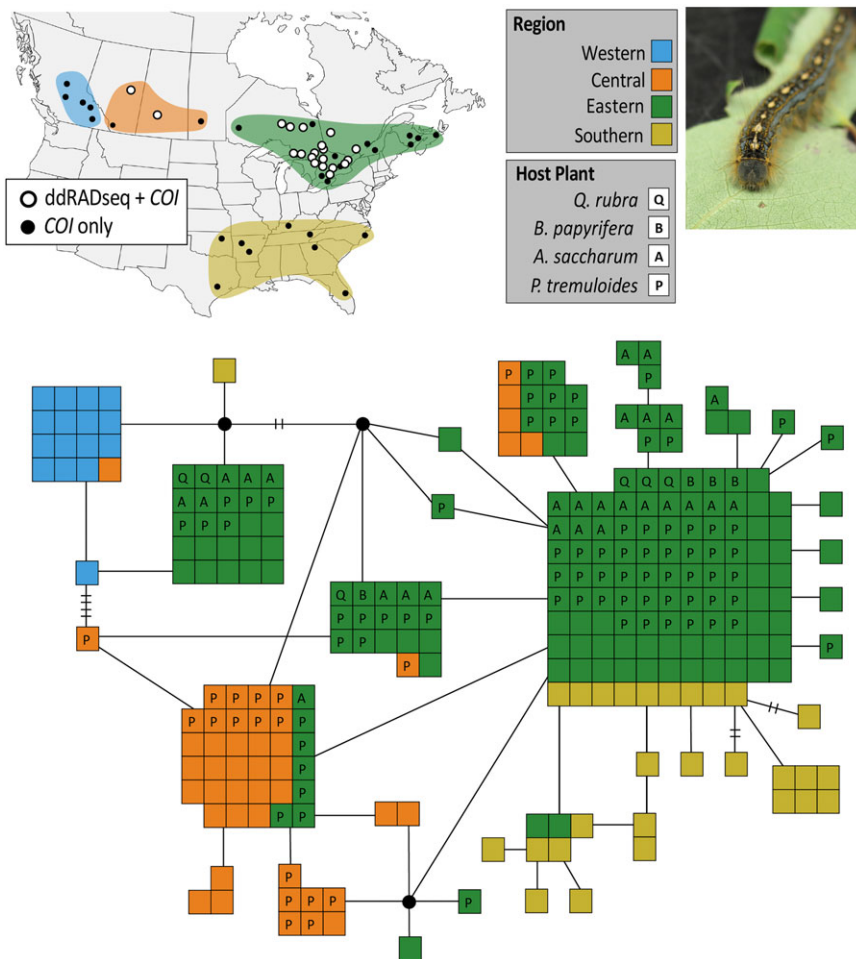


Figure 2. Templeton, Crandall, and Sing haplotype network of 269 *Malacosoma disstria* mitochondrial *cytochrome c oxidase subunit 1* (*COI*) sequences from across Canada and the United States of America. Sequence data from this study ($n = 130$) were combined with data from Lait and Hebert (2018; $n = 139$). Each square represents a single individual, with colour representing a major geographic region and letters indicating larval host plant (when known). Individuals were clustered into distinct *COI* haplotypes, with black dots representing unsampled, potential haplotypes and small black lines representing additional mutational differences.

different hosts sharing identical haplotypes and with larvae on the same host having different haplotypes (Fig. 2).

Discussion

Widespread species are shaped by geographically distinct ecological and evolutionary forces, often giving rise to regionally structured populations with unique genetic signatures. Mitochondrial variation is often insufficient for resolving complex genetic patterns, and genome-wide markers are needed to adequately quantify spatial variation. Using genome-wide SNPs, we show that *M. disstria* east of the Rocky Mountains in Canada are structured into two distinct genetic clusters, similar to other widespread forest species in Canada (e.g., *Choristoneura fumiferana*; Lumley *et al.* 2020). Although we observed geographically structured clusters in our genomic data,

discontinuous sampling between clusters means that we cannot conclusively resolve whether they are connected by a cline of genomic divergence or are separated by a discrete barrier to gene flow. Regardless, our genomic results contradict mitochondrial inferences of homogeneous population structure east of the Rocky Mountains (Lait and Hebert 2018). The lack of concordance between mitochondrial variation and genome-wide markers is not uncommon and points to the importance of using thousands of unlinked markers to assess intraspecific variation in widespread species.

Population genomics of the forest tent caterpillar

Temperate forest species can have complex evolutionary histories arising from shifting distributions during the glacial and interglacial cycles of the Quaternary Period (Hewitt 2004). Although the boreal forest currently represents an extensive contiguous biome, glacial cycles fragmented these habitats and allowed populations to adapt and diverge in isolation and during subsequent postglacial spread (Weir and Schluter 2004; Jaramillo-Correa *et al.* 2009). Combinations of drift and adaptation in these regions have led to intra- and interspecific diversification in many forest species. Similar population structures among different forest species reflect the common demographic responses to historic bioclimatic change and ecological co-associations (Garrick *et al.* 2021). For example, midcontinent biogeographic breaks, similar to what we observed in *M. disstria*, have been detected in other forest animals, including *Choristoneura fumiferana* (Lepidoptera: Tortricidae) (Lumley *et al.* 2020), *Lynx canadensis* (Carnivora: Felidae) (Stenseth *et al.* 1999; Rueness *et al.* 2003), *Ursus americanus* (Carnivora: Ursidae) (Puckett *et al.* 2015; but see Bradburd *et al.* 2018), *Campanula americana* (Campanulaceae) (Prior *et al.* 2020), and *Eptesicus fuscus* (Chiroptera: Vespertilionidae) (Yi and Latch 2022). Many boreal trees also show similar biogeographic patterns, including *Picea glauca* (Pinaceae) (de Lafontaine *et al.* 2010), *Abies balsamifera* (Pinaceae) (Cinget *et al.* 2015), and two *Betula* species (Thomson *et al.* 2015). Herbivorous insects, such as *M. disstria*, would have tracked the fragmentation and expansion of their host plants, leaving similar genetic signatures within their populations (Smith *et al.* 2011; Satler and Carstens 2017). Another lepidopteran forest pest, the eastern spruce budworm, *Choristoneura fumiferana*, has a midcontinental break that separated the species into central and eastern populations (Lumley *et al.* 2020). A similar spatial structure was observed in *P. glauca* and *A. balsamifera* (de Lafontaine *et al.* 2010; Cinget *et al.* 2015), the primary host plants for *C. fumiferana*. Each of these authors suggests that this biogeographic pattern is a result of migration from distinct glacial refugia, creating a common history of expansion and spread (Lumley *et al.* 2020). Recent analyses of *P. tremuloides*, an important host for *M. disstria*, showed a midcontinent division between northeastern and northwestern populations (Goessen *et al.* 2022), similar to *M. disstria*. Goessen *et al.* (2022) proposed that these two northern *P. tremuloides* populations arose following distinct colonisation routes out of a single common refugium, possibly south of the Great Lakes (Jaramillo-Correa *et al.* 2009; Ding *et al.* 2017). It is conceivable that *M. disstria* followed a similar evolutionary trajectory, dispersing from a single refugium (see also Lait and Hebert 2018) and then evolving into distinct regional populations.

Discriminating population structure in widespread species is not trivial. Although we observed a discrete break between central and eastern groups of *M. disstria*, these clusters might represent a genomic cline rather than discrete populations (Bradburd *et al.* 2018). Genomic clines may develop through drift and adaptation so that genetic variation becomes spatially autocorrelated; as samples get further apart, genetic differences between samples also increase (Wright 1943). This phenomenon is common in nature (Schwartz and McKelvey 2008; Meirmans 2012; Perez *et al.* 2018), more so if dispersal is limited (Wright 1943; Slatkin 1985). Many analytical approaches used to identify population structure can artificially delimit clusters in continuous genetic data (Perez *et al.* 2018), particularly with patchy or uneven sampling (Meirmans 2019). This analytical conundrum has been termed the “cluster *versus* cline dilemma” (Guillot

et al. 2009). Recent work on *Populus balsamifera* clearly demonstrates this issue. Keller *et al.* (2010) described two discrete clusters of *P. balsamifera* in central and eastern Canada, similar to the pattern we observed in *M. disstria*. However, when Meirmans *et al.* (2017) increased sampling between these hypothesised groups, they resolved a genetic cline rather than two distinct clusters. Distinguishing between a genetic cline and discrete populations depends on sampling. Uneven sampling across a distribution can strongly influence the analytical delimitation of genomic variation (Meirmans 2015, 2019). Our sampling of *M. disstria* was patchy, particularly between our central and eastern clusters. Improved sampling of this region is required to resolve whether the observed genetic structure is in fact discrete or continuous. Interestingly, *P. tremuloides*, a host plant for *M. disstria*, does show discrete population structure across this part of the range (Goessen *et al.* 2022). We are interested to see, with increased sampling, whether *M. disstria* mirrors the evolutionary history of its host plant.

Our sampling reflects variation in the diversity of available hosts, with higher diversity occurring in eastern Canada (*e.g.*, *Acer* and *Quercus*) than in central Canada (*P. tremuloides*). We did not observe any relationship between population structure and host association for either genome-wide SNPs or mitochondrial haplotypes. The majority of variation in our genomic data set was spatially structured (central *versus* eastern Canada), even among individuals sampled from a single host (*P. tremuloides*; Fig. 1). However, recent work on the eastern population of *M. disstria* identified signals of host-associated genomic divergence within *M. disstria* that was not apparent in patterns of population structuring (MacDonald *et al.* 2022). It is unclear whether *M. disstria* is currently undergoing divergence related to host association that will eventually be reflected in patterns of population clustering or whether gene flow among host-associated groups is simply too high to allow discrete host-associated structuring to precipitate. Given the polyphagous nature of this species, it would be worthwhile to examine the historical biogeography of *M. disstria* in relation to regional host association across its entire North American range using whole-genome sequence data. Another widespread forest insect pest, *Neodiprion lecontei* (Hymenoptera: Diprionidae), has distinct genetic structuring due to Pleistocene drift and host-associated divergence (Bagley *et al.* 2017). Bagley *et al.* (2017) showed that distinct geographic clusters varied in their patterns of host-associated divergence. Given that MacDonald *et al.* (2022) also detected host-associated divergence in the eastern cluster of *M. disstria*, it would be valuable to assess whether similar patterns exist in other populations at broader spatial scales, contrasting spatially structured and host-associated genomic divergences across the species' entire range. Earlier reciprocal transplant experiments clearly showed that functional differences exist between southern and northern populations of *M. disstria* (Parry and Goyer 2004), suggesting that at least some host-associated divergence exists among populations. Therefore, it would be valuable to expand our genomic sampling to these regional populations to assess the impact of geography and host association on genomic divergence.

Mitochondrial discordance

Mitochondrial DNA has been widely used to explore species-level variation and biogeography for more than 20 years (Avice 2000). However, there is mounting evidence that mitochondrial variation is not ideal for inferring population-level variation (Ballard and Whitlock 2004; Toews and Brelsford 2012; Teske *et al.* 2018, Morón-López *et al.* 2022) and that a shift towards unlinked genome-wide nuclear markers is occurring (Garrick *et al.* 2015). We observed discordance between the population structure inferred from genome-wide SNPs and that shown in the mitochondrial *COI* data. We resolved clear differences between central and eastern populations (Fig. 1) using SNPs, but no such patterns were resolved among *COI* haplotypes. In fact, a number of *COI* haplotypes were shared between geographic regions (Fig. 2). Previous work described *COI* admixture in the eastern region (Lait and Hebert 2018); however, we found no evidence of this

substructure in our genomic data. We also did not see correspondence between *COI* variation and host association.

Discordance between nuclear and mitochondrial genomes is a well-known phenomenon (reviewed in Toews and Brelsford 2012) and has been described in a range of taxa (e.g., Wielstra and Arntzen 2020; Cairns *et al.* 2021; Morón-López *et al.* 2022; Yi and Latch 2022). Mitochondrial genomes, due to differences in inheritance, recombination, and effective population size, exhibit distinct responses to selection, demographic changes, and drift (Toews and Brelsford 2012; Bonnet *et al.* 2017). Because the mitochondrial genome is maternally inherited as a single independent unit with little to no recombination, it is susceptible to asymmetrical introgression (Linnen and Farrell 2007) and selective sweeps (Gompert *et al.* 2008; Wendt *et al.* 2022). These evolutionary events may limit the power of mitochondrial DNA to resolve intraspecific variation or to create a distinct evolutionary history when compared with the nuclear genome. This may be particularly prevalent in species with large populations (Gillespie 2000) or in species that have experienced rapid range expansions (Petit and Excoffier 2009). Similar discordant results were detected in *Speyeria aglaja* (Lepidoptera: Nymphalidae) (Polic *et al.* 2022), where distinct geographic lineages were detected with genome-wide SNPs but not with *COI* haplotypes. Polic *et al.* (2022) suggested that gene flow or selective sweeps homogenised haplotype diversity and erased the spatial variation in the mitochondrial genome. We hypothesise that similar events may have occurred within *M. disstria*. We acknowledge, however, that other evolutionary processes, such as adaptive selection (Bazin *et al.* 2006; Pavlova *et al.* 2013), sex-biased dispersal (Folt *et al.* 2019), or cytoplasmic incompatibilities (Hurst and Jiggins 2005), may have also led to incongruent mitochondrial and nuclear variation.

Future directions

Malacosoma disstria is a widespread species found throughout Canada and the United States of America, and our sampling represents only a portion of its total distribution. The combination of postglacial history, host association, and contemporary dispersal associated with irruptive outbreaks (e.g., James *et al.* 2015) could lead to a complex population structure within this species. We know that distinct mitochondrial haplotypes persist in *M. disstria* populations west of the Rocky Mountains (Lait and Hebert 2018), and this structure was not captured in our study due to the spatial extent of our sampling. Previous work has quantified functional variation among other *M. disstria* populations. For example, southern populations of *M. disstria* have distinct regional host associations (Parry and Goyer 2004) and show latitudinal clines in reproductive traits (Parry *et al.* 2001). Given that genomic differentiation was linked to both host association and variation in environmental conditions (MacDonald *et al.* 2022), it is plausible that additional population structure remains undetected throughout the range of *M. disstria*. Increased sampling across the entire range of *M. disstria* is needed to fully describe the population variation in this species and would provide insight to the historical biogeography and evolutionary history of this important forest pest.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.4039/tce.2023.13>.

Acknowledgements. The authors gratefully acknowledge all the collaborators and field crews who assisted with specimen collection for this project: Emma Despland, Vanessa Chaimbrone, Lia Fricano, Joshua Jarry, Mike Francis, Ariel Ilic, Kristin Hicks, Chris McVeety, Tyler Nelson, Christi Jaeger, Anne-Sophie Caron, Tyler Wist, and Reshma Jose. They also thank Sophie Dang, Erin Campbell, Tyler Nelson, Alice (Yuehong) Liu, Eric Lemieux, Kevin Ong, and Meng Zhang for assistance with laboratory procedures, bioinformatics, and larval rearing. Funding was provided by Natural Sciences and Engineering Council Discovery Grant RGPIN-2018-04920 to F.A.H.S., with support from Natural Resources Canada through A.D.R., and a NSERC Alexander Graham

Bell Scholarship - Doctoral (CGS-D) and UCLA La Kretz Center for California Conservation Postdoctoral Fellowship to Z.G.M.

References

- Avise, J.C. 2000. *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Massachusetts, United States of America.
- Avise, J.C., Arnold, J., Ball, M., Bermingham, E., Lamb, T., Neigel, J.E., *et al.* 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**: 489–522.
- Bagley, R.K., Sousa, V.C., Niemiller, M.L., and Linnen, C.R. 2017. History, geography and host use shape genome wide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, **26**: 1022–1044. <https://doi.org/10.1111/mec.13972>.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., *et al.* 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS One*, **3**: e3376. <https://doi.org/10.1371/journal.pone.0003376>.
- Ballard, J.W. and Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, **13**: 729–744. <https://doi.org/10.1046/j.1365-294x.2003.02063.x>.
- Bandelt, H., Forster, P., and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**: 37–48.
- Batzer, H.O., Martin, M.P., Mattson, W.J., and Miller, W.E. 1995. The forest tent caterpillar in aspen stands: distribution and density estimation of four life stages in four vegetation strata. *Forest Science*, **41**: 99–121.
- Bazin, E., Glemin, S., and Galtier, N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**: 570–572.
- Bonnet, T., Leblois, R., Rousset, F., and Crochet, P.A. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, **71**: 2140–2158. <https://doi.org/10.1111/evo.13296>.
- Bradburd, G.S., Coop, G.M., and Ralph, P.L. 2018. Inferring continuous and discrete population genetic structure across space. *Genetics*, **210**: 33–52. <https://doi.org/10.1534/genetics.118.301333>.
- Brown, C.E. 1965. Mass transport of forest tent caterpillar moths, *Malacosoma disstria* Hübner, by a cold front. *The Canadian Entomologist*, **97**: 1073–1075. <https://doi.org/10.4039/Ent971073-10>.
- Cairns, N.A., Cicchino, A.S., Stewart, K.A., Austin, J.D., and Lougheed, S.C. 2021. Cytonuclear discordance, reticulation and cryptic diversity in one of North America's most common frogs. *Molecular Phylogenetics and Evolution*, **156**: 107042. <https://doi.org/10.1016/j.ympev.2020.107042>.
- Campbell, E.O., MacDonald, Z.G., Gage, E.V., Gage, R.V., and Sperling, F.A.H. 2022. Genomics and ecological modelling clarify species integrity in a confusing group of butterflies. *Molecular Ecology*, **31**: 2400–2417. <https://doi.org/10.1111/mec.16407>.
- Carling, M.D. and Brumfield, R.T. 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PLOS One*, **2**: e160. <https://doi.org/10.1371/journal.pone.0000160>.
- Cinget, B., de Lafontaine, G., Gerardi, S., and Bousquet, J. 2015. Integrating phylogeography and paleoecology to investigate the origin and dynamics of hybrid zones: insights from two widespread North American firs. *Molecular Ecology*, **24**: 2856–2870. <https://doi.org/10.1111/mec.13194>.
- Clement, M., Posada, D., and Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**: 1657–1659.
- Cooke, B.J., MacQuarrie, C.J.K., and Lorenzetti, F. 2011. The dynamics of forest tent caterpillar outbreaks across east-central Canada. *Ecography*, **34**: 1–14.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics*, **27**: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.

- de Lafontaine, G., Turgeon, J., and Carine, M. 2010. Phylogeography of white spruce (*Picea glauca*) in eastern North America reveals contrasting ecological trajectories. *Journal of Biogeography*, **37**: 741–751. <https://doi.org/10.1111/j.1365-2699.2009.02241>.
- Despland, E. and Hamzeh, S. 2004. Ontogenetic changes in social behaviour in the forest tent caterpillar, *Malacosoma disstria*. *Behavioral Ecology and Sociobiology*, **56**: 177–184. <https://doi.org/10.1007/s00265-004-0767-8>.
- Despland, E. and Noseworthy, M. 2006. How well do specialist feeders regulate nutrient intake? Evidence from a gregarious tree-feeding caterpillar. *The Journal of Experimental Biology*, **209**: 1301–1309. <https://doi.org/10.1242/jeb.02130>.
- Ding, C., Schreiber, S.G., Roberts, D.R., Hamann, A., and Brouard, J.S. 2017. Post-glacial biogeography of trembling aspen inferred from habitat models and genetic variance in quantitative traits. *Scientific Reports*, **7**: 4672. <https://doi.org/10.1038/s41598-017-04871-7>.
- Dupuis, J.R., Roe, A.D., and Sperling, F.A.H. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, **21**: 4422–4436. <https://doi.org/10.1111/j.1365-294X.2012.05642.x>.
- Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**: 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Evenden, M.L., Whitehouse, C.M., and Jones, B.C. 2015. Resource allocation to flight in an outbreaking forest defoliator *Malacosoma disstria*. *Environmental Entomology*, **44**: 835–845. <https://doi.org/10.1093/ee/nvv055>.
- Fitzgerald, T.D. 1995. *The Tent Caterpillars*. Cornell University Press, Ithaca, New York, United States of America.
- Folt, B., Bauder, J., Spear, S., Stevenson, D., Hoffman, M., Oaks, J.R., *et al.* 2019. Taxonomic and conservation implications of population genetic admixture, mito-nuclear discordance, and male-biased dispersal of a large endangered snake, *Drymarchon couperi*. *PLOS One*, **14**: e0214439. <https://doi.org/10.1371/journal.pone.0214439>.
- Funk, D.J. and Omland, K.E. 2003. Species level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**: 397–423. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132421>.
- Galtier, N., Nabholz, B., Glemin, S., and Hurst, G.D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**: 4541–4550. <https://doi.org/10.1111/j.1365-294X.2009.04380.x>.
- Garrick, R.C., Bonatelli, I.A., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., *et al.* 2015. The evolution of phylogeographic data sets. *Molecular Ecology*, **24**: 1164–1171. <https://doi.org/10.1111/mec.13108>.
- Garrick, R.C., Hyseni, C., Arantes, Í.C., Zachos, L.G., Zee, P.C., Oliver, J.C., and Lozier, J. 2021. Is phylogeographic congruence predicted by historical habitat stability, or ecological co-associations? *Insect Systematics and Diversity*, **5**: 1–18 <https://doi.org/10.1093/isd/ixab018>.
- Gillespie, J.H. 2000. The neutral theory in an infinite population. *Gene*, **261**: 11–18.
- Goessen, R., Isabel, N., Wehenkel, C., Pavy, N., Tischenko, L., Touchette, L., *et al.* 2022. Coping with environmental constraints: geographically divergent adaptive evolution and germination plasticity in the transcontinental *Populus tremuloides*. *Plants, People, Planet*, **4**: 638–654. <https://doi.org/10.1002/ppp3.10297>.
- Gompert, Z., Forister, M.L., Fordyce, J.A., and Nice, C.C. 2008. Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, **17**: 5231–5244. <https://doi.org/10.1111/j.1365-294X.2008.03988.x>.
- Gompert, Z., Lucas, L.K., Buerkle, C.A., Forister, M.L., Fordyce, J.A., and Nice, C.C. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**: 4555–4573. <https://doi.org/10.1111/mec.12811>.

- Gray, D.R. and Ostaff, D.P. 2012. Egg hatch of forest tent caterpillar (Lepidoptera: Lasiocampidae) on two preferred host species. *The Canadian Entomologist* **144**: 756–763. <https://doi.org/10.4039/tce.2012.73>.
- Guillot, G., Leblois, R., Coulon, A., and Frantz, A.C. 2009. Statistical methods in spatial genetics. *Molecular Ecology*, **18**: 4734–4756. <https://doi.org/10.1111/j.1365-294X.2009.04410.x>.
- Hebert, P.D., Cywinska, A., Ball, S.L., and deWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**: 313–321. <https://doi.org/10.1098/rspb.2002.2218>.
- Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**:183–195. <https://doi.org/10.1098/rstb.2003.1388>.
- Hey, J. and Nielsen, R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**: 747–760. <https://doi.org/10.1534/genetics.103.024182>.
- Hurst, G.D. and Jiggins, F.M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **272**: 1525–1534. <https://doi.org/10.1098/rspb.2005.3056>.
- James, P.M.A., Cooke, B., Brunet, B.M.T., Lumley, L.M., Sperling, F.A.H., Fortin, M.J., *et al.* 2015. Life-stage differences in spatial genetic structure in an irruptive forest insect: implications for dispersal and spatial synchrony. *Molecular Ecology*, **24**: 296–309. <https://doi.org/10.1111/mec.13025>.
- Jaramillo-Correa, J.P., Beaulieu, J., Khasa, D.P., and Bousquet, J. 2009. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research*, **39**: 286–307. <https://doi.org/10.1139/x08-181>.
- Jombart, T. 2008. ADEGENET: an R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**: 1403–1405.
- Katoh, K., Rozewicki, J., and Yamada, K.D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, **20**: 1160–1166. <https://doi.org/10.1093/bib/bbx108>.
- Keller, S.R., Olson, M.S., Silim, S., Schroeder, W., and Tiffin, P. 2010. Genomic diversity, population structure, and migration following rapid range expansion in the Balsam poplar, *Populus balsamifera*. *Molecular Ecology*, **19**: 1212–1226. <https://doi.org/10.1111/j.1365-294X.2010.04546.x>.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**: 1179–1191. <https://doi.org/10.1111/1755-0998.12387>.
- Lait, L.A. and Hebert, P.D.N. 2018. Phylogeographic structure in three North American tent caterpillar species (Lepidoptera: Lasiocampidae): *Malacosoma americana*, *M. californica*, and *M. disstria*. *PeerJ*, **6**: e4479. <https://doi.org/10.7717/peerj.4479>.
- Leigh, J.W. and Bryant, D. 2015. POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**: 1110–1116. <https://doi.org/10.1111/2041-210X.12410>.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [preprint]. arXiv: 1303.3997. <https://doi.org/10.48550/arXiv.1303.3997>.
- Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Linnen, C.R. and Farrell, B.D. 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**: 1417–1438. <https://doi.org/10.1111/j.1558-5646.2007.00114.x>.
- Lumley, L.M., Pouliot, E., Laroche, J., Boyle, B., Brunet, B.M.T., Levesque, R.C., *et al.* 2020. Continent-wide population genomic structure and phylogeography of North America’s most

- destructive conifer defoliator, the spruce budworm (*Choristoneura fumiferana*). *Ecology and Evolution*, **10**: 914–927. <https://doi.org/10.1002/ece3.5950>.
- MacDonald, Z.G., Dupuis, J.R., Davis, C.S., Acorn, J.H., Nielsen, S.E., and Sperling, F.A.H. 2020. Gene flow and climate-associated genetic variation in a vagile habitat specialist. *Molecular Ecology*, **29**: 3889–3906. <https://doi.org/10.1111/mec.15604>.
- MacDonald, Z.G., Snape, K.L., Roe, A.D., and Sperling, F.A.H. 2022. Host association, environment, and geography underlie genomic differentiation in a major forest pest. *Evolutionary Applications*, **15**: 1749–1765. <https://doi.org/10.1111/eva.13466>.
- Maddison, W.P. and Maddison, D.R. 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61. Available from <http://mesquiteproject.org> [accessed 15 December 2019].
- Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics*, **26**: 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>.
- Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., *et al.* 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**: 1817–1828. <https://doi.org/10.1101/gr.159426.113>.
- McClure, M. and Despland, E. 2010. Collective foraging patterns of field colonies of *Malacosoma disstria* caterpillars. *The Canadian Entomologist*, **142**: 473–480. <https://doi.org/10.4039/n10-001>.
- McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., and Brumfield, R.T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**: 526–538. <https://doi.org/10.1016/j.ympev.2011.12.007>.
- Meirmans, P.G. 2012. AMOVA-based clustering of population genetic data. *Journal of Heredity*, **103**: 744–750. <https://doi.org/10.1093/jhered/ess047>.
- Meirmans, P.G. 2015. Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, **24**: 3223–3231. <https://doi.org/10.1111/mec.13243>.
- Meirmans, P.G. 2019. Subsampling reveals that unbalanced sampling affects STRUCTURE results in a multi-species dataset. *Heredity*, **122**: 276–287. <https://doi.org/10.1038/s41437-018-0124-8>.
- Meirmans, P.G., Godbout, J., Lamothe, M., Thompson, S.L., and Isabel, N. 2017. History rather than hybridization determines population structure and adaptation in *Populus balsamifera*. *Journal of Evolutionary Biology*, **30**: 2044–2058. <https://doi.org/10.1111/jeb.13174>.
- Miller, W. 2006. Forest tent caterpillar: mating, oviposition and adult congregation at town lights during a Northern Minnesota outbreak. *Journal of the Lepidopterists' Society*, **60**: 156–160.
- Morón-López, J., Vergara, K., Sato, M., Gajardo, G., and Ueki, S. 2022. Intraspecies variation of the mitochondrial genome: an evaluation for phylogenetic approaches based on the conventional choices of genes and segments on mitogenome. *PLOS One*, **17**: e0273330. <https://doi.org/10.1371/journal.pone.0273330>.
- Nicol, R.W., Arnason, J.T., Helson, B., and Abou-Zaid, M.M. 1997. Effect of host and nonhost trees on the growth and development of the forest tent caterpillar, *Malacosoma disstria* (Lepidoptera: Lasiocampidae). *The Canadian Entomologist*, **129**: 991–999. <https://doi.org/10.4039/Ent129991-6>.
- O'Connell, K.A., Mulder, K.P., Maldonado, J., Currie, K.L., and Ferraro, D.M. 2019. Sampling related individuals within ponds biases estimates of population structure in a pond-breeding amphibian. *Ecology and Evolution*, **9**: 3620–3636. <https://doi.org/10.1002/ece3.4994>.
- Parry, D. and Goyer, R.A. 2004. Variation in the suitability of host tree species for geographically discrete populations of forest tent caterpillar. *Environmental Entomology*, **33**: 1477–1487.
- Parry, D., Goyer, R.A., and Lenhard, G.J. 2001. Macrogeographic clines in fecundity, reproductive allocation, and offspring size of the forest tent caterpillar *Malacosoma disstria*. *Ecological Entomology*, **26**: 281–291. <https://doi.org/10.1046/j.1365-2311.2001.00319.x>.

- Parry, D., Spence, J.R., and Volney, W.J.A. 1998. Budbreak phenology and natural enemies mediate survival of first-instar forest tent caterpillar (Lepidoptera: Lasiocampidae). *Environmental Entomology*, **27**: 1368–1374.
- Pavlova, A., Amos, J.N., Joseph, L., Loynes, K., Austin, J.J., Keogh, J.S., *et al.* 2013. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution*, **67**: 3412–3428. <https://doi.org/10.1111/evo.12107>.
- Perez, M.F., Franco, F.F., Bombonato, J.R., Bonatelli, I.A.S., Khan, G., Romeiro-Brito, M., *et al.* 2018. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions*, **24**: 1883–1889. <https://doi.org/10.1111/ddi.12816>.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS One*, **7**: e37135. <https://doi.org/10.1371/journal.pone.0037135>.
- Petit, R.J. and Excoffier, L. 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution*, **24**: 386–393. <https://doi.org/10.1016/j.tree.2009.02.011>.
- Poland, J.A. and Rife, T.W. 2012. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome Journal*, **5**: 92. <https://doi.org/10.3835/plantgenome2012.05.0005>.
- Polic, D., Yildirim, Y., Lee, K.M., Franzen, M., Mutanen, M., Vila, R., and Forsman, A. 2022. Linking large-scale genetic structure of three Argynniini butterfly species to geography and environment. *Molecular Ecology*, **31**: 4381–4401. <https://doi.org/10.1111/mec.16594>.
- Prior, C.J., Layman, N.C., Koski, M.H., Galloway, L.F., and Busch, J.W. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. *Molecular Ecology*, **29**: 4473–4486. <https://doi.org/10.1111/mec.15650>.
- Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945–959.
- Puckett, E.E., Etter, P.D., Johnson, E.A., and Eggert, L.S. 2015. Phylogeographic analyses of American black bears (*Ursus americanus*) suggest four glacial refugia and complex patterns of postglacial admixture. *Molecular Biology and Evolution*, **32**: 2338–2350. <https://doi.org/10.1093/molbev/msv114>.
- R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from www.R-project.org [accessed 10 June 2021].
- Rochette, N.C., Rivera-Colon, A.G., and Catchen, J.M. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, **28**: 325–345. <https://doi.org/10.1111/mec.15253>.
- Roe, A.D. and Sperling, F.A.H. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*, **44**: 325–345. <https://doi.org/10.1016/j.ympev.2006.12.005>.
- Rubinoff, D., Cameron, S., and Will, K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for “barcoding” identification. *Journal of Heredity*, **97**: 581–594.
- Rueness, E.K., Stenseth, N.C., O’Donoghue, M., Boutin, S., Ellegren, H., and Jakobsen, K.S. 2003. Ecological and genetic spatial structuring in the Canadian lynx. *Nature*, **425**: 69–72. <https://doi.org/10.1038/nature01942>.
- Satler, J.D. and Carstens, B.C. 2017. Do ecological communities disperse across biogeographic barriers as a unit? *Molecular Ecology*, **26**: 3533–3545. <https://doi.org/10.1111/mec.14137>.
- Satler, J.D., Carstens, B.C., Garrick, R.C., Espíndola, A., and Mardulyn, P. 2021. The phylogeographic shortfall in hexapods: a lot of leg work remaining. *Insect Systematics and Diversity*, **5**: 1–18. <https://doi.org/10.1093/isd/ixab015>.
- Schowalter, T.D. 2017. Biology and management of the forest tent caterpillar (Lepidoptera: Lasiocampidae). *Journal of Integrated Pest Management*, **8**: 24 <https://doi.org/10.1093/jipm/pmx022>.

- Schwartz, M.K. and McKelvey, K.S. 2008. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**: 441–452. <https://doi.org/10.1007/s10592-008-9622-1>.
- Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**: 393–430. <https://doi.org/10.1146/annurev.es.16.110185.002141>.
- Smith, C.I., Tank, S., Godsoe, W., Levenick, J., Strand, E., Esque, T., and Pellmyr, O. 2011. Comparative phylogeography of a coevolved community: concerted population expansions in Joshua trees and four yucca moths. *PLOS One*, **6**: e25628. <https://doi.org/10.1371/journal.pone.0025628>.
- Stenseth, N.C., Chan, K.S., Tong, H., Boonstra, R., Boutin, S., Krebs, C.J., *et al.* 1999. Common dynamic structure of Canada lynx populations within three climatic regions. *Science*, **285**: 1071–1073. <https://doi.org/10.126/science.285.5430.1071>.
- Struble, D.L. 1970. A sex pheromone in the forest tent caterpillar. *Journal of Economic Entomology*, **63**: 295–296.
- Templeton, A.R., Routman, E., and Phillips, C.A. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**: 767–782.
- Teske, P.R., Golla, T.R., Sandoval-Castillo, J., Emami-Khoyi, A., van der Lingen, C.D., von der Heyden, S., *et al.* 2018. Mitochondrial DNA is unsuitable to test for isolation by distance. *Scientific Reports*, **8**: 8448. <https://doi.org/10.1038/s41598-018-25138-9>.
- Thomson, A.M., Dick, C.W., Dayanandan, S., and Carine, M. 2015. A similar phylogeographical structure among sympatric North American birches (*Betula*) is better explained by introgression than by shared biogeographical history. *Journal of Biogeography*, **42**: 339–350. <https://doi.org/10.1111/jbi.12394>.
- Toews, D.P. and Brelsford, A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**: 3907–3930. <https://doi.org/10.1111/j.1365-294X.2012.05664.x>.
- Weir, J.T. and Schluter, D. 2004. Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society London B: Biological Sciences*, **271**: 1881–1887. <https://doi.org/10.1098/rspb.2004.2803>.
- Wielstra, B. and Arntzen, J.W. 2020. Extensive cytonuclear discordance in a crested newt from the Balkan Peninsula glacial refugium. *Biological Journal of the Linnean Society*, **130**: 578–585.
- Wendt, M., Kulaneck, D., Varga, Z., Rakosy, L., and Schmitt, T. 2022. Pronounced mito-nuclear discordance and various Wolbachia infections in the water ringlet *Erebia pronoe* have resulted in a complex phylogeographic structure. *Scientific Reports*, **12**: 5175. <https://doi.org/10.1038/s41598-022-08885-8>.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wright, S. 1943. Isolation by distance. *Genetics*, **28**: 114–138.
- Yi, X. and Latch, E.K. 2022. Nuclear phylogeography reveals strong impacts of gene flow in big brown bats. *Journal of Biogeography*, **49**: 1061–1074.
- Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**: 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>.

Cite this article: Roe, A.D., MacDonald, Z.G., Snape, K.L., and Sperling, F.A.H. 2023. Genome-wide markers show continental structuring and mitonuclear discordance in the forest tent caterpillar (*Malacosoma disstria* Hübner) (Lepidoptera: Lasiocampidae). *The Canadian Entomologist*. <https://doi.org/10.4039/tce.2023.13>.