Received: 19 April 2019 | Revised: 7 August 2019 | Accepted: 12 August 2019

DOI: 10.1111/jbi.13702

RESEARCH PAPER

Journal of Biogeography | WILEY

# Native or non-native? Historical biogeography of an emergent forest pest, *Matsucoccus macrocicatrices*

Thomas D. Whitney[1,2] | Kamal J. K. Gandhi[1] | Rima D. Lucardi[2]

[1]Puyallup Research and Extension Center, Washington State University, Puyallup, WA 98371, USA

[2]USDA Forest Service, Southern Research Station, 320 E. Green Street, Athens, GA 30602, USA

**Correspondence**
Thomas D. Whitney, Puyallup Research and Extension Center, Washington State University, Puyallup, WA 98371, USA.
Email: thomasdantaswhitney@gmail.com

**Funding information**
USDA Forest Service, Southern Research Station, Grant/Award Number: 13-CA-11330129-056 and 16-CS-11330129-045; Southern Region (8)-Forest Health Protection; USDA Agricultural and Food Research Initiative, Foundation Grant; D.B. Warnell School of Forestry and Natural Resources, University of Georgia

Handling Editor: Vincent Merckx

## Abstract

**Aim:** A historically benign insect herbivore, *Matsucoccus macrocicatrices*, has recently been linked to dieback and mortality of eastern white pine (*Pinus strobus* L.). Previous reports indicated that its native range was restricted to New England, USA and southeastern Canada. Now, the insect occurs throughout an area extending from the putative native range, southward to Georgia, and westward to Wisconsin. Our goal was to evaluate whether its current distribution was due to recent introductions consistent with invasion processes. We considered two hypotheses: (a) if recent expansion into adventive regions occurred, those populations would have reduced genetic diversity due to founder effect(s); alternatively (b) if *M. macrocicatrices* is native and historically co-occurred with its host tree throughout the North American range, then populations would have greater overall genetic diversity and a population structure indicative of past biogeographical influences.

**Location:** Eastern North America.

**Methods:** We developed nine *M. macrocicatrices*-specific microsatellite markers de novo and genotyped 390 individuals from 22 populations sampled across the range of eastern white pine in the USA. We assessed genetic variability, relatedness, and population structure.

**Results:** There were no signatures of founder effects. The only differences in genetic diversity occurred latitudinally, where the number of rare alleles and observed heterozygosity was highest in the southern range extent. Analyses of population structure indicated three distinct genetic clusters separated by the Great Lakes and the Blue Ridge Mountains.

**Main Conclusions:** The seemingly sudden ecological shift from benign herbivore to significant pest led us to suspect that *M. macrocicatrices* was non-native. However, our findings suggest that this insect is native and has likely co-occurred with its host tree since the last glacial maximum. Our study demonstrates the importance of historical biogeographical reconstruction to inform how to approach an emergent pest.

KEYWORDS

Blue Ridge Mountains, eastern white pine bast scale, forest health, Great Lakes, microsatellites, native pest, *Pinus strobus*, population genetics, Southern Appalachians

  

## 1 | INTRODUCTION

Non-native species lack the long evolutionary history that native species have within a local community, and hence communities can suffer greater damage from non-natives due to naiveté (Paolucci, MacIsaac, & Ricciardi, 2013; Salo, Korpimaki, Banks, Nordstrom, & Dickman, 2007; Simberloff, Souza, Nunez, Barrios-Garcia, & Bunn, 2012). However, endemism does not preclude a species from developing pestiferous behaviours. Although rarer, native species can become pests within their native ranges similar to non-native species through expansion into adventive ranges (Dodds et al., 2018; Hassan & Ricciardi, 2014; Simberloff et al., 2012).

Unifying all organisms causing serious ecological and economic damage is the release from evolutionary constraints and/or the exploitation of new niche opportunities (Carey, Sanderson, Barnas, & Olden, 2012). For instance, the absence of co-evolved natural enemies (Keane & Crawley, 2002) or host/prey defenses (Gandhi & Herms, 2010; Paolucci et al., 2013) can allow non-natives to establish and thrive in novel environments, but for a native species, these constraints on their populations generally remain intact (Tong, Wang, & Chen, 2018). Instead, the reasons certain native species elevate to pest status are often multi-faceted, sometimes involving positive population responses to climate change (Nackley, West, Skowno, & Bond, 2017), anthropogenic habitat alterations (Carrete et al., 2010) and/or host-shifts following other non-native introductions (Lefort et al., 2014). Reconstructing the historical origin of an emergent pest species can provide an evolutionary context to its contemporary interactions (Richardson & Ricciardi, 2013; Sakai et al., 2001), an important first step in control and conservation efforts.

In this study, we evaluated the population genetic variability and distribution of the eastern white pine bast scale, *Matsucoccus macrocicatrices* Richards (Hemiptera: Matsucoccidae), a small sap-sucking insect currently associated with the novel dieback phenomenon of eastern white pine (*Pinus strobus* L.) in North America (Costanza, Whitney, McIntire, Livingston, & Gandhi, 2018; Mech et al., 2013; Figure 1a,b). This insect creates deep feeding wounds during its second-instar cyst stage, which is hypothesized to facilitate subcortical infection of trees by pathogens, primarily the native *Caliciopsis pinea* Peck (Schulz et al., 2018). This fungus requires an entry point, such as a bark fissure or insect feeding site, to successfully colonize a host (Funk, 1963). Once established in the cambium, it causes the formation of cankers on the bark (Figure 1c), which leads to hallmark symptoms, including the girdling of stems in young trees and the bottom-up branch dieback in older trees (Figure 1d; Asaro, Chamberlin, Rose, Mooneyham, & Rose, 2018; Costanza et al., 2018). The pathogenic effects of *C. pinea* have long been known (Ray, 1936), but the severity and scope of the current symptoms are unprecedented (Costanza et al., 2018). A similar scenario to beech bark disease in American beech (*Fagus grandifola* Ehrh.) may also be occurring in eastern white pine, where the feeding behaviour of a non-native sap-sucking insect has allowed fungal pathogens to infect and kill host trees at an increased rate (Houston, 1994). Although a causal mechanism has not yet been identified, recent research has found the incidence of *M. macrocicatrices*, Caliciopsis cankers and dieback symptoms in eastern white pine to be highly correlated (Schulz et al., 2018; Whitney et al., 2018).

Prior to 2011, *M. macrocicatrices* was considered a benign herbivore with a limited distribution. The only recorded specimens were collected from eastern white pine in the northeastern USA (Massachusetts, New Hampshire, and Vermont) and southeastern Canada (New Brunswick, Nova Scotia, Ontario, and Quebec; Mech et al., 2013; Richards, 1960; Watson, Underwood, & Reid, 1960). However, it has now been observed throughout the North American range of eastern white pine linked to host-tree damage and mortality (Mech et al., 2013; Schulz et al., 2018). Other *Matsucoccus* spp. have become pests outside their native ranges, such as the Japanese

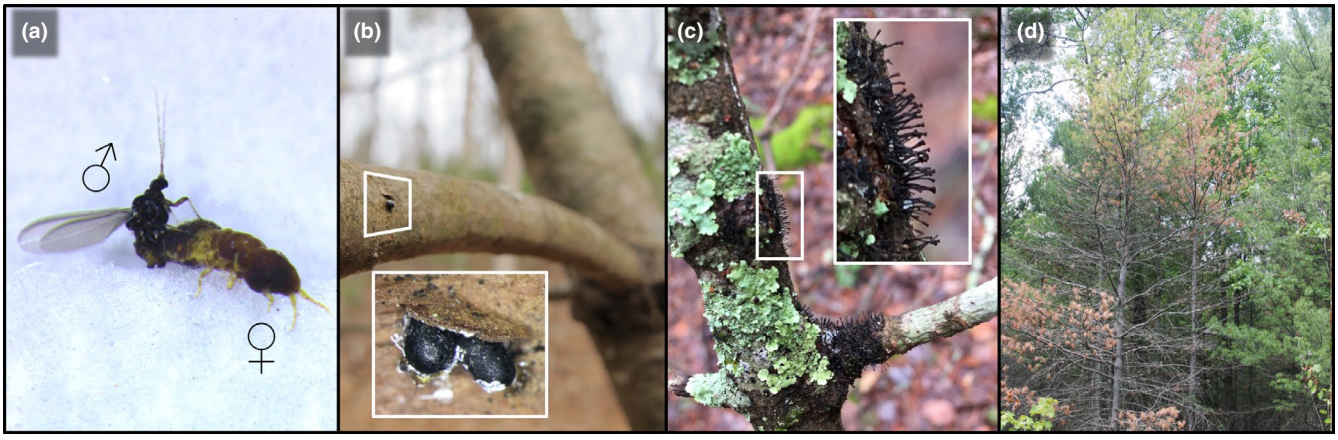


**FIGURE 1** The insect-pathogen complex associated with eastern white pine dieback. The eastern white pine bast scale (*Matsucoccus macrocicatrices*) is an insect with (a) a sexually dimorphic adult stage lasting mere weeks and (b) a second-instar cyst stage lasting 1–2 years. As a juvenile cyst, which resembles a small, black pearl, *M. macrocicatrices* will colonize and feed on tree sap within branch nodes, under lichen and in bark crevices (inset), where feeding wounds are hypothesized to facilitate infection by (c) *Caliciopsis pinea* (inset shows the characteristic "eyelash-like" fruiting bodies), which drive canker development and leads to bottom-up branch dieback and mortality (d). Photo credit: Joe O'Brien (USDA Forest Service, d)

pine bast scale (*Matsucoccus matsumurae* Kuwana), maritime pine bast scale (*Matsucoccus feytaudi* Ducasse) and Israeli pine bast scale (*Matsucoccus josephi* Bodenheimer et Harpaz; Bean & Godwin, 1971; Kerdelhúe, Boivin, & Burban, 2014; Mendel, 1998). In these cases, release from natural enemies and/or host defenses were attributed as the cause for invasion (Jactel et al., 2006; Mendel, 1998). Whether *M. macrocicatrices* has similarly expanded its range to enemy-free areas with naïve host provenances or has become pestiferous within its native range due to abiotic or biotic shifts, remains unknown.

Microsatellites are frequently used in population genetic studies to identify the origin of pest arthropods (e.g. Havill et al., 2016; Zemanova, Knop, & Heckel, 2016; Zhang, Edwards, Kang, & Fuller, 2012). As *M. macrocicatrices* is now well-established south of Massachusetts and west of Lake Erie, where no records existed prior to 2011 and 2015, respectively (Mech et al., 2013; Michigan Department of Natural Resources, 2015), we developed microsatellites de novo to learn if this insect species was new to regions outside its putative native range. We tested two competing hypotheses: (a) if populations of *M. macrocicatrices* established outside of its purported native range (New England) are the result of recent introduction(s) and colonization, then we expected to observe reductions in genetic diversity consistent with founder events. Alternatively, (b) if *M. macrocicatrices* has historically co-occurred with its host outside its purported native range, then we expected to observe similar levels of genetic diversity. This hypothesis assumes that, like its eastern white pine host, the insect existed in Southern Appalachian refugial populations during the last glacial maximum, recolonized northward as glaciers receded, and re-accumulated genetic diversity over thousands of years (Nadeau et al., 2015). Furthermore, we also expected to observe prominent population structure where geographical barriers, such as the Great Lakes, may have limited *M. macrocicatrices* gene flow over time.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection

We sampled 22 sites throughout the range of eastern white pine for *M. macrocicatrices* in the USA (Table 1). Immature cysts were collected between 2015 and 2018 during the winters and springs when the insects are near the end of their second-instar, relatively large (0.5–1.0 mm), and easiest to locate when sessile and embedded in tree bark (Figure 1a). Sampling occurred in one of two ways: (a) for 14 of the 22 sites, branches and stems of eastern white pine trees, sized 1–12 cm diameter at breast height (DBH), were shipped overnight to the University of Georgia (Athens, Georgia, USA; Table 1). Individual scale insects were then located with a stereo microscope and immediately preserved in 95% ethanol at −20°C. (b) For the remaining 8 of 22 sites, individual cysts were located and removed in situ from eastern white pine bark, preserved in 95% ethanol immediately, and stored at −20°C within 3 days of collection (Table 1). Sites were separated by ≥25 km. We sampled 11–20 individual *M. macrocicatrices* from between one and nine trees per site (referred to as population, hereafter). For one Michigan population ($n$ = 8) and the Wisconsin population ($n$ = 6), sampling was conducted in June 2018, narrowly after most of the insects had already moulted. We instead collected the voided cuticles (i.e. exoskeletons) in lieu of live cysts. Cuticles that produced adequate genomic DNA purity and yield were used for microsatellite analyses. Insects were confirmed as *M. macrocicatrices* by amplifying and sequencing the 28S barcode region (Appendix S1).

### 2.2 | Molecular analysis

#### 2.2.1 | DNA extraction

All DNA extractions of individual *M. macrocicatrices* cysts and cuticles were performed with the Qiagen DNEasy® Blood and Tissue Extraction Kit (Qiagen Inc.) following the manufacturer's protocol with two minor modifications: (a) we pierced each *M. macrocicatrices* cyst with a flame-sterilized insect pin and proceeded with overnight lysis, which allowed us to retain the cuticles for vouchering and still achieve adequate genomic DNA yield and (b) we decreased final DNA volumes for each sample to 100 µl total (two elution steps of 50 µl).

#### 2.2.2 | Microsatellite amplification

The microsatellite marker discovery procedure using shotgun sequence reads, as well as polymerase chain reaction conditions are detailed in Appendix S2. We developed 12 robust primer pairs specific to *M. macrocicatrices* (Table S2.1 in Appendix S2), which we amplified for all 390 total individuals. Amplicon sizes were determined on a 3,730 capillary sequencer (Applied Biosystems) at the Arizona State University DNA Core Lab using GeneScan LIZ 500 size standard (Applied Biosystems). Allele sizes were scored using the microsatellite plugin for GENEIOUS version 10.2.3 (Biomatters).

### 2.3 | Statistical analysis

#### 2.3.1 | Genetic diversity

All microsatellite loci for each population were tested for the following in GENEPOP version 4.2 (Raymond & Rousset, 1995): linkage disequilibrium with the probability test, deviations from Hardy–Weinberg equilibrium with exact tests, and null allele frequency with the Brookfield (1996) method. Genetic diversity was estimated using effective number of alleles ($A_E$), mean frequency of private alleles ($A_P$), mean number of locally common alleles (≥5%) occurring in ≤50% of populations ($A_{LC}$), observed heterozygosity ($H_O$) and unbiased expected heterozygosity ($H_E$) in GENALEX version 6.503 (Peakall & Smouse, 2006, 2012). Rarefied allelic richness ($A_R$) and inbreeding coefficients ($F_{IS}$) were calculated in the R package 'hierfstat' (Goudet, 2005). Generalized linear models were conducted in R version 3.5.1 (R Core Team, 2018) to evaluate the association between latitude and longitude with genetic diversity. Latitude and longitude were

**TABLE 1** Population information for microsatellite analysis of *Matsucoccus macrocicatrices*, including location data and estimates of genetic diversity

| Site ID | Site name | County, state | *N* | Latitude | Longitude | Elev. (m) | $A_N$ | $A_R$ | $A_E$ | $A_P$ | $A_{LC}$ | $F_{IS}$ | $H_O$ | $H_E$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GA1[a] | Mill Creek | Murray, GA | 12 | 34.8729 | −84.7236 | 287 | 4.56 | 2.88 | 2.70 | 0.00 | 2.44 | 0.02 | 0.63 | 0.64 |
| GA2[a] | Rock Creek | Gilmer, GA | 20 | 34.7806 | −84.3312 | 591 | 5.78 | 3.17 | 3.32 | 0.11 | 3 | 0.21 | 0.55 | 0.68 |
| GA3[b] | Boggs Creek | Lumpkin, GA | 20 | 34.7008 | −83.8860 | 558 | 6.11 | 2.97 | 2.87 | 0.11 | 3 | 0.04 | 0.61 | 0.64 |
| GA4[b] | Panther Creek | Habersham, GA | 20 | 34.6996 | −83.4195 | 454 | 4.78 | 2.76 | 2.73 | 0.00 | 2.33 | 0.08 | 0.56 | 0.61 |
| TN[a] | Tellico Plains | Monroe, TN | 20 | 35.3362 | −84.1471 | 573 | 6.56 | 3.16 | 3.51 | 0.00 | 3.33 | 0.14 | 0.58 | 0.66 |
| NC1[a] | Glenn Falls | Macon, NC | 20 | 35.0359 | −83.2352 | 1,113 | 4.78 | 2.73 | 2.54 | 0.00 | 2.67 | 0.05 | 0.56 | 0.59 |
| NC2[a] | Silver Mine | Madison, NC | 16 | 35.9101 | −82.7923 | 679 | 3.78 | 2.34 | 2.08 | 0.11 | 1.67 | 0.29 | 0.33 | 0.45 |
| VA1[a] | Price Ridge | Bland, VA | 20 | 37.1567 | −80.9578 | 710 | 5.44 | 2.77 | 3.28 | 0.11 | 2.22 | 0.15 | 0.45 | 0.53 |
| VA2[b] | Deerfield | Augusta, VA | 20 | 38.2034 | −79.3351 | 612 | 5.33 | 2.87 | 3.15 | 0.22 | 2.11 | 0.20 | 0.46 | 0.58 |
| VA3[b] | Falling Springs | Alleghany, VA | 19 | 38.1028 | −78.7652 | 784 | 5.44 | 2.88 | 3.13 | 0.11 | 2.44 | 0.07 | 0.54 | 0.58 |
| WV[b] | Watoga State Park | Pocahontas, WV | 20 | 38.1110 | −80.1076 | 840 | 4.11 | 2.29 | 2.46 | 0.00 | 1.56 | 0.05 | 0.38 | 0.40 |
| PA1[b] | Cook Forest State Park | Clarion, PA | 19 | 41.3535 | −79.2193 | 453 | 4.11 | 2.55 | 2.59 | 0.22 | 1.44 | 0.16 | 0.44 | 0.52 |
| PA2[b] | Rothrock State Forest | Centre, PA | 20 | 40.7764 | −77.6195 | 460 | 6.00 | 2.79 | 3.03 | 0.33 | 2.44 | 0.06 | 0.51 | 0.54 |
| PA3[b] | Tioga State Forest | Tioga, PA | 20 | 41.5272 | −77.4506 | 283 | 4.78 | 2.45 | 2.93 | 0.00 | 1.78 | 0.11 | 0.38 | 0.43 |
| NH1[b] | Bear Brook State Park | Merrimack, NH | 20 | 43.1645 | −71.3902 | 102 | 4.78 | 2.62 | 2.51 | 0.11 | 1.67 | 0.13 | 0.46 | 0.53 |
| NH2[b] | Litchfield State Forest | Hillsborough, NH | 20 | 42.8592 | −71.4603 | 65 | 5.89 | 2.69 | 2.87 | 0.11 | 2.67 | 0.12 | 0.44 | 0.49 |
| NH3[b] | Mast Yard State Forest | Merrimack, NH | 20 | 43.2392 | −71.6528 | 112 | 5.33 | 2.75 | 2.93 | 0.11 | 2.22 | 0.10 | 0.48 | 0.53 |
| ME[b] | University of Maine Forest | Penobscot, ME | 20 | 44.9307 | −68.6850 | 38 | 5.44 | 2.72 | 3.30 | 0.22 | 2 | 0.04 | 0.50 | 0.52 |
| MI1[b] | Grouse Trail | Crawford, MI | 11 | 44.7028 | −84.4086 | 332 | 3.56 | 2.45 | 2.35 | 0.00 | 1.56 | 0.10 | 0.45 | 0.50 |
| MI2[b] | Frederic | Crawford, MI | 19 | 44.7862 | −84.7387 | 382 | 3.89 | 2.27 | 2.03 | 0.00 | 1.56 | 0.15 | 0.40 | 0.47 |
| MI3[a] | Hiawatha National Forest | Delta, MI | 8 | 45.9095 | −86.8392 | 214 | 4.11 | 3.11 | 2.98 | 0.11 | 1.67 | 0.39 | 0.41 | 0.64 |
| WI[a] | Lake Owen | Bayfield, WI | 6 | 46.3243 | −91.2372 | 410 | 2.78 | 2.45 | 1.99 | 0.11 | 1.33 | 0.48 | 0.25 | 0.46 |

*Note:* $A_N$: mean number of alleles; $A_E$: effective number of alleles; $A_R$: rarefied allelic richness; $A_P$: number of private alleles per locus; $A_{LC}$: mean number of locally common alleles (≥5%) occurring in ≤50% of populations; $F_{IS}$: inbreeding coefficient; $H_O$: observed heterozygosity; $H_E$: unbiased expected heterozygosity.

USA states: GA = Georgia; ME = Maine; MI = Michigan; NC = North Carolina; NH = New Hampshire; PA = Pennsylvania; TN = Tennessee; VA = Virginia; WI = Wisconsin; WV = West Virginia.

[a]Sampled from shipped wood material.

[b]Sampled in situ.

included in the models simultaneously as covariates. To specifically test for clinical decays in genetic diversity according to geographical distance from the putative native range, we created a Euclidean distance (km) matrix among all populations from one population in New Hampshire (NH1). We assigned a value of 0 km to the four populations located within the insect's putative native range (NH1, NH2, NH3 and ME). Regressions were then conducted on the indices as stated above. All models included the number of genotyped individuals as a covariate to control for uneven population sizes. All indices met the assumptions of normality except for $A_{R,}$ $F_{IS}$ and $A_{LC}$, which were log-transformed.

The program BOTTLENECK (Piry, Luikart, & Cornuet, 1999) was used to detect if signals of recent bottleneck/founder event(s) existed within in our dataset. This program tests for deviations from mutation-drift equilibrium with the assumption that allelic richness decreases faster than heterozygosity in shrinking populations. All 22 populations were tested separately, permuted 1,000 times. We used the single-step mutation model (SMM) and the two-phase model (TPM) with 95% single-step mutations and 5% multi-step mutations (Piry et al., 1999). Significant excesses in heterozygosity for each population were determined with the one-tailed Wilcoxon signed-rank test.

### 2.3.2 | Population structure

We used the Bayesian clustering algorithm STRUCTURE version 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) to infer subgroup assignments for *M. macrocicatrices*. For all simulations we did not use a location prior, and we assumed an admixture model with allele frequencies correlated among groups (Falush, Stephens, & Pritchard, 2003). Each run utilized 25,000 burn-in, followed by 50,000 Markov Chain Monte Carlo (MCMC) iterations, replicated 20 times for each number of clusters assumed (*K*).

Hierarchical groupings of individual *M. macrocicatrices* were simulated in separate STRUCTURE runs as follows: (1) the entire dataset of 22 populations, with *K* ranging from 1 to 22; (2) simulations to evaluate substructure within resulting major clusters, including (2a) the Southern Appalachians ("*SApps*"; 7 populations in Georgia, Tennessee and North Carolina) with *K* = 1 through 7, (2b) the Northeast ("*NEast*"; 11 populations in Virginia, West Virginia, Pennsylvania, New Hampshire, and Maine) with *K* = 1 through 11, and (2c) the Great Lakes ("*GLakes*"; 4 populations in Michigan and Wisconsin) with *K* = 1 through 5. Optimal *K*, or the most likely number of clusters for each grouping, was determined by the Evanno, Regnaut, and Goudet (2005) method implemented in STRUCTURE HARVESTER (Earl & Vonholdt, 2012). Populations were assigned to the cluster with the highest corresponding mean posterior probability of ancestry.

We also inferred optimal population structure from analyses of molecular variance (AMOVA) using ARLEQUIN version 3.5 (Excoffier & Lischer, 2010) to determine the hierarchical partitioning of genetic variance using pre-defined population structure from STRUCTURE results. We conducted six AMOVAs with 10,000 permutations to test: (a) no genetic structure, (b) genetic structure where *K* = 2, (c) genetic structure where *K* = 3, (d) *NEast* populations only, (e) *SApps* populations only, and (f) *GLakes* populations only.

We performed multiple principal coordinates analyses (PCoA) using Nei's unbiased genetic distances (Nei, 1978) in GENALEX. We also calculated pairwise $F_{ST}$ (Weir & Cockerham, 1984) and Slatkin's (1995) linearized pairwise $F_{ST}$ values in ARLEQUIN to evaluate genetic differentiation between populations. The pairwise $F_{ST}$ matrix was used to generate an unrooted neighbour-joining tree with the 'neighbour' package in PHYLIP version 3.695 (Felsenstein, 1989). The linearized $F_{ST}$ matrix and a pairwise matrix of log-transformed geographical distances (km) were also used in a Mantel test (Mantel, 1967) to detect isolation-by-distance (IBD). Mantel tests may falsely detect IBD in instances of hierarchical structure with distinct barriers to gene flow (Meirmans, 2012), so we conducted additional Mantel tests within clusters. We also performed a partial Mantel test controlling for cluster assignment with a covariate matrix containing binary values for each pairwise relationship: 0 for pairs of populations belonging to the same cluster and 1 for those belonging to separate clusters. In another partial Mantel test, we examined the association between genetic distance and cluster assignment, using the geographical distance matrix as a covariate. All Mantel and partial Mantel tests were performed with 100,000 permutations in the R package 'vegan' (Oksanen et al., 2018).

Based on the finding of a potential barrier to gene flow existing in the Blue Ridge mountains (see Section 3), we further evaluated the link between eastern white pine density and the genetic connectivity of *M. macrocicatrices* by assessing the least-cost paths between populations in Georgia, North Carolina, Tennessee, Virginia, and West Virginia. We sought to test tree-host connectivity through the application of circuit theory (McRae & Beier, 2007) for the purpose of comparing pairwise genetic distances but not for modelling gene flow. Remote sensing data of eastern white pine from forest inventory and analysis (FIA, USDA Forest Service) were used to create a relative density raster in R where each pixel (size = 250 m) holds a value equal to the percentage of eastern white pine comprising the total composition of trees ≥12.7 cm DBH. We created a cost-surface raster with each pixel holding a resistance value based on its corresponding tree density value. Pixels with 0% eastern white pine were assigned a high resistance value of 200 and all other pixels were assigned resistance values of 1–100, inversely proportional to their relative density of host trees (100%–1%). We assessed the least-cost paths of the cost-surface raster between the 11 populations adjacent to the Blue Ridge geographical barrier in the R package 'gdistance' (van Etten, 2017). We conducted another partial Mantel test, controlling for cluster assignment as above, to assess the correlation between linearized $F_{ST}$ and pairwise least-cost distance.

## 3 | RESULTS

### 3.1 | Microsatellite loci quality

All loci had null allele rates of less than 0.1 averaged across all populations except for three, which we then removed from all further

analyses (Table S2.1 in Appendix S2). Of the 198 locus-population combinations, exact tests revealed a significant departure from Hardy–Weinberg equilibrium in 47 pairs, but with no clear concentration in any particular locus or population. None of the nine remaining loci showed significant linkage disequilibrium. Every individual *M. macrocicatrices* had a unique multilocus genotype and no individual was homozygous across all loci, which negates the possibility *M. macrocicatrices* has haplodiploid sex-determination.

## 3.2 | Genetic diversity

Estimated indices of genetic diversity for each population ($A_E$, $A_R$, $A_P$, $A_{LC}$, $F_{IS}$, $H_O$ and $H_E$) are summarized in Table 1. Our analyses found some evidence of latitudinal, but not longitudinal clines, with mean number of locally common alleles ($A_{LC}$; $R^2 = 0.41$, $df = 18$, $t = -2.78$, $p = .01$) and observed heterozygosity ($H_O$; $R^2 = 0.33$, $df = 18$, $t = -2.57$, $p = .02$) both decreasing as latitude increased. Linear models showed no correlation between distance from the putative native range (the four populations in New England) and genetic diversity, except for $A_{LC}$, which as observed to increase with distance from our New Hampshire reference population (NH1; $R^2 = 0.25$, $df = 19$, $t = 2.12$, $p = .047$).

No signatures of bottleneck were detected in any of the populations we sampled according to one-tailed Wilcoxon tests (Table S3.1 in Appendix S3). Heterozygote deficiency, however, which can be indicative of population expansion, was detected in 13 of the 22 populations and in all three pooled groups according to two-tailed Wilcoxon tests for at least one of the two models (SMM and TPM).

## 3.3 | Population structure

### 3.3.1 | Range-wide population structure

The optimal number of clusters for all *M. macrocicatrices* ($N = 390$) in the USA was $K = 3$ (Figure 2a). Although $\Delta K$ indicated optimal $K = 2$ from STRUCTURE results, the plateau in the ln Pr($X|K$) curve was strongest when $K = 3$ (Figure S3.2 in Appendix S3). Furthermore, the clear clustering visualized in the PCoA (Figure 3), as well as results from AMOVAs (Table 2: tests A–C), Mantel tests, and pairwise $F_{ST}$ (Table 3), all support a three-cluster model ($K = 3$). The three clusters were regionally distinct and were designated as follows: populations located in (a) Virginia and northward were defined as the "*NEast*" cluster, (b) those located in North Carolina and southward were defined as "*SApps*" cluster, and (c) those from Michigan and Wisconsin were defined as "*GLakes*" (Figure 2a).

AMOVAs resulted in significant genetic differentiation between defined groups (Table 2). Range-wide tests using cluster assignment priors inferred from STRUCTURE accounted for more overall differentiation ($F_{ST} = 0.325$ for $K = 2$, 0.324 for $K = 3$) than the test assuming no population structure ($F_{ST} = 0.253$ for $K = 1$). The two-cluster and three-cluster AMOVA tests resulted in nearly the same $F_{ST}$, but genetic variation was partitioned differently. Twice as much variation was partitioned between groups ($F_{CT} = 0.235$) as among populations within groups ($F_{SC} = 0.116$) for the three-cluster model, whereas there was only 1.3 times as much genetic variation partitioned between groups ($F_{CT} = 0.201$) as among populations within groups ($F_{SC} = 0.155$) for the two-cluster model. Standard and partial mantel tests revealed that the association between overall pairwise genetic and geographical distance was significant ($R = .509$, $p < .001$) but not when controlling for clustering assignments ($R = .046$, $p < .30$). There was also an association between genetic distance and cluster assignment, controlling for genetic distance ($R = .653$, $p < .001$).

Corresponding with the three-cluster model, pairwise $F_{ST}$ revealed abrupt changes in genetic distance (a) isolating all four *GLakes* populations from the rest of the dataset, and (b) dividing the *NEast* and *SApps* populations in the Blue Ridge Mountains (Table 3). Host-tree connectivity could not explain the presence of this second potential barrier to gene flow, as pairwise $F_{ST}$ between populations in Georgia, North Carolina, Tennessee, Virginia and West Virginia was not associated with pairwise least-cost distance based on eastern white pine density (Table S3.2 in Appendix S3), according to a partial Mantel test controlling for cluster assignment ($R = .179$, $p = .882$).

The pooled population heterozygosity of the *SApps* individuals was highest among the three clusters ($H_O = 0.549$, $H_E = 0.700$), followed by *NEast* ($H_O = 0.458$, $H_E = 0.567$), and lastly *GLakes* individuals ($H_O = 0.397$, $H_E = 0.528$). There was significant genetic distance between each of the clusters, as informed by pairwise $F_{ST}$ values (Table S3.3 in Appendix S3). The neighbour-joining tree (Figure S3.4 in Appendix S3) grouped the NC2 population with the *NEast* populations, but otherwise corroborated the three-cluster model. The *NEast* and *GLakes* populations shared a node, which was joined with the remaining populations, indicating they arose from a common *SApps* ancestor.

### 3.3.2 | Regional population structure

Results from within-cluster STRUCTURE runs are shown in Figure 2b. Within the *NEast* cluster, we found the optimal $K = 2$ (Figure S3.3 in Appendix S3), where the three Virginia and single West Virginia populations comprised one subgroup and populations from Pennsylvania, New Hampshire and Maine comprised the other subgroup. Within the *SApps* cluster, we found the optimal $K = 3$ (Figure S3.3 in Appendix S3), with the first subgroup consisting of the three populations in northeastern Georgia and western North Carolina (populations GA3, GA4 and NC1), the second subgroup consisting of the three populations in the northwestern Georgia and southeastern Tennessee (populations: GA1, GA2 and TN), and the third subgroup consisting solely of individuals from the population NC2. Within the *GLakes* cluster, we found the optimal $K = 4$ (Figure S3.3 in Appendix S3), but every population appeared to be of a mostly mixed ancestry.

The AMOVAs conducted within each cluster revealed that genetic differentiation was partitioned similarly and was overall comparable in both the *NEast* cluster ($F_{ST} = 0.150$, $F_{CT} = 0.091$, $F_{SC} = 0.065$) and the *SApps* cluster ($F_{ST} = 0.158$, $F_{CT} = 0.095$, $F_{SC} = 0.069$). The *GLakes* cluster had comparatively lower genetic differentiation

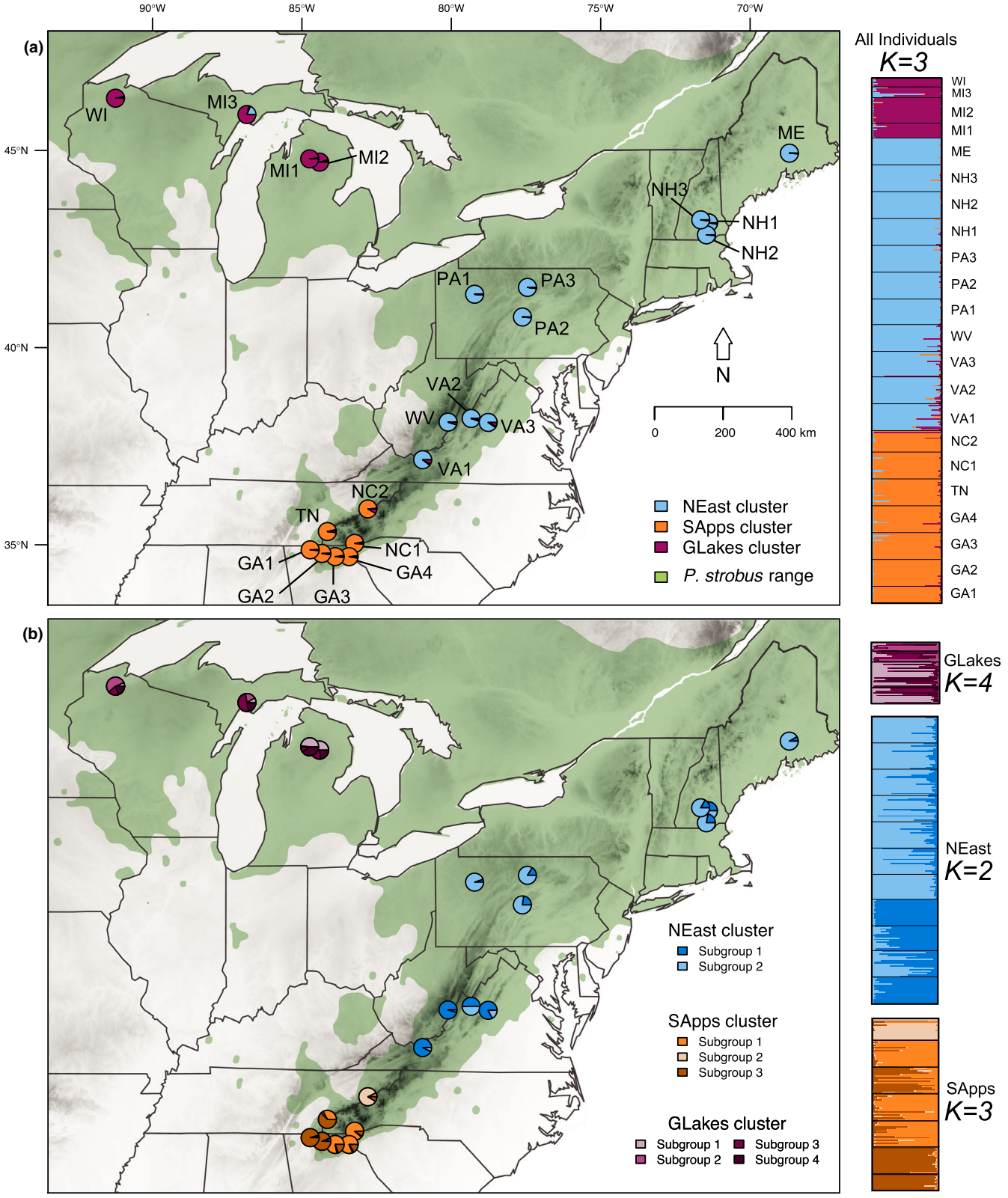


**FIGURE 2** STRUCTURE results for *Matsucoccus macrocicatrices* sampled across its USA range: (a) run including all samples; (b) runs including only populations from each of the three inferred clusters: "*NEast*" (Maine, New Hampshire, Pennsylvania, Virginia, and West Virginia), "*SApps*" (Georgia, North Carolina, and Tennessee), and "*GLakes*" (Michigan and Wisconsin)

($F_{ST}$ = 0.071). Among pairwise populations, $F_{ST}$ values were all statistically significant except for four: two among New Hampshire populations and two among Michigan populations (Table 3). The NC2 population from North Carolina was the most highly differentiated, with pairwise $F_{ST}$ values ranging from 0.230 to 0.533. Mantel tests revealed significant IBD within each cluster (*NEast*: $R$ = .335,

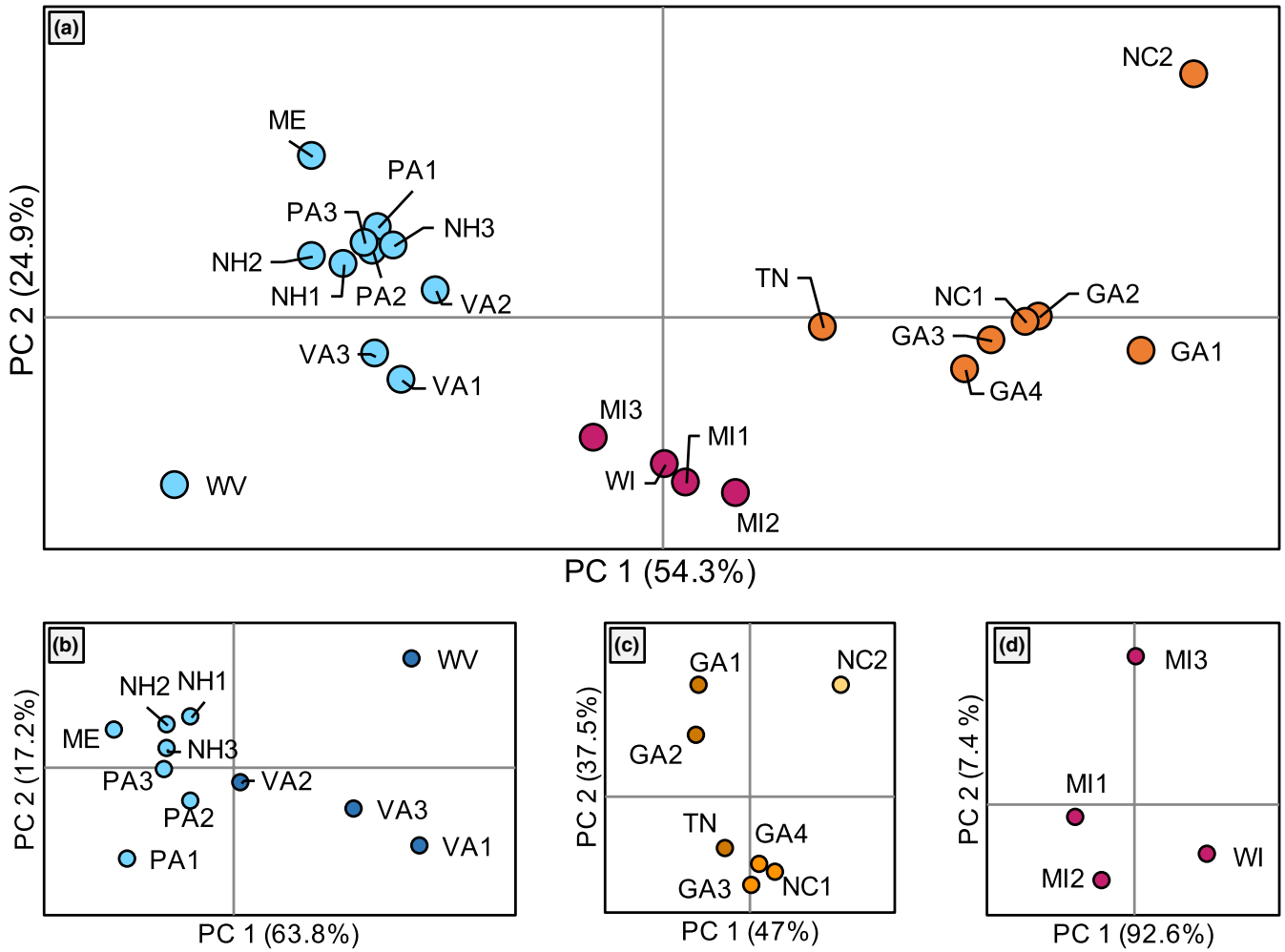


**FIGURE 3** Principal coordinates analysis (PCoA) based on Nei's unbiased genetic distances of (a) all *Matsucoccus macrocicatrices* populations sampled in the USA, (b) "*NEast*" populations (Maine, New Hampshire, Pennsylvania, Virginia, and West Virginia), (c) "*SApps*" populations (Georgia, North Carolina, and Tennessee), and (d) "*GLakes*" populations (Michigan and Wisconsin)

$p$ = .005; *SApps*: $R$ = .833, $p$ < .001; and *GLakes*: $R$ = .835, $p$ = .04; Figure S3.1 in Appendix S3).

## 4 | DISCUSSION

Non-native and native species that become pestiferous often do so by escaping different evolutionary constraints, such as exploiting an enemy-free or defense-free space. Determining whether *M. macrocicatrices* is new outside its purported native range in New England may offer perspective into its sudden association with novel dieback symptoms and mortality of its host tree. Based on evidence presented herein, we propose the insect is native throughout the North American range of eastern white pine and that the two organisms have likely co-occurred since the last glacial maximum.

### 4.1 | Evidence for nativity

The genetic landscape of *M. macrocicatrices* was not consistent with that of an exotic species recently introduced to a new range. Source populations are usually genetically rich, whereas founder populations are usually genetically depauperate (Nei, Maruyama, & Chakraborty, 1975). We found high levels of global genetic diversity (e.g., $H_E$ = 0.43–0.68), especially when compared to a congener, *M. feytaudi*, where there are both source populations and recent, non-native founder populations in Europe (e.g., $H_E$ = 0.25–0.58; Kerdelhúe et al., 2014). There was no evidence of a recent bottleneck in any population, nor was there a longitudinal cline in genetic diversity despite the most easterly populations being within the purported native range (New Hampshire and Maine). We also did not find the expected decay in genetic diversity when assessing Euclidean distance from these populations. In fact, the most southerly populations tended to be the most genetically rich, as both the mean number of locally common alleles and the observed heterozygosity per population were negatively associated with latitude. Glacial history may provide some context to our resulting patterns. Both palynological and molecular phylogeographical evidence indicate that refugial populations of eastern white pine survived in the mid-Atlantic and at the southernmost portion of the Appalachian Mountain range during the last glacial

**TABLE 2** Analyses of molecular variance for *Matsucoccus macrocicatrices*

| Test | Group structure | Source of variation | Degrees of freedom | Sum of squares | Variance components | Percent variation | Fixation indices |
|---|---|---|---|---|---|---|---|
| | Range-wide | | | | | | |
| A | No structure | Among populations | 21 | 632 | 0.78 | 25.3 | $F_{ST}$ = **0.253** |
| | | Within populations | 758 | 1,760 | 2.32 | 74.7 | |
| | | Total | 779 | 2,392 | | | |
| B | *K* = 2 | Between groups | 1 | 284 | 0.69 | 20.1 | $F_{CT}$ = **0.201** |
| | | Among populations | 20 | 348 | 0.43 | 12.4 | $F_{SC}$ = **0.155** |
| | | Within populations | 758 | 1,760 | 2.32 | 67.5 | $F_{ST}$ = **0.325** |
| | | Total | 779 | 2,392 | 3.44 | | |
| C | *K* = 3 | Between groups | 2 | 381 | 0.81 | 23.4 | $F_{CT}$ = **0.235** |
| | | Among populations | 19 | 250 | 0.3 | 8.9 | $F_{SC}$ = **0.116** |
| | | Within populations | 758 | 1,760 | 2.32 | 67.7 | $F_{ST}$ = **0.324** |
| | | Total | 779 | 2,392 | 3.43 | | |
| | Within cluster | | | | | | |
| D | *NEast* | Among subgroups | 1 | 56 | 0.24 | 9.1 | $F_{CT}$ = **0.091** |
| | | Among populations | 9 | 75 | 0.15 | 5.9 | $F_{SC}$ = **0.065** |
| | | Within populations | 425 | 937 | 2.21 | 85 | $F_{ST}$ = **0.150** |
| | | Total | 435 | 1,068 | 2.6 | | |
| E | *SApps* | Among subgroups | 2 | 65 | 0.3 | 9.5 | $F_{CT}$ = **0.095** |
| | | Among populations | 4 | 40 | 0.2 | 6.2 | $F_{SC}$ = **0.069** |
| | | Within populations | 249 | 660 | 2.65 | 84.3 | $F_{ST}$ = **0.158** |
| | | Total | 255 | 765 | 3.15 | | |
| F[a] | *GLakes* | Among populations | 3 | 15 | 0.15 | 7.1 | $F_{ST}$ = **0.071** |
| | | Within populations | 84 | 163 | 1.94 | 92.9 | |
| | | Total | 87 | 178 | 2.09 | | |

Significant *F*-statistics are bold ($p < .05$).
[a]No priors were set, because the *K* = 4 result from STRUCTURE implied a largely mixed ancestry among populations.

maximum (Davis, 1983; Nadeau et al., 2015). If *M. macrocicatrices*, being obligate on its host tree, co-occurred during northward recolonization following glacial thaw, then the southernmost populations would likely have retained more ancestral genetic variation (Hewitt, 1999).

The patterns of genetic differentiation also failed to substantiate that *M. macrocicatrices* is non-native outside of New England. In addition to restricted genetic exchange and genetic drift, sufficient time is required for populations to differentiate, and thus, our results consistently suggest *M. macrocicatrices* is well-established within its entire, current distribution. Populations were highly structured overall and delineated into three distinct, regional and genetic clusters: *NEast*, *SApps and GLakes*. STRUCTURE had high support for clustering *SApps* and *GLakes* together, but with only 44 total individuals analyzed from four populations in the Great Lakes region, uneven sampling may have influenced the results. STRUCTURE analysis tends to merge distinct, but small, subpopulations together when sampling is biased (Puechmaille, 2016). However, results from AMOVA (Table 2) and pairwise genetic distances (Table 3) strongly suggest *SApps* and *GLakes* are separate groups.

## 4.2 | Barriers to gene flow

The limited dispersal ability of *M. macrocicatrices* may help to explain their overall high genetic differentiation. The main dispersal stage for this species is the first-instar "crawler", capable of walking short distances and being wind-dispersed longer distances (Costanza et al., 2018). Congeners can stay airborne for up to 0.5 km (Bean & Godwin, 1955) and could theoretically exceed 85 km in passive flight given optimal conditions (Hanks & Denno, 1998). Arthropods evolving in heterogenous landscapes with frequent patches of unsuitable habitat tend to avoid passive, aerial dispersal, due to the high risk of mortality (Bonte et al., 2012; Bonte, Vandenbroecke, Lens, & Maelfait, 2003). *Matsucoccus* spp. are no different, with <20% of individuals observed to disperse from their natal trees (McClure, 1977; Stephens & Aylor, 1978; Unruh & Luck, 1987). As a host tree, eastern white pine only occasionally grows in pure stands and is more commonly found as a highly scattered super-canopy tree (Abrams, 2001). We found evidence of IBD within each genetic cluster according to Mantel tests, indicating that long-distance dispersal of *M. macrocicatrices* between patches is rare (McClure, 1976).

**TABLE 3** Population pairwise $F_{ST}$ values for *Matsucoccus macrocicatrices*

| Population | GA1 | GA2 | GA3 | GA4 | TN | NC1 | NC2 | VA1 | VA2 | VA3 | WV | PA1 | PA2 | PA3 | NH1 | NH2 | NH3 | ME | MI1 | MI2 | MI3 | WI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GA1 | | | | | | | | | | | | | | | | | | | | | | |
| GA2 | 0.027 | | | | | | | | | | | | | | | | | | | | | |
| GA3 | 0.151 | 0.105 | | | | | | | | | | | | | | | | | | | | |
| GA4 | 0.168 | 0.118 | 0.028 | | | | | | | | | | | | | | | | | | | |
| TN | 0.102 | 0.082 | 0.053 | 0.056 | | | | | | | | | | | | | | | | | | |
| NC1 | 0.185 | 0.153 | 0.049 | 0.091 | 0.091 | | | | | | | | | | | | | | | | | |
| NC2 | 0.264 | 0.253 | 0.242 | 0.230 | 0.251 | 0.232 | | | | | | | | | | | | | | | | |
| VA1 | 0.326 | 0.288 | 0.285 | 0.259 | 0.253 | 0.311 | 0.418 | | | | | | | | | | | | | | | |
| VA2 | 0.306 | 0.273 | 0.259 | 0.260 | 0.209 | 0.282 | 0.352 | 0.112 | | | | | | | | | | | | | | |
| VA3 | 0.316 | 0.275 | 0.271 | 0.263 | 0.233 | 0.302 | 0.402 | 0.055 | 0.058 | | | | | | | | | | | | | |
| WV | 0.409 | 0.371 | 0.393 | 0.389 | 0.335 | 0.417 | 0.533 | 0.135 | 0.161 | 0.125 | | | | | | | | | | | | |
| PA1 | 0.323 | 0.306 | 0.320 | 0.325 | 0.258 | 0.351 | 0.388 | 0.206 | 0.079 | 0.151 | 0.295 | | | | | | | | | | | |
| PA2 | 0.330 | 0.299 | 0.304 | 0.308 | 0.252 | 0.334 | 0.383 | 0.143 | 0.032 | 0.093 | 0.210 | 0.042 | | | | | | | | | | |
| PA3 | 0.412 | 0.356 | 0.362 | 0.376 | 0.308 | 0.388 | 0.476 | 0.212 | 0.072 | 0.167 | 0.269 | 0.079 | 0.035 | | | | | | | | | |
| NH1 | 0.331 | 0.302 | 0.312 | 0.325 | 0.254 | 0.334 | 0.397 | 0.171 | 0.041 | 0.110 | 0.184 | 0.080 | 0.024 | 0.055 | | | | | | | | |
| NH2 | 0.361 | 0.321 | 0.327 | 0.348 | 0.275 | 0.350 | 0.428 | 0.190 | 0.044 | 0.118 | 0.213 | 0.064 | 0.025 | 0.020 | 0.001* | | | | | | | |
| NH3 | 0.325 | 0.287 | 0.302 | 0.315 | 0.252 | 0.330 | 0.387 | 0.177 | 0.044 | 0.114 | 0.212 | 0.056 | 0.027 | 0.024 | 0.017 | 0.007* | | | | | | |
| ME | 0.336 | 0.293 | 0.335 | 0.356 | 0.293 | 0.363 | 0.393 | 0.229 | 0.100 | 0.168 | 0.272 | 0.078 | 0.063 | 0.063 | 0.047 | 0.038 | 0.032 | | | | | |
| MI1 | 0.301 | 0.297 | 0.259 | 0.248 | 0.222 | 0.311 | 0.434 | 0.254 | 0.263 | 0.247 | 0.314 | 0.360 | 0.336 | 0.426 | 0.332 | 0.370 | 0.348 | 0.400 | | | | |
| MI2 | 0.301 | 0.303 | 0.281 | 0.277 | 0.242 | 0.327 | 0.438 | 0.290 | 0.307 | 0.296 | 0.350 | 0.382 | 0.372 | 0.445 | 0.365 | 0.397 | 0.375 | 0.423 | 0.016* | | | |
| MI3 | 0.211 | 0.208 | 0.198 | 0.207 | 0.165 | 0.253 | 0.377 | 0.161 | 0.170 | 0.146 | 0.257 | 0.246 | 0.222 | 0.359 | 0.226 | 0.264 | 0.249 | 0.281 | 0.076 | 0.055 | | |
| WI | 0.295 | 0.247 | 0.229 | 0.238 | 0.212 | 0.259 | 0.432 | 0.176 | 0.211 | 0.183 | 0.307 | 0.309 | 0.288 | 0.401 | 0.289 | 0.309 | 0.288 | 0.336 | 0.181 | 0.123 | 0.042* | |
| Mean | 0.275 | 0.246 | 0.241 | 0.247 | 0.209 | 0.272 | 0.367 | 0.222 | 0.173 | 0.195 | 0.293 | 0.224 | 0.196 | 0.253 | 0.200 | 0.213 | 0.199 | 0.233 | 0.285 | 0.303 | 0.208 | 0.255 |

All pairwise relationships showed significant differentiation ($p < .05$) except those indicated with "*".

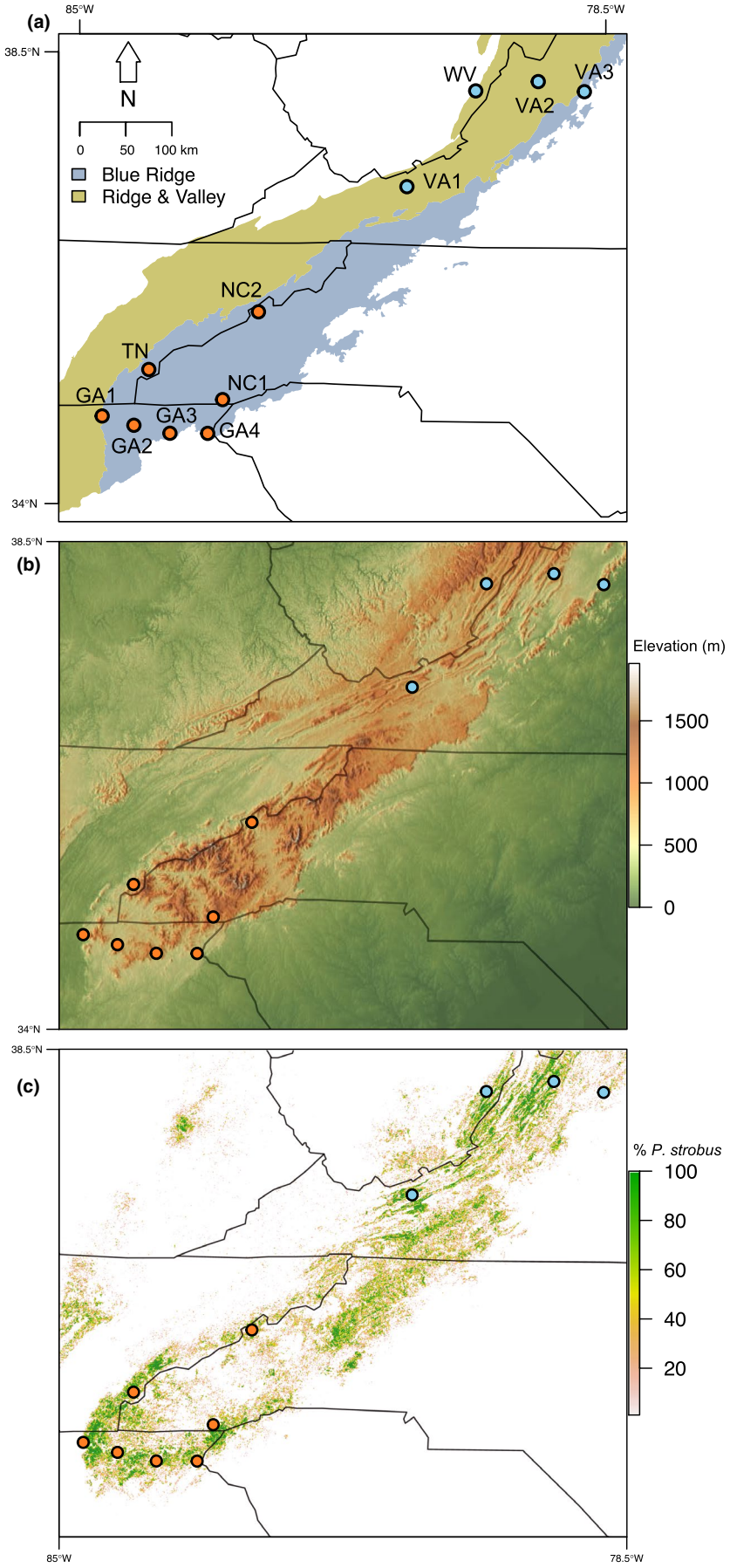


**FIGURE 4** Abiotic variables that may influence the barrier to *Matsucoccus macrocicatrices* gene flow located in the Blue Ridge mountains, USA, including (a) Level III ecoregions (U.S. Environmental Protection Agency, 2013), (b) elevation and (c) host-tree density using remote sensing data, where each pixel indicates the percentage of eastern white pine compared to total tree species for individuals ≥12.7 cm DBH, at 250-m resolution (FIA, USDA Forest Service)

Between the three clusters, there were sudden increases in genetic distance not simply explained by geographical distance. We identified two main barriers to gene flow that were likely responsible. One barrier isolates the *GLakes* cluster, suggesting the Great Lakes act as a physical barrier to successful dispersal. Large water bodies present a high risk of mortality for passive, aerial dispersed arthropods and can lead to vicariance (Hawes, Worland, Convey, & Bale, 2007; Kuntner & Agnarsson, 2011). Genetic divergence between USA populations located in the Great Lakes and the northeastern states have also been observed in active-dispersing terrestrial animals (e.g. Bagley, Sousa, Niemiller, & Linnen, 2017; Hapeman, Latch, Rhodes, Swanson, & Kilpatrick, 2017).

The second barrier to *M. macrocicatrices* gene flow is located in between North Carolina (population NC2) and Virginia (population VA1) where the Blue Ridge Mountains and the Ridge and Valley ecoregions (U.S. Environmental Protection Agency, 2013) meet (Figure 4a). Population NC2 was the most genetically isolated in our study, perhaps because it lies in the French Broad River basin in North Carolina with imposing mountains to its southwest and northeast. It is noteworthy that the least-cost path (based on eastern white pine density) from this population to the nearest-neighbour *SApps* populations is roughly equal to its least-cost path to the southernmost *NEast* population (VA1). However, genetic distance was much greater between *M. macrocicatrices* in North Carolina and Virginia. Unique features of the area near the North Carolina-Virginia border, other than host-tree density, must therefore be contributing to the restriction of gene flow. Geological attributes of the Blue Ridge Mountains, such as their irregularity and precipitous changes in elevation from 450 m to over 2,000 m (Figure 4b), may be factors contributing to hindered gametic exchange. Significant genetic structure in the Blue Ridge, and especially among populations on either side of the French Broad River in North Carolina, has been observed in several other taxa such as snakes (Fontanella, Feldman, Siddall, & Burbrink, 2008), salamanders (Crespi, Rissler, & Browne, 2003), centipedes (Garrick, Newton, & Worthington, 2018), and harvestmen (Hedin & McCormack, 2017). It stands to reason that the terrain would also make it difficult for passive, wind-dispersed animals to be carried freely among suitable habitat patches of its host. The long, parallel mountains within the Ridge and Valley ecoregion channel wind along their axes (Whiteman & Doran, 1993), whereas in the Blue Ridge ecoregion, prevailing winds travel perpendicular to mountain axes (Raichle & Carson, 2009). Thus, impeding winds and irregular terrain, rather than just host plant distribution and density (Figure 4c), may explain why the Blue Ridge Mountains appear to significantly impede gene flow in the area between the *NEast* and *SApps* clusters.

## 5 | CONCLUSIONS

With high range-wide genetic diversity, no signatures of recent founder events and clear genetic clusters separated by distinct geographical barriers, we reject the hypothesis that *M. macrocicatrices* is a non-native invader within the North American range of eastern white pine. Hence, host trees currently experiencing dieback symptoms and mortality have likely co-evolved with this insect. Its small size, sessile nature and seemingly benign impacts probably allowed it to remain undetected until the recent emergence of eastern white pine dieback symptoms in the mid-2000s. Costanza et al. (2018) reviews several ecological disturbance factors contributing to this phenomenon—such as climate change, land use, site conditions and forest management—which may be contributing to sudden *M. macrocicatrices* population growth. Assumed to be just one of over 250 innocuous herbivores of eastern white pine (Wendel & Smith, 1990), it currently remains a mystery why this native species has recently been associated with severe tree injury and mortality. The pathogenic fungus thought to exploit *M. macrocicatrices* feeding wounds and drive canker formation, *C. pinea*, is also native (Ray, 1936). This system presents a unique opportunity to understand how a native insect–pathogen complex, perhaps nonexistent or rare in the past, can become a transregional forest health concern. Excluding the possibility of a non-native invasion narrows the search for why and how a species might become pestiferous. Our work has demonstrated the utility of establishing the origin of a pest in guiding ecosystem conservation.

### ACKNOWLEDGEMENTS

We are especially grateful to those who assisted in sampling for this project: Kim Adams (SUNY-ESF), Arya Aghdasi (University of Georgia), Brittany Barnes (University of Georgia), Aaron Bergdahl (Maine Forest Service), Joe Bither (Maine Forest Service), Allison Kanoti (Maine Forest Service), Kyle Lombard (New Hampshire Division of Forests and Lands), Kaitlin Mooneyham (Virginia Department of Forestry), Joe O'Brien (USDA Forest Service), Jill Rose (West Virginia Department of Agriculture), Ashley Schulz (Arkansas State University), Tim Tomon (Pennsylvania DCNR), Jen Weimer (New Hampshire DNCR), and Jiangming Yao (Guangxi University). We also thank Nathan Havill (USDA Forest Service), Troy Kieran (University of Georgia) and Joseph Nairn (University of Georgia) for assistance with microsatellite development and Lyn Cook (University of Queensland) and Penny Gullan (Australian National University) for suggestions on DNA extraction. Nathan Havill, Joe Nairn and Kenneth Ross provided helpful comments on a previous version of this paper. This research was supported by the USDA Forest Service, Southern Research Station (13-CA-11330129-056 and 16-CS-11330129-045), Southern Region (8)-Forest Health Protection, USDA Agriculture and Food Research Initiative, Foundation Grant, and D.B. Warnell School of Forestry and Natural Resources, University of Georgia.

### DATA AVAILABILITY STATEMENT

Microsatellite genotype data can be found in the Table S2.2 in Appendix S2.

## ORCID

*Thomas D. Whitney* https://orcid.org/0000-0003-3359-9964

## REFERENCES

Abrams, M. D. (2001). Eastern white pine versatility in the presettlement forest. *BioScience*, *51*(11), 967–979. https://doi.org/10.1641/0006-3568(2001)051[0967:ewpvit]2.0.co;2

Asaro, C., Chamberlin, L. A., Rose, J. A., Mooneyham, K., & Rose, A. K. (2018). Mortality of eastern white pine (*Pinus strobus* L.) in association with a novel scale insect-pathogen complex in Virginia and West Virginia. *Forest Ecology and Management*, *423*, 37–48. https://doi.org/10.1016/j.foreco.2017.12.032

Bagley, R. K., Sousa, V. C., Niemiller, M. L., & Linnen, C. R. (2017). History, geography and host use shape genomewide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, *26*(4), 1022–1044. https://doi.org/10.1111/mec.13972

Bean, J. L., & Godwin, P. A. (1955). Description and bionomics of a new red pine scale, *Matsucoccus resinosae*. *Forest Science*, *1*(2), 164–176.

Bean, J. L., & Godwin, P. A. (1971). Red Pine Scale. *Forest Pest Leaflet*, *10*, 1–6.

Bonte, D., Van Dyck, H., Bullock, J. M., Coulon, A., Delgado, M., Gibbs, M., … Travis, J. M. J. (2012). Costs of dispersal. *Biological Reviews*, *87*(2), 290–312. https://doi.org/10.1111/j.1469-185X.2011.00201.x

Bonte, D., Vandenbroecke, N., Lens, L., & Maelfait, J. P. (2003). Low propensity for aerial dispersal in specialist spiders from fragmented landscapes. *Proceedings of the Royal Society B-Biological Sciences*, *270*(1524), 1601–1607. doi:https://doi.org/10.1098/rspb.2003.2432.

Brookfield, J. F. Y. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, *5*(3), 453–455. https://doi.org/10.1046/j.1365-294X.1996.00098.x

Carey, M. P., Sanderson, B. L., Barnas, K. A., & Olden, J. D. (2012). Native invaders: Challenges for science, management, policy, and society. *Frontiers in Ecology and the Environment*, *10*(7), 373–381. https://doi.org/10.1890/110060

Carrete, M., Lambertucci, S. A., Speziale, K., Ceballos, O., Travaini, A., Delibes, M., … Donázar, J. A. (2010). Winners and losers in human-made habitats: Interspecific competition outcomes in two neotropical vultures. *Animal Conservation*, *13*(4), 390–398. https://doi.org/10.1111/j.1469-1795.2010.00352.x

Costanza, K. K. L., Whitney, T. D., McIntire, C. D., Livingston, W. H., & Gandhi, K. J. K. (2018). A synthesis of emerging health issues of eastern white pine (*Pinus strobus*) in eastern North America. *Forest Ecology and Management*, *423*, 3–17. https://doi.org/10.1016/j.foreco.2018.02.049

Crespi, E. J., Rissler, L. J., & Browne, R. A. (2003). Testing Pleistocene refugia theory: Phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. *Molecular Ecology*, *12*(4), 969–984. https://doi.org/10.1046/j.1365-294X.2003.01797.x

Davis, M. B. (1983). Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden*, *70*(3), 550–563. https://doi.org/10.2307/2992086

Dodds, K. J., Aoki, C. F., Arango-Velez, A., Cancelliere, J., D'Amato, A. W., DiGirolomo, M. F., & Rabaglia, R. J. (2018). Expansion of southern pine beetle into northeastern forests: Management and impact of a primary bark beetle in a new region. *Journal of Forestry*, *116*(2), 178–191. https://doi.org/10.1093/jofore/fvx009

Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587.

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (Verson 3.2). *Cladistics*, *5*, 164–166.

Fontanella, F. M., Feldman, C. R., Siddall, M. E., & Burbrink, F. T. (2008). Phylogeography of *Diadophis punctatus*: Extensive lineage diversity and repeated patterns of historical demography in a trans-continental snake. *Molecular Phylogenetics and Evolution*, *46*(3), 1049–1070. https://doi.org/10.1016/j.ympev.2007.10.017

Funk, A. (1963). Studies in the genus *Caliciopsis*. *Canadian Journal of Botany*, *41*, 503–543. https://doi.org/10.1139/b63-044

Gandhi, K. J. K., & Herms, D. A. (2010). Direct and indirect effects of alien insect herbivores on ecological processes and interactions in forests of eastern North America. *Biological Invasions*, *12*(2), 389–405. https://doi.org/10.1007/s10530-009-9627-9

Garrick, R. C., Newton, K. E., & Worthington, R. J. (2018). Cryptic diversity in the southern Appalachian Mountains: Genetic data reveal that the red centipede, *Scolopocryptops sexspinosus*, is a species complex. *Journal of Insect Conservation*, *22*(5–6), 799–805. https://doi.org/10.1007/s10841-018-0107-3

Goudet, J. (2005). HIERFSTAT: A package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8278.2004.00828.x

Hanks, L. M., & Denno, R. F. (1998). Dispersal and adaptive deme formation in sedentary coccoid insects. In S. Y. Strauss (Eds.),*Genetic structure and local adaptation in natural insect populations* (pp. 239–262). Boston, MA: Springer.

Hapeman, P., Latch, E. K., Rhodes, O. E., Swanson, B., & Kilpatrick, C. W. (2017). Genetic population structure of fishers (*Pekania pennanti*) in the Great Lakes region: Remnants and reintroductions. *Canadian Journal of Zoology*, *95*(11), 869–876. https://doi.org/10.1139/cjz-2016-0325

Hassan, A., & Ricciardi, A. (2014). Are non-native species more likely to become pests? Influence of biogeographic origin on the impacts of freshwater organisms. *Frontiers in Ecology and the Environment*, *12*(4), 218–223. https://doi.org/10.1890/130188

Havill, N. P., Shiyake, S., Galloway, A. L., Foottit, R. G., Yu, G. Y., Paradis, A., … Caccone, A. (2016). Ancient and modern colonization of North America by hemlock woolly adelgid, *Adelges tsugae* (Hemiptera: Adelgidae), an invasive insect from East Asia. *Molecular Ecology*, *25*(9), 2065–2080. https://doi.org/10.1111/mec.13589

Hawes, T. C., Worland, M. R., Convey, P., & Bale, J. S. (2007). Aerial dispersal of springtails on the Antarctic Peninsula: Implications for local distribution and demography. *Antarctic Science*, *19*(1), 3–10. https://doi.org/10.1017/s0954102007000028

Hedin, M., & McCormack, M. (2017). Biogeographical evidence for common vicariance and rare dispersal in a southern Appalachian harvestman (Sabaconidae, *Sabacon cavicolens*). *Journal of Biogeography*, *44*(7), 1665–1678. https://doi.org/10.1111/jbi.12973

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1–2), 87–112. https://doi.org/10.1006/bijl.1999.0332

Houston, D. R. (1994). Major new tree disease epidemics: Beech bark disease. *Annual Review of Phytopathology*, *32*, 75–87. https://doi.org/10.1146/annurev.py.32.090194.000451

Jactel, H., Menassieu, P., Vetillard, F., Gaulier, A., Samalens, J. C., & Brockerhoff, E. G. (2006). Tree species diversity reduces the invasibility of maritime pine stands by the bast scale, *Matsucoccus feytaudi*

(Homoptera: Margarodidae). *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, *36*(2), 314–323. https://doi.org/10.1139/x05-251

Keane, R. M., & Crawley, M. J. (2002). Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution*, *17*(4), 164–170. https://doi.org/10.1016/s0169-5347(02)02499-0

Kerdelhúe, C., Boivin, T., & Burban, C. (2014). Contrasted invasion processes imprint the genetic structure of an invasive scale insect across southern Europe. *Heredity*, *113*(5), 390–400. https://doi.org/10.1038/hdy.2014.39

Kuntner, M., & Agnarsson, I. (2011). Phylogeography of a successful aerial disperser: The golden orb spider *Nephila* on Indian Ocean islands. *BMC Evolutionary Biology*, *11*, 119. https://doi.org/10.1186/1471-2148-11-119

Lefort, M. C., Boyer, S., De Romans, S., Glare, T., Armstrong, K., & Worner, S. (2014). Invasion success of a scarab beetle within its native range: Host range expansion versus host-shift. *Peerj*, *2*,e262. https://doi.org/10.7717/peerj.262

Mantel, N. (1967). Detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

McClure, M. S. (1976). Colonization and establishment of red pine scale, *Matsucoccus resinosae* (Homoptera: Margarodidae) in a Connecticut plantation. *Environmental Entomology*, *5*(5), 943–947.

McClure, M. S. (1977). Population-dynamics of red pine scale, *Matsucoccus resinosae* (Homoptera: Margarodidae): Influence of resinosis. *Environmental Entomology*, *6*(6), 789–795. https://doi.org/10.1093/ee/6.6.789

McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the USA*, *104*(50), 19885–19890. https://doi.org/10.1073/pnas.0706568104

Mech, A. M., Asaro, C., Cram, M. M., Coyle, D. R., Gullan, P. J., Cook, L. G., & Gandhi, K. J. K. (2013). *Matsucoccus macrocicatrices* (Hemiptera: Matsucoccidae): First report, distribution, and association with symptomatic eastern white pine in the southeastern United States. *Journal of Economic Entomology*, *106*(6), 2391–2398. https://doi.org/10.1603/ec13251

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Mendel, Z. (1998). Biogeography of *Matsucoccus josephi* (Homoptera: Matsucoccidae) as related to host resistance in *Pinus brutia* and *Pinus halepensis*. *Canadian Journal of Forest Research – Revue Canadienne De Recherche Forestiere*, *28*(3), 323–330. https://doi.org/10.1139/cjfr-28-3-323

Michigan Department of Natural Resources (2015). 2015 forest health highlights. Retrieved from https://fhm.fs.fed.us/fhh/fhh_15/MI_FHH_2015.pdf

Nackley, L. L., West, A. G., Skowno, A. L., & Bond, W. J. (2017). The nebulous ecology of native invasions. *Trends in Ecology & Evolution*, *32*(11), 814–824. https://doi.org/10.1016/j.tree.2017.08.003

Nadeau, S., Godbout, J., Lamothe, M., Gros-Louis, M.-C., Isabel, N., & Ritland, K. (2015). Contrasting patterns of genetic diversity across the ranges of *Pinus monticola* and *P. strobus*: A comparison between eastern and western North American postglacial colonization histories. *American Journal of Botany*, *102*(8), 1342–1355. https://doi.org/10.3732/ajb.1500160

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*(3), 583–590.

Nei, M., Maruyama, T., & Chakraborty, R. (1975). Bottleneck effect and genetic variability in populations. *Evolution*, *29*(1), 1–10. https://doi.org/10.2307/2407137

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H. (2018). vegan: community ecology package (R package). Retrieved from https://CRAN.R-project.org/package=vegan

Paolucci, E. M., MacIsaac, H. J., & Ricciardi, A. (2013). Origin matters: Alien consumers inflict greater damage on prey populations than do native consumers. *Diversity and Distributions*, *19*(8), 988–995. https://doi.org/10.1111/ddi.12073

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, *28*(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, *90*(4), 502–503. https://doi.org/10.1093/jhered/90.4.502

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Puechmaille, S. J. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. https://doi.org/10.1111/1755-0998.12512.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Raichle, B. W., & Carson, W. R. (2009). Wind resource assessment of the Southern Appalachian Ridges in the Southeastern United States. *Renewable & Sustainable Energy Reviews*, *13*(5), 1104–1110. https://doi.org/10.1016/j.rser.2007.12.005

Ray, W. W. (1936). Pathogenicity and cultural experiments with *Caliciopsis pinea*. *Mycologia*, *28*(3), 201–208. https://doi.org/10.2307/3754267

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*(3), 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Richards, W. R. (1960). A new species of *Matsucoccus* Cockerell (Homoptera: Coccoidea). *Canadian Entomology*, *92*(3), 179–181.

Richardson, D. M., & Ricciardi, A. (2013). Misleading criticisms of invasion science: A field guide. *Diversity and Distributions*, *19*(12), 1461–1467. https://doi.org/10.1111/ddi.12150

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., … Weller, S. G. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics*, *32*, 305–332. https://doi.org/10.1146/annurev.ecolsys.32.081501.114037

Salo, P., Korpimaki, E., Banks, P. B., Nordstrom, M., & Dickman, C. R. (2007). Alien predators are more dangerous than native predators to prey populations. *Proceedings of the Royal Society B-Biological Sciences*, *274*(1615), 1237–1243. https://doi.org/10.1098/rspb.2006.0444.

Schulz, A. N., Mech, A. M., Cram, M. M., Asaro, C., Coyle, D. R., Lucardi, R. D., … Gandhi, K. J. K. (2018). Association of *Caliciopsis pinea* Peck and *Matsucoccus macrocicatrices* Richards with eastern white pine (*Pinus strobus* L.) seedling dieback. *Forest Ecology and Management*, *423*, 70–83. https://doi.org/10.1016/j.foreco.2018.03.013

Simberloff, D., Souza, L., Nunez, M. A., Barrios-Garcia, M. N., & Bunn, W. (2012). The natives are restless, but not often and mostly when disturbed. *Ecology*, *93*(3), 598–607. https://doi.org/10.1890/11-1232.1

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457–462.

Stephens, G. R., & Aylor, D. E. (1978). Aerial dispersal of red pine scale, *Matsucoccus resinosae* (Homoptera: Margarodidae). *Environmental Entomology*, *7*(4), 556–563. https://doi.org/10.1093/ee/7.4.556

Tong, X., Wang, R., & Chen, X.-Y. (2018). Expansion or Invasion? A response to Nackley et al. *Trends in Ecology & Evolution*, *33*(4), 234–235. https://doi.org/10.1016/j.tree.2018.02.002

Unruh, T. R., & Luck, R. F. (1987). Deme formation in scale insects: A test with the pinyon needle scale and a review of other evidence. *Ecological Entomology*, *12*(4), 439–449. https://doi.org/10.1111/j.1365-2311.1987.tb01025.x

U.S. Environmental Protection Agency (2013). Level III ecoregions of the continental United States: Corvallis, Oregon, U.S. EPA – National Health and Environmental Effects Research Laboratory, map scale 1:7,500,000. Retrieved from https://www.epa.gov/eco-research/level-iii-and-iv-ecore gions-continental-united-states

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(13), 1–21. https://doi.org/10.18637/jss.v076.i13

Watson, W. Y., Underwood, G. R., & Reid, J. (1960). Notes on *Matsucoccus macrocicatrices* Richards (Homoptera: Margarodidae) and its association with *Septobasidium pinicola* Snell in eastern Canada. *Canadian Entomologist*, *92*(9), 662–667.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wendel, G. W., & Smith, H. C. (1990). Pinus strobus L. eastern white pine. In Burns, R. M., & Honkala, B. H. (Eds.), *Silvics of North America* (pp. 476–488).Washington DC: USDA Forest Service.

Whiteman, C. D., & Doran, J. C. (1993). The relationship between overlying synoptic-scale flows and winds within a valley. *Journal of Applied Meteorology*, *32*(11), 1669–1682. https://doi.org/10.1175/1520-0450(1993)032<1669:trboss>2.0.co;2

Whitney, T. D., Cram, M. M., Barnes, B. F., Yao, J. M., Lucardi, R. D., & Gandhi, K. J. K. (2018). Tree-level distribution of a novel insect-pathogen complex and its potential contribution to eastern white pine dieback. *Forest Ecology and Management*, *423*, 49–58. https://doi.org/10.1016/j.foreco.2018.02.002

Zemanova, M. A., Knop, E., & Heckel, G. (2016). Phylogeographic past and invasive presence of *Arion* pest slugs in Europe. *Molecular Ecology*, *25*(22), 5747–5764. https://doi.org/10.1111/mec.13860

Zhang, B., Edwards, O. R., Kang, L., & Fuller, S. J. (2012). Russian wheat aphids (*Diuraphis noxia*) in China: Native range expansion or recent introduction? *Molecular Ecology*, *21*(9), 2130–2144. https://doi.org/10.1111/j.1365-294X.2012.05517.x

## BIOSKETCH

**Thomas D. Whitney** is interested in forest entomology and forest health. He applies various interdisciplinary methodologies in his research, including landscape and population genetics, to understand the ecology and evolution of forest pest species.

**Kamal J. K. Gandhi** is a forest entomologist and conducts research on all aspects of community ecology and population dynamics of forest insects. She especially studies disturbance and chemical ecology, and she works on a variety of insect taxa with a focus on terrestrial ecosystems.

**Rima D. Lucardi** is a plant ecologist interested in biodiversity conservation through invasive species prevention and management. Her research seeks to conserve rare species while mitigating threats from the establishment and spread of invasives. She utilizes landscape and population biology with genetic approaches in her research program.

Author contributions: T.D.W., R.D.L. and K.J.K.G. designed the study, T.D.W. and R.D.L. contributed to sampling, T.D.W. performed the research and analyses, T.D.W., R.D.L. and K.J.K.G. provided useful comments, T.D.W. wrote the paper.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Whitney TD, Gandhi KJK, Lucardi RD. Native or non-native? Historical biogeography of an emergent forest pest, *Matsucoccus macrocicatrices*. *J Biogeogr.* 2019;00:1–15. https://doi.org/10.1111/jbi.13702