Weak spatial-genetic structure in a native invasive, the southern pine beetle (*Dendroctonus frontalis*), across the eastern United States

Ryan C. Garrick[1*], Ísis C. Arantes[1], Megan B. Stubbs[1], Nathan P. Havill[2*]

[1] Department of Biology, University of Mississippi, Oxford, MS 38677, USA
[2] USDA Forest Service, Northern Research Station, 51 Mill Pond Road, Hamden, CT 06514, USA

Corresponding Authors:
Ryan C. Garrick[*]
Department of Biology, University of Mississippi, Oxford, MS 38677, USA
Email address: rgarrick@olemiss.edu
Nathan P. Havill[*]
USDA Forest Service, Northern Research Station, Hamden, CT 06514, USA
Email address: nathan.p.havill@usda.gov

Abstract

The southern pine beetle, *Dendroctonus frontalis*, is a native pest of pine trees that has recently expanded its range into the northeastern United States. Understanding its colonization, dispersal, and connectivity will be critical for mitigating negative economic and ecological impacts in the newly invaded areas. Characterization of spatial-genetic structure can contribute to this; however, previous studies have reached different conclusions about regional population genetic structure, with one study reporting a weak east-west pattern, and the most recent reporting an absence of structure. Here we systematically assessed several explanations for the absence of spatial-genetic structure. To do this, we developed nine new microsatellite markers and combined them with an existing 24-locus data matrix for the same individuals. We then reanalyzed this full dataset alongside datasets in which certain loci were omitted with the goal of creating more favorable signal to noise ratios. We also partitioned the data based on the sex of *D. frontalis* individuals, and then employed a broad suite of genotypic clustering and isolation-by-distance (IBD) analyses. We found that neither inadequate information content in the molecular marker set, nor unfavorable signal-to-noise ratio, nor insensitivity of the analytical approaches could explain the absence of structure. Regardless of dataset composition, there was little

37 evidence for clusters (i.e., distinct geo-genetic groups) or clines (i.e., gradients of increasing
38 allele frequency differences over larger geographic distances), with one exception: significant
39 IBD was repeatedly detected using an individual-based measure of relatedness whenever datasets
40 included males (but not for female-only datasets). This is strongly indicative of broad-scale
41 female-biased dispersal, which has not previously been reported for *D. frontalis*, in part owing to
42 logistical limitations of direct approaches (e.g., capture-mark-recapture). Weak spatial-genetic
43 structure suggests long-distance connectivity and that gene flow is high, but additional research
44 is needed to understand range expansion and outbreak dynamics in this species using alternate
45 approaches.

47 ## Introduction

48

49   The southern pine beetle, *Dendroctonus frontalis* (Zimmermann 1868), is an
50 economically important pest of pine trees across eastern North America. Outbreaks have been
51 particularly devastating to hard pines (Gernandt et al. 2005), and multimillion-dollar losses in
52 timber and pulpwood have been recorded (Price et al. 1997; Pye et al. 2011). The species'
53 impacts are now moving north, with pitch pine ecosystems in northeastern North America newly
54 at risk (Lesk et al. 2017; Dodds et al. 2018). Given the destruction that can be caused by *D.*
55 *frontalis*, and the potential for this species to continue extending its range northward owing to a
56 warming climate, a deeper understanding of the population biology of this beetle, including
57 development and application of molecular tools that can provide insights into dispersal,
58 connectivity, and local mating dynamics, is of immediate importance.

59

60   Owing to the pest status of *Dendroctonus frontalis*, its natural history and population
61 ecology are relatively well understood. All life stages are either fully or partly dependent on the
62 inner bark or phloem of the host tree (Hain et al. 2011). Initially, adult pioneer female beetles are
63 attracted to odors from stressed host trees (e.g., owing to crowding, drought, disease, lightning
64 strikes, or other storm damage), and upon arrival by flight, they initiate an attack and release
65 aggregation pheromones that attract males (Sullivan 2011). Monogamous reproduction typically
66 occurs in nuptial chambers within a newly attacked tree, although some females arrive already
67 gravid. Up to 30 eggs are deposited in serpentine galleries that are excavated by the female
68 parent, who also deposits mycangial fungal spores in the galleries, leading to fungal growth that
69 provides nutritional sustenance to the early stage larval offspring. After developing to the adult
70 stage, offspring emerge from the outer bark of the tree, and then disperse to nearby hosts, which
71 they attack (Hain et al. 2011). The combination of tree odors and aggregation pheromones can
72 stimulate a mass attack that results in "infestation spots" that enlarge within a forest (Sullivan
73 2011). Such large-scale irruptions cause widespread pine tree morality among healthy trees.
74 These epidemic outbreak years have occurred at irregular intervals ranging from every 5–10
75 years from the early 1970s to 2000s in the species' native range in the south, where as many as
76 nine overlapping generations per year can be completed. Historically, climatic conditions in the

77   northern regions prevent more than just a few generations per year, with the species' latitudinal
78   range limit thought to be constrained by winter lows, where temperatures of -16 °C cause almost
79   100% mortality (Hain et al. 2011).
80

81   There are several reasons to expect spatial-genetic structure across the range of *D.*
82   *frontalis* in the eastern United States. For example, its distribution spans the southern
83   Appalachian Mountains—a well-known biogeographic and genetic barrier for diverse species
84   (Soltis et al. 2006), including arthropods (Garrick 2011). Also, given that *D. frontalis* is native to
85   the southern United States, with a broad geographic range that extends south to Honduras and
86   west to Arizona, there should have been ample time for such structure to evolve. Capture-mark-
87   recapture and flight mill studies have estimated mean dispersal distances of 0.5 km to 3.4 km
88   (Turchin & Thoeny 1993; Kinn et al. 1994). While this is considered quite a long distance
89   because of the species' small size ($\leq$ 3 mm), given its expansion into the northeastern United
90   States over the past decade or more via source-sink stepping-stone dynamics, genetic isolation-
91   by-distance (IBD) might be detectable. Indeed, earlier work by Schrey et al. (2011) identified a
92   weak east-west division roughly coinciding with the southern Appalachian Mountains, as well as
93   IBD across the eastern United States, using eight microsatellite loci.
94

95   Somewhat surprisingly, a recent investigation of spatial-genetic structure in *D. frontalis*
96   by Havill et al. (2019) found broad-scale differentiation between eastern and western North
97   America but did not detect distinct geo-genetic groups nor IBD within eastern North America.
98   The discrepancy with earlier findings is not clearly attributable to differences in the underlying
99   genetic marker set, or the extent of geographic sampling. Havill et al. (2019) used 24
100   microsatellite loci that included the eight loci from Schrey et al. (2011), and the maximum
101   pairwise distances among collection localities were very similar (1658 km vs. 1513 km,
102   respectively). Also, in both studies, per-site sample sizes were quite large (typically $\geq$ 25
103   individuals). Thus, it would seem that biologically meaningful genetic distance estimates should
104   have been attainable. From an applied management perspective, it is important to reconcile these
105   contrasting inferences about spatial-genetic structure among *D. frontalis* in eastern North
106   America because they have different ramifications for impacts upon pine forest ecosystems in
107   newly colonized regions. For example, if distinct genetic clusters exist, multiple gene pools may
108   converge and recombine at the wave front, producing an invasive population with elevated
109   genetic variation and adaptive potential (e.g., Kolbe et al. 2004). Conversely, if the species is
110   essentially panmictic across eastern North America, then the risk of rapid adaptive evolution
111   (e.g., leading to larger population sizes and/or faster expansion speeds in novel environments;
112   Szücs et al. 2017) may be lowered, and mitigation measures previously developed for outbreaks
113   in the native range should be readily transferable to the northeastern United States. That said,
114   forecasting responses to novel climates and associated no-analog communities is challenging
115   (Williams & Jackson 2007), likely necessitating adaptive management.
116

117   There are several plausible explanations for an unexpected absence of clusters or clines.
118 Some can be broadly classified as technical issues, such as inadequate information content or
119 unfavorable signal-to-noise ratio in the molecular marker set, and/or insensitivity of the chosen
120 analytical approaches or violation of key assumptions. Alternatively, species-specific biological
121 phenomena may mask or overwrite signatures of spatial-genetic structure. For example, when
122 using biparentally inherited markers, a life history trait such as sex-biased dispersal could
123 prevent detection of philopatry that is exhibited by only half of the members of a population.
124 Likewise, a demographic event such as recent rapid range expansion could create non-
125 equilibrium conditions that re-partition genetic variation (perhaps ephemerally) to be consistent
126 with broad-scale panmixia (i.e., similar to impacts of postglacial expansion into newly available
127 habitats; Hewitt 1996, 2004, but see Excoffier & Ray 2008).
128
129   While the explanations for a lack of detectable spatial-genetic structure given above are
130 not mutually exclusive (e.g., technical limitations and biological influences may act
131 concurrently), and acknowledging the potential for genuine panmixia (i.e., a null hypothesis that
132 may be true), the goal here was to systematically explore evidence for each of them. This was
133 approached by augmenting, sub-setting, and reanalyzing Havill et al.'s (2019) dataset.
134 Specifically, we enhanced overall information content by developing and screening additional
135 microsatellite loci, and elevated signal over noise either by identifying and omitting loci most
136 likely to be compromised by null alleles and homoplasy, or following Russello et al. (2012), by
137 retaining only those loci that showed the greatest magnitude of differentiation among sampling
138 sites. For each of these, we also created male-only and female-only partitions of the data. To
139 circumvent potential analytical inefficiencies and/or restrictive assumptions, we used a suite of
140 methods/metrics that each leverage different types of signal in the data, thereby reducing the
141 overall risk of false negatives.
142
143   Our exploration of the extent to which inferences about the type, and strength, of spatial-
144 genetic structure may be impacted by technical and/or biological factors improves our
145 understanding of the biology of a destructive native invasive species. This work also contributes
146 new molecular markers that can be used to investigate other aspects of the biology of *D.*
147 *frontalis*, such as questions about parentage and relatedness, family group structure and kin
148 clustering, as well as inbreeding dynamics.
149

Materials & Methods

151
152 General approaches to assess lack of structure
153
154 *1. Inadequate information content.* The power of microsatellites to detect spatial-genetic
155 structure increases as independent loci are added (Allendorf et al. 2013), but in some cases, at
156 least 30 loci may be needed to detect weak structure (e.g., Duchesne & Turgeon 2012). To
157 address this, nine new microsatellite markers were developed here, and used to genotype the

158   same individuals included in Havill et al.'s (2019) 24-locus data matrix. Briefly, those samples
159   were collected with permission from the U.S. Department of Agriculture Forest Service by
160   numerous colleagues (see field permit statement) between 2013 and 2017, using Lindgren funnel
161   traps baited with frontalin and alpha-pinene. Specimens were identified as *D. fontalis* following
162   Armendáriz-Toledano and Zuñiga's (2017) key, and individual sex was determined based on the
163   presence (female) or absence (male) of a mycangium on the pronotum. The new and old datasets
164   were combined, creating what we refer to as the "augmented dataset."

166   *2. Unfavorable signal-to-noise ratio*. The utility of microsatellite data can be compromised by
167   technical artefacts, and/or affected (either negatively of positively) by inherent features of the
168   loci themselves. To *reduce overall noise*, we omitted loci most likely to suffer from a higher
169   frequency of non-amplifying null alleles and/or allele size homoplasy based on a proxy for
170   potential mutation rate. Herein, this is the "low noise dataset." As an alternative strategy, we
171   attempted to *enhance overall signal* by retaining only those loci that maximized genetic
172   differentiation, as measured by F_{ST} (Weir & Cockerham 1984) among sampling sites. The
173   thresholds for inclusion in this "high signal dataset" was chosen based on observed natural
174   breaks in the distribution of F_{ST} values.

176   *3. Sex-biased dispersal*. Different degrees of site fidelity between males and females can affect
177   detection of population differentiation when using biparentally inherited microsatellite markers.
178   Impacts are most pronounced when sex-specific dispersal rates and distances are highly
179   asymmetric, and when sampling includes many individuals that have dispersed but not yet
180   reproduced (Prugnolle & de Meeûs 2002). To evaluate this, we analyzed male-only ($n = 185$
181   individuals from 9 sites) and female-only ($n = 70$, 8 sites) partitions, in addition to combined
182   datasets ($n = 255$, 9 sites; Figure 1).

184   *4. Insensitivity of analysis, or violation of assumptions*. Genotypic clustering methods differ in
185   their incorporation of geo-spatial information, and when differentiation is weak, this impacts
186   accuracy (Chen et al. 2007). Also, sensitivity of IBD analyses can depend on the chosen genetic
187   distance metric (Séré et al. 2017). To at least partly account for this, each of the above datasets
188   were analyzed using several different clustering and IBD approaches that represent different
189   strengths and limitations of a broader suite of existing methods and metrics. That said, we
190   recognize that clusters vs. clines are not mutually exclusive (Rosenberg et al. 2005), and so we
191   interpret these outcomes with caution, and in the context of our sampling limitations.

193   Development and validation of new microsatellite loci

195   In the present study, we extended upon the assessment of a suite of microsatellite loci
196   identified by Havill et al. (2019). Specifically, nine additional loci were selected from among
197   those identified using the software QDD (Meglecz et al. 2014) following low coverage whole
198   genome sequencing of one male *D. frontalis* from Homochitto National Forest, Mississippi

199   (NCBI BioProject Accession Number PRJNA493650). To assess the suitability of these new
200   loci, multiplex polymerase chain reaction (PCR) amplifications were performed for sets of three
201   loci, each with a different 5' fluorescent label on the forward primer and mostly non-overlapping
202   allele size ranges (Table 1). Reactions were conducted in 15 µL volumes that contained 7.5 µL
203   Type-it® Microsatellite PCR Kit master mix (Qiagen, Valencia CA), 1.5 µL dH$_2$O, 0.75 µL of
204   each primer (10 mM), and 1.5 µL genomic DNA. The following "touchdown" thermal cycling
205   conditions were used: 95ºC for 2 min (1 cycle), 95ºC for 30 sec, 61ºC minus 2ºC per successive
206   cycle for 30 sec, 72ºC for 45 sec (5 cycles), 95ºC for 30 sec, 51ºC for 30 sec, 72ºC for 45 sec (30
207   cycles), and 60ºC for 30 min (1 cycle).

208

209   Amplified fragments were run on an ABI 3730 sequencer (Life Technologies) with a Liz
210   500 ladder (Gel Company, San Francisco CA) at Yale University's DNA Analysis Facility on
211   Science Hill. Genotype scoring was performed using GENEIOUS v10.0.5 (Kearse et al. 2012).
212   Following Havill et al. (2019), we assessed Mendelian inheritance patterns for each of the new
213   loci in two local populations (i.e., 27 individuals from Sicily Island, Louisiana, and 28
214   individuals from Tombigbee, Mississippi). For each locus and population, evidence for
215   departures from Hardy-Weinberg Equilibrium (HWE) were examined via exact tests (Guo &
216   Thompson 1992), with significance assessed using 10,000 Markov Chain permutations.
217   Observed (H$_O$) and expected (H$_E$) heterozygosity were also calculated. Independent segregation
218   of alleles among loci was assessed using all *D. frontalis* samples (255 individuals) by testing for
219   departures from Linkage Equilibrium (LE) for all possible pairs of 33 loci (i.e., nine reported
220   here, plus 24 from Havill et al. 2019), again using 10,000 Markov Chain permutations. Given
221   that nine different sampling sites were included, Fisher's combined probability test was used to
222   calculate *P*-values across populations for each locus pair. These analyses were performed in
223   GENEPOP v4.5.1 (Rousset 2008), with sequential Bonferroni correction (Holm 1979) of *P*-
224   values associated with LE tests.

225

226   Identification and omission of loci that may elevate stochastic noise

227

228   MICRO-CHECKER v2.2.3 (van Oosterhout et al., 2004) was used to determine whether
229   there was evidence for null alleles, and if so, to estimate their frequencies. For each locus in each
230   local population, a null allele was inferred to exist if there was a significant excess of
231   homozygotes that were evenly distributed across all homozygote classes. This was determined
232   by Fisher's combined probability test, with *P*-values determined using 1,000 permutations.
233   Frequency of the null allele (*r*) was then was estimated using the method of Chakraborty et al.
234   (1992), which assumes that missing data are not attributable to homozygous nulls, and that
235   populations are panmictic. To identify inbred populations for removal, we relied on the tendency
236   for inbreeding to generate homozygote excess at many loci (cf. locus-specific impacts of null
237   alleles). Here we defined outbred populations as those in which fewer than seven loci (i.e., <
238   20%) showed significantly (*P* < 0.05) positive F_{IS} values. For each locus and population, the null

239 hypothesis of panmixia was compared to an alternative hypothesis of heterozygote deficiency
240 using a U-test (Rousset & Raymond 1995), performed in GENEPOP with 10,000 Markov Chain
241 permutations. After omitting outbred populations, we rank-ordered the loci based on the level of
242 noise contributed by putative null alleles. To ensure that both prevalence and severity of the null
243 allele at a locus were considered, the r value from each population with a significant excess of
244 homozygotes was summed ($r_{cumulative}$).
245
246   The number of contiguous repeat units may be a useful proxy for microsatellite mutation
247 rate, given that DNA structure affects the opportunity for slipped-strand mispairing (Bhargava &
248 Fuentes 2010, and references therein). Here we used the reported positive association between
249 repeat number and mutation rate to identify which loci might be predisposed to having alleles
250 identical in state but not by descent. Briefly, based on one sequence per locus generated during
251 marker development (Schrey et al. 2007; Havill et al. 2019; this study), we first estimated the
252 PCR amplicon length for each locus given the locations of forward and reverse primer annealing
253 sites. Next, the observed number of contiguous repeats was scaled to reflect what should be
254 present within the median-sized allele (i.e., a reference allele identified from the eastern United
255 States gene pool as whole). Rank-ordering of loci based on the potential level of stochastic noise
256 contributed by homoplasy followed the inferred number of repeats of each reference allele.
257
258   The omission of loci that may elevate overall noise was based on joint consideration of
259 null alleles and homoplasy. Along each of these two axes, we iteratively omitted the lowest
260 ranked locus one at a time, until a two-sample t-test (one-tailed) determined that a significant (P
261 < 0.05) reduction in the mean value of the original metric used for rank-ordering (compared to
262 that of the 33-locus dataset) had been achieved. To assess whether the low noise dataset was
263 comprised of a subset of loci with inadvertently low information content, we calculated Paetkau
264 et al.'s (1995) Probability of Identity (PI; the chance that two individuals randomly drawn from a
265 panmictic population have the same genotype) using GENALEX v6.5 (Peakall & Smouse 2006,
266 2012). We then compared mean PI between the reduced vs. full dataset, using a t-test as above.
267
268 Identification and retention of loci that may enhance overall signal
269
270   For species in which population divergence is shallow, "outlier" microsatellite loci (e.g.,
271 those displaying unusually strong differentiation, potentially due to effects of divergent selection
272 on linked genomic regions) may be particularly useful for identification of population structure
273 (Russello et al. 2012). This is because recent divergence (e.g., over post-glacial timescales or
274 shorter) may be slow to register in most loci, particularly when it is coupled with large effective
275 populations sizes. To explore the potential benefit of focusing on loci with the strongest
276 signatures of spatial-genetic structure in *D. frontalis*, we calculated global F_{ST} (Weir &
277 Cockerham 1984) using the augmented dataset, for each of the 33 loci separately, and then
278 plotted the distribution of these values in rank order. Natural breaks were identified qualitatively,

279   based on the magnitude of change (i.e., $\Delta F_{ST} = F_{ST\ locus\ n} - F_{ST\ locus\ n+1}$, where n is the rank order
280   position). These breaks defined the thresholds for retention vs. exclusion of loci from the high
281   signal dataset.

282

283   Genotypic clustering analyses

284

285   To infer the number of natural genetic groups (K) and their members, we analyzed
286   microsatellite datasets using three approaches: STRUCTURE v2.3.4 (Pritchard et al. 2000),
287   BAPS v6.0 (Corander et al. 2003), and Discriminant Analysis of Principal Components (DAPC;
288   Jombart et al. 2010). The first two are Bayesian approaches and assume HWE and LE within
289   "true" clusters. However, STRUCTURE is strictly individual-based and does not make use of
290   geo-referenced samples (although priors on population of origin can be applied), but it has been
291   shown to tolerate clinal variation relatively well (Chen et al. 2007). BAPS can be run using either
292   individuals or groups of individuals as the basic units of analysis, and it explicitly considers
293   spatial coordinates, which can help when differentiation is weak. DAPC can also be used to
294   identify natural groups of individuals *a posteriori*, but this method does not attribute the
295   underlying cause of such groups to population genetic processes, nor does it incorporate spatial
296   information. In all clustering analyses, we assessed K values from one up to and including the
297   total number of sampling sites (i.e., maximum $K = 8$ or 9, depending on the dataset under
298   consideration).

299

300   *STRUCTURE*. All runs employed the correlated allele frequency and admixture ancestry models.
301   Estimated -ln likelihood scores were obtained for each value of K, with 20 replicates each. A
302   burn-in of 50,000 Markov chain Monte Carlo (MCMC) generations and run length of 200,000
303   MCMC generations were used, with other parameters set as default. The best-fit value of K was
304   chosen using Evanno et al.'s (2005) ΔK method implemented in STRUCTURE HARVESTER
305   v0.6.94 (Earl & vonHoldt 2012). However, the ΔK method is unable to assess support for $K=1$,
306   and it may be subject to bias with respect to identifying only the top (e.g., $K=2$) level of
307   hierarchical structure (see Janes et al. 2017). Accordingly, we also used a visual assessment of
308   Ln Pr(X|K) plots, from which the smallest K that captured the major structure in the data was
309   accepted, following Pritchard and Wen (2003).

310

311   *BAPS*. To enable spatial clustering of individuals, we created non-redundant GPS coordinates via
312   the addition of fine-scale jitter (i.e., a random number at the 4th and 5th decimal places, equivalent
313   to ≤ 11.1 m at the equator). For spatial clustering of groups of individuals, each of the 8 or 9
314   sampling locations (depending on the dataset) defined an *a priori* group, and the original spatial
315   coordinates were used. Accordingly, the complexity of Voronoi tessellations differed
316   considerably between the two approaches. Spatial data were analyzed with associated
317   microsatellite genotype data, with 10 replicates per run. The best-fit K was identified via log
318   marginal likelihood scores (Cheng et al. 2013).

319

320   *DAPC*. Individual-based multilocus microsatellite genotypes were first transformed via principal
321   component (PC) analysis to identify uncorrelated variables to be used for subsequent
322   discriminant analysis. The best-fit number of clusters was determined via *k*-means clustering,
323   with the lowest Bayesian Information Criterion (BIC) score used to choose *K*. Following Miller
324   et al. (2020), an initial preliminary DAPC was conducted to determine the optimal number of
325   PCs to retain, followed by a second final DAPC using the chosen number of PCs. In the initial
326   assessment, the optimal number of PCs was determined using the cross-validation procedure,
327   considering 20 to 300 PCs (in increments of 20) with 30 replicates at each level of PC retention,
328   and 90% of the data comprising the training set. Pre-defined groups based on sampling location
329   information were used only for the purpose of quantifying the number of *de novo* clusters that
330   were inferred to be present at a given sampling location. DAPC was implemented using
331   ADEGENET v2.1.2 (Jombart 2008) in R v3.6.1 (R Core Team 2020).

332

333   Isolation-by-distance analyses

334

335   To assess evidence for subtle gradients in genetic variation across the landscape, we
336   conducted IBD analyses of our microsatellite datasets using either sampling sites or individuals
337   as the basic units. The rationale for these two approaches was that population allele frequencies
338   typically can change rather gradually over time and therefore provide insights into historical
339   processes, whereas diploid genotypes are much more labile owing to reshuffling of alleles every
340   generation in sexually reproducing species, and therefore have the potential to closely track
341   recent or on-going processes (Sunnucks 2000; Garrick et al. 2010; Epps & Keyghobadi 2015).

342

343   *Population-based metrics*. Three measures of genetic distance were used: Cavalli-Sforza and
344   Edwards' (1967) chord distance (D_c), which has desirable properties for detecting IBD in the
345   presence of null alleles (Séré et al. 2017); Nei's (1972) standard genetic distance (D_s), which
346   increases linearly with geographic distance if mutation follows an infinite allele model and
347   occurs at a constant rate; and Weir and Cockerham's (1984) estimate of F_{ST} [linearized as
348   $F_{ST}/(1- F_{ST})$ following Slatkin (1995), to account for the high mutation rate of microsatellite loci.

349

350   *Individual-based metrics*. Three metrics for genetic similarity were used: the kinship coefficient
351   (D_{kf}, Cavalli-Sforza & Bodmer 1971); proportion of shared alleles (D_{ps}, Bowcock et al. 1994), or
352   relatedness ($\hat{r}$, Lynch & Ritland 1999). Briefly, for a given locus, D_{kf} is the probability that a
353   randomly chosen allele from one individual is identical-by-descent to that of second individual
354   (summed over all loci and alleles). Conversely, when comparing two mutlilocus genotypes, D_{ps}
355   is calculated as number of shared alleles summed over loci / (2× number of loci compared).
356   Unlike the other two metrics, Lynch and Ritland's (1999) methods-of-moments estimator of
357   relatedness makes use of information on population allele frequencies (e.g. a shared rare allele is

358 considered more likely to be identical-by-descent than a shared common allele). Given that $\hat{r}$ is
359 asymmetrical, following Ritland (2000), we used the average of both directions.
360
361   MICROSATELLITE ANALYZER v4.05 (Dieringer & Schlötterer 2003) was used to
362 calculate D_c, D_{kf}, and D_{ps}, whereas GENALEX was used to calculate D_s, F_{ST} and $\hat{r}$, as well as
363 linear pairwise geographic distance matrices based on latitude and longitude coordinates, using a
364 modified Haversine Formula. For all datasets and distance metrics, the significance of IBD was
365 assessed via Mantel (1967) tests with 999 permutations.
366

Results

368
369 Development and validation of new microsatellite loci
370
371   For the nine new loci developed here, amplification success across all eastern United
372 States *D. frontalis* samples was reasonably high (mean of 4.2% missing data per locus). Three
373 loci showed significant departures from HWE in the direction of homozygote excess. However,
374 in all cases this was limited to just one of the two local collection sites (Table 1), suggesting that
375 there were no intrinsic issues with allelic inheritance (i.e., the new microsatellite loci appear to
376 be autosomal, diploid, and single copy). Levels of with-population polymorphism were modest
377 (mean of ~4.3 alleles per new locus for the two exemplar local populations). Across all eastern
378 United States *D. frontalis* samples, there was a mean of 8.1 alleles per new locus (Table 1).
379 Based on our assessment of LE for all 33 loci (528 locus pairs), 22 pairs were significant at the P
380 < 0.05-level, but none remained so after sequential Bonferroni correction to account for multiple
381 tests. Accordingly, there was no strong evidence for non-independence among loci in the
382 augmented dataset. Preliminary analyses using all loci indicated weak genetic structure among
383 sampling sites (F_{ST} = 0.009, P = 0.001).
384
385 Identification and omission of loci that may elevate overall noise
386
387 *Null alleles*. Two local populations had seven or more loci with significantly positive F_{IS} values:
388 Homochitto, MS, and Woolford, MD. These were considered inbred and therefore omitted from
389 consideration when assessing null alleles. MICRO-CHECKER identified 13 loci that may have a
390 null allele (Table 2). However, prevalence was low (i.e., restricted to a single population) for 10
391 of these. Of the remaining loci, two had null alleles in four populations, and one had nulls in two
392 populations. Across the 20 cases of putative null alleles (out of 33 loci × 7 retained populations =
393 231 potential cases), mean estimated r was 0.245 (range: 0.139–0.431). Rank-ordering followed
394 by iterative removal revealed that omission of the five worst loci was required to significantly
395 reduce the mean $r_{cumulative}$ value (t = 1.720, d.f. = 40, P = 0.047). The removal of these loci did
396 not cause a concomitant reduction in mean information content, as measured by PI (t = -0.435,
397 d.f. = 57, P = 0.332).

398

399   *Homoplasy*. For 31 loci, the available allele sequence trimmed at the 5' ends of both primers
400   yielded a predicted PCR product that was within the empirically determined size range for
401   eastern United States *D. frontalis*. For two loci with unexpectedly short or long amplicons, an
402   alternative priming site (i.e., a region of high sequence similarity located up or downstream of
403   the original target site) was readily identifiable (Table 3). Three loci constrained two distinct
404   microsatellite regions. These were either composed of different (i.e., compound) repeat motifs
405   located immediately adjacent to one other (locus Dfr-16; Schrey et al. 2007), or the same (i.e.,
406   impure) repeat motif interrupted by a 2-bp (locus Dfr-24; Schrey et al. 2007) or 3-bp (locus
407   SPB1534; this study) insertion. In all three cases, only the longest repeat region was used as a
408   proxy for potential for homoplasy. Following re-scaling (see Methods), across 33 loci the
409   average value of the median contiguous number of repeats was 7.6 (range: 4.945–14.901; Table
410   3). The five most concerning loci needed to be removed in order to significantly reduce the
411   number of repeat units ($t = 1.819$, d.f. = 50, $P = 0.037$), but their omission did not also reduce
412   mean PI ($t = -0.780$, d.f. = 58, $P = 0.219$).

413

414   Of the two sets of five omitted loci, one (Dfr-24) was common to both approaches used
415   to reduce noise. Accordingly, the largest subset of loci with reduced null allele and homoplasy
416   issues was made up of 24 loci. Reassessment of the impacts of our strategic removal of nine loci
417   to create the low noise dataset showed that the reduction in mean $r_{\text{cumulative}}$ remained close to
418   significant ($t = 1.520$, d.f. = 42, $P = 0.068$), and so did reduction in mean number of repeats ($t = $
419   1.499, d.f. = 53, $P = 0.070$). As before, there was no inadvertent change in mean PI ($t = -1.180$,
420   d.f. = 51, $P = 0.122$). Although our low noise dataset happened to contain the same number of
421   loci originally analyzed by Havill et al. (2019), their compositions differed by eight loci.

422

423   Identification and retention of loci that may enhance overall signal

424

425   Based on a visual examination of the ΔF_{ST} plot, several inflection points marking
426   pronounced decreases in F_{ST} among adjacent rank-ordered loci were apparent (Figure 2). Aside
427   from steep declines that would have set a threshold for retention of too few loci for meaningful
428   population genetic analyses (i.e., ≤ 5 loci), there were two other inflection points: one at the
429   transition from the 10th to 11th ranked locus, and another at the transition from 19th to 20th.
430   Accordingly, we created two alternative versions of the high sensitivity dataset. One contained
431   10 loci with the highest F_{ST} values, and the other was an expansion of this, where we retained the
432   19 highest ranked loci (Table 4). The 10-locus and 19-locus high signal datasets contained 3 and
433   6 loci, respectively, that were not shared with the 24-locus low noise dataset. Compared to Havill
434   et al.'s (2019) original dataset, the 10-locus and 19-locus high signal datasets contained 2 and 5
435   novel loci, respectively. Ultimately, all datasets analyzed in the present paper differed from each
436   other and from the previous study, thereby enabling meaningful comparisons.

437

438   Genotypic clustering analyses

439

440   *STRUCTURE*. Across all 12 datasets (i.e., ADS, LNDS, and HSDS with 10 or 19 loci; males and
441   females combined, or separated) there were five instances where the two alternative approaches
442   for identifying the best-fit value of K were in conflict. In these cases, Pritchard and Wen's (2003)
443   qualitative method inferred $K=1$ whereas Evanno et al.'s (2005) ΔK method inferred $K=2, 3,$ or
444   4, depending in the dataset (Table 5; but note that ΔK is not capable of assessing $K=1$). More
445   generally, irrespective of whether there was agreement or conflict between methods, wherever
446   $K>1$ was inferred, the optimal clustering solution almost always contained "ghost clusters"
447   (*sensu* Guillot et al. 2005; i.e., those for which no individuals were strongly assigned with
448   membership coefficient of $Q>0.50$). In some cases, all clusters were nonsensical. For instance,
449   based on ΔK, the best-fit value was $K=3$ for LNDS males plus females, and LNDS females-only,
450   yet not a single individual was strongly assigned; Table 5). Additionally, when $K>1$ there was
451   also an absence of geographic localization of clusters. Ultimately, as with Havill et al.'s (2019)
452   STRUCTURE analysis of a 24-locus *D. frontalis* dataset, our reanalysis using augmented and/or
453   sub-setted datasets did not provide any indications of discrete population structure.

454

455   *BAPS*. The optimal partition of geo-referenced molecular data was $K=1$ for all datasets, using
456   both individual- and population-based clustering approaches (Table 5). Given this outcome, we
457   performed post-hoc HWE exact tests in GENEPOP, treating all individuals as members of a
458   single population. Each of the 12 datasets showed highly significant departures from HWE (all
459   $P<0.0005$), indicating that eastern United States *D. frontalis* cannot simply be characterized as
460   panmictic.

461

462   *DAPC*. Across all datasets, there were only two instances where $K=1$ was inferred (Table 5; but
463   note that DAPC was not intended for this purpose; see Miller et al. 2020). However, wherever
464   $K\geq2$, there were strong indications that the groups were artifacts rather true reflections of
465   population structure. For example, there were no cases where a given local sampling site
466   contained members of only a single inferred cluster. Indeed, in seven of the datasets with best-fit
467   $K\geq2$, all sampling sites contained representatives of all inferred clusters (Table 5); such levels of
468   coexistence among putatively distinct gene pools has no plausible biological explanation.

469

470   Isolation-by-distance analyses

471

472   *Population-based metrics*. Regardless of which microsatellite dataset was analyzed, or which
473   genetic distance metric was used, no significant IBD was detected (all $P\geq0.130$; Table 6).

474

475   *Individual-based metrics*. All instances of significant IBD were limited to Lynch and Ritland's
476   (1999) relatedness (note that such relationships are negative). Furthermore, this significant IBD
477   was almost always associated with datasets that included males, with the one exception being the

478   female-only 19-locus high sensitivity dataset (Table 6). Despite relatively strong significance (all
479   $P=0.001$, except for the aforementioned female-only dataset, which had $P=0.031$; Table 6), the
480   nature of relationships between relatedness and geographic distance were consistently weak
481   [augmented dataset with males plus females: slope = -2×10^{-6}, correlation coefficient (r) = 0.036;
482   augmented dataset with males only: slope = -2×10^{-6}, r = 0.037; low noise dataset with males plus
483   females: slope = -2×10^{-6}, r = 0.020; low noise dataset with males only: slope = -2×10^{-6}, r =
484   0.030; and even weaker for the 19-locus high signal dataset with females only: slope = -3×10^{-6}, r
485   = 0.001]. Overall, these outcomes are consistent with female-biased dispersal; however, males do
486   not appear to be strongly philopatric.

487

488   **Discussion**

489

490   Spatially explicit population genetic analyses of microsatellite data from invasive insect
491   species can identify the geographic origin(s) and number of independent introductions into newly
492   invaded areas, and provide insights into the relative importance of natural vs. human-mediated
493   dispersal (e.g., fire ants, Ascunce et al. 2011; termites, Perdereau et al. 2013; hemlock woolly
494   adelgid, Havill et al. 2016). These inferences have implications for management, such as
495   enabling targeted control (e.g., of an invasive "bridgehead" population that is the primary source
496   of subsequent spread [Lombaert et al. 2010] and which may have evolved higher invasiveness
497   [Whitney & Gabler 2008]), or enhanced surveillance of "stowaways" where transportation
498   networks are involved in accelerating spread (e.g., via inadvertent movement of contaminated
499   live plants, wood packing materials, or firewood; Meurisse et al. 2019). Notably, these insights
500   often depend on the existence of discrete populations, yet a long-standing challenge in
501   population genetics is the "clusters vs. clines" problem. This refers to the tendency for clustering
502   methods to fail (or mislead) when structure manifests as continuous gradients of genetic
503   differentiation, and conversely, true signatures of IBD can be distorted if sampling traverses
504   abrupt genetic breaks or is discontinuous and uneven (e.g., Frantz et al. 2009; Bradburd et al.
505   2018). While clusters and clines are not mutually exclusive (see Rosenberg et al. 2005), an
506   understanding of the predominant form of spatial-genetic structure informs the choice of
507   appropriate analyses for reconstructing population history. In the case of *D. frontalis*, gaining
508   insights into the species' recent and rapid range expansion to the northeastern United States is of
509   considerable interest owing to the economic and ecological damage this may cause. Accordingly,
510   here we systematically explored alternative explanations for the apparent absence of clusters or
511   clines reported by Havill et al. (2019).

512

513   *Clusters*

514

515   Regardless of whether we increased the number of microsatellite loci (augmented
516   dataset) or analyzed subsets of loci with more favorable signal-to-noise ratio (low noise and high
517   signal datasets), no spatially abrupt genetic subdivisions were detected. Using simulations,

518   Cullingham et al. (2020) showed that in the presence of even quite low levels of on-going gene
519   flow (i.e., migration rate, $m \geq 0.005$), the best-fit number of clusters inferred using
520   STRUCTURE may be systematically underestimated, leading to erroneous inferences of $K=1$.
521   Following the recommendations of those authors and others, we compared outcomes across
522   several analytical methods with different assumptions and found that the inferred absence of
523   discrete clusters was robust (Table 5). We interpret this to indicate that either *D. frontalis* does
524   not exhibit discrete population structure, or gene flow is sufficiently high that such structure is
525   not detectable with our data. Other types of molecular markers, such as single nucleotide
526   polymorphisms (SNPs) assayed via sub-genomic sampling methods that are applicable to non-
527   model species (e.g., Elshire et al. 2011) could reveal subtle fine-scale spatial-genetic structure.
528   For example, compared to inferences from microsatellites, SNPs have identified additional
529   structure in the mountain pine beetle, *D. ponderosae* (Batista et al. 2016; Janes et al. 2018). That
530   said, although only 4× to 12× more bi-allelic SNPs than multi-allelic microsatellite loci might
531   offer comparable resolution of population structure (e.g., Liu et al. 2005; but see Haasl &
532   Payseur 2011 for a potentially much higher ratio), microsatellites have desirable properties for
533   assignment tests, kinship analyses and estimating heterozygosity, and therefore remain valuable
534   (Narum et al. 2008).
535
536   *Clines*
537
538   When using sampling sites as the unit of analysis, we found that microsatellite dataset
539   composition did not impact conclusions about the absence of range-wide gradients of genetic
540   differentiation. This outcome was also robust to the chosen genetic distance metric (Table 6).
541   Range expansion is often considered in the context of a series of founder events that each give
542   rise to new population with reduced genetic variation along the moving wave of advance, leading
543   to genetic differentiation from the source population. However, when range expansion is recent
544   or on-going, multi-directional, and/or effective population sizes are consistently large, there can
545   be a lag time or weakening of bottleneck effects (also see Roques et al. 2012 for an example of
546   how Allee effects at a low-density wave front can prevent successive loss of genetic diversity).
547   Likewise, if occasional long-distance dispersal is involved, new populations can establish in
548   leaps and bounds, far from the parent population (Nichols & Hewitt 1994; Hewitt 1996, 1999;
549   Ibrahim et al. 1996). Under any of these scenarios, IBD is not expected to be immediately
550   detectable. Indeed, the current expansion of *D. frontalis* to the northeastern United States is both
551   recent and on-going: the species spread throughout New Jersey by the mid- to late-2000's, to
552   Long Island in New York by 2014, and was trapped in Connecticut, Rhode Island and
553   Massachusetts by 2015–2016 (Dodds et al. 2018). Effective population sizes certainly have the
554   potential to be very large, but little is known about non-outbreak populations, making it difficult
555   to speculate about whether genetic drift would be a strong driver of allele frequency differences
556   over increasing geographic distances. However, the capacity for long-distance dispersal by *D.*
557   *frontalis* has been well-documented: this could occur via active flight over several kilometers, or

558   via passive above-canopy wind-assisted dispersal over tens of kilometers (Jones et al. 2019, and
559   references therein). Ultimately, lack of IBD is not entirely unexpected, at least not in the newly
560   invaded portion of this species' current geographic range. However, other explanations exist. For
561   example, detection of IBD might be scale-dependent, whereby its signal fades at increasingly
562   larger geographic distances owing to an upper bound on the maximum attainable genetic
563   distance (Castric & Bernatchez 2003). Such non-linearity might be revealed by iterative
564   reanalysis using different distance thresholds (van Strien et al. 2015), provided that straight-line
565   geographic distance (cf. isolation-by-environment; Wang & Summers 2010) is the primary driver
566   of genetic divergence.
567
568   *Female-biased dispersal*
569
570   Despite overall weak spatial-genetic structure among eastern United States *D. fontalis*,
571   we did find significant individual-based IBD attributable to males (Table 6). Although this
572   inference of female-biased dispersal was limited to Lynch & Ritland's (1999) measure of
573   relatedness, it was robust to different compositions of microsatellite loci across datasets. Indeed,
574   this finding is concordant with life history traits of *D. frontalis*. Females are the pioneering sex,
575   who locate a suitable host before recruiting males via release of aggregation pheromones (Jones
576   et al. 2019). Furthermore, females are larger than males, and given that body mass is correlated
577   with stored energy reserves, their greater capacity for flight (mean distance of 3.4 km vs. 2.7 km
578   for males) in a flight mill experiment was attributed to sexual size dimorphism (Kinn et al.
579   1994). Notably, the notion that males are not strongly philopatric is supported by capture-mark-
580   recapture data, which showed no significant differences in dispersal between the sexes over
581   distances up to 1 km (Turchin & Thoeny 1993).
582
583   There are several reasons that could explain why individual-based IBD was detectable in
584   our study, yet population-based IBD was not. For instance, this may simply reflect an issue of
585   sample size (i.e., 2,415–32,385 pairwise comparisons among individuals vs. only 28–36 among
586   sampling sites depending on the dataset; but note the elevated pseudo-replication associated with
587   the former). Alternatively, a difference in the timescales over which spatial-genetic structure has
588   evolved and/or is assessed may be responsible. Indeed, individual-based genetic distances may
589   better reflect contemporary processes such as intra-generational dispersal, whereas population-
590   based metrics primarily measure historical connectivity given that most signal comes from the
591   accumulated multi-generational effects of dispersal and gene flow (Sunnucks 2000; Garrick et al.
592   2010; Epps & Keyghobadi 2015). In the context of this study, however, these explanations are
593   plausible only if the other two individual-based genetic distance metrics that we assessed (i.e.,
594   D_{kf} and D_{ps}) had low power. One potential advantage of Lynch & Ritland's (1999) $\hat{r}$ is that loci
595   with more alleles—particularly those with rare alleles—provide more information about
596   relatedness, such that highly polymorphic genetic markers such as microsatellites are very
597   powerful (Ritland 2000). That said, this remains speculative, as simulations have shown D_{ps} to

598   be more sensitive than $\hat{r}$ under a variety of IBD scenarios, albeit under simplified conditions
599   (Shirk et al. 2016).

600

601   *Reconciliation with previous work*

602

603   Our findings for *D. frontalis* are inconsistent with those of Schrey et al. (2011) who
604   reported both clusters and clines, in the form of an east-west division approximately coincident
605   with southern Appalachian Mountains and population-based IBD across the eastern United
606   States. Those authors used eight microsatellite loci, all of which were included in Havill et al.'s
607   (2019) 24-locus dataset, as well as the present study's 33-locus augmented dataset. Generally,
608   trade-offs between the number of loci and number of individuals generally favors adding loci
609   (Landguth et al. 2012), but given the shared loci, it seems unlikely that the marker set alone can
610   account for contrasting inferences. However, geographic and temporal sampling did differ.
611   Whereas Schrey et al.'s (2011) study design included 19 sites with an average of 63 individuals
612   per site (range: 26–100) collected between fall 2004 to spring 2005, ours included 9 sites with an
613   average of 28 individuals (range: 21–31) collected >10 years later (i.e., summer 2016 to summer
614   2017, except for Ponte Vedra, Florida, collected in summer 2013). Thus, although the overall
615   spatial scale of sampling was similar between studies (see Introduction), effort allocation was
616   not. Indeed, Landguth and Schwartz (2014) have shown that the optimal allocation of n
617   individuals for accurate estimation of F_{ST} changes depending on the extent of gene flow
618   limitation, with misallocations leading to either over- or under-estimation of F_{ST}, depending on
619   the circumstances. Furthermore, direct comparison between studies may be impacted by the fact
620   that Schrey et al. (2011) sampled *D. frontalis* during years of relatively low population densities
621   (albeit following a significant outbreak in 2001–2002) whereas our specimens were collected at a
622   time of considerable increase in trap captures in Louisiana and Alabama, and an active outbreak
623   in Mississippi (B. Sullivan, pers. comm.).

624

625   During *D. frontalis* outbreak years, pulse irruptions have unusual population dynamics
626   whereby beetles emerge from infested trees in several waves and inter-tree distance is an
627   important determinant of attack success, with crowded healthy trees being at greater risk owing
628   to the spill-over effects of aggregation pheromones (Hain et al. 2011). Conversely, in endemic
629   (non-outbreak) situations, *D. frontalis* persists at low densities and generally must travel further
630   after emergence to encounter trees already under stress (e.g., from lightning strike damage or
631   disease). Indeed, seasonal differences within a given year can also impact dispersal behavior,
632   whereby individuals have the greatest potential for long-distance dispersal in Fall and the lowest
633   potential in mid-Summer, owing to temporal differences in fat content levels (Turchin & Thoeny
634   1993). Ultimately, the impacts of inter- and intra-year differences in sampling across studies may
635   at least partly account for contrasting conclusions about the nature and strength of spatial-genetic
636   structure.

637

638   The idea that preexisting structure may have been "overwritten" either by a change in
639   outbreak status (i.e., from endemic to epidemic) and/or on-going expansion processes
640   underscores the importance of considering both and ecological and geographic context. For
641   example, James et al. (2015) showed that while the spruce budworm (*Choristoneura*
642   *fumiferana*), a cyclical irruptive pest in North America, exhibits weak spatial-genetic structure,
643   there is considerable nuance to this general finding. For example, the legacy of high connectivity
644   during past outbreaks can persist for some time afterwards, high admixture can be driven by
645   migrant adults sourced from relatively few sites, and different life stages (i.e., adults vs. larvae)
646   can concurrently exhibit different degrees of spatial genetic structure. Although the periodicity of
647   spruce budworm outbreaks is considerably longer than that of *D. frontalis* (~35 years vs. ~5−10
648   years, respectively) and the synchrony of irruptive populations may occur at different spatial
649   scales, there may be considerable similarity in the underlying processes that generate apparent
650   weak spatial-genetic structure. Future work on *D. frontalis* could attempt to distinguish between
651   migrant and resident individuals, and model range expansion dynamics and associated source-
652   sink dynamics, explicitly incorporating temporal effects.
653

Conclusions
655

656   The lack of discrete population differentiation and substantial IBD in *D. frontalis* in the
657   eastern United States has implications for management as the species' impacts spread north. In
658   contrast to alien pests that invade a region via long distance introduction by human transport, *D.*
659   *frontalis* is rapidly spreading into a previously unoccupied region contiguous to its native range.
660   For alien species, gene flow between native and alien populations only occurs via additional
661   introductions. Conversely, for *D. frontalis*, the gene flow between long established populations
662   in the southeastern United States and invading populations in New Jersey, New York, and
663   Connecticut is likely to be extensive and on-going. While it is still an open question whether *D.*
664   *frontalis* had been present in the northeastern areas at low densities but is only now causing
665   impacts due to population increases with less winter mortality, or if it truly represents a recent
666   arrival, the management implications could rely less on looking for differences in beetle
667   genetics, and more on environmental and host tree differences between the south and north.
668   Indeed, this is an approach that seems to be working (Dodds et al. 2018).
669

Acknowledgements
671

672   We thank USDA Forest Service for beetle specimens, DeAdra Newman for laboratory
673   assistance, Yale University's DNA Analysis Facility on Science Hill for genotyping services,
674   and Brian Sullivan, Dimitrios Avtzis, Marion Javal, and Patrick James for constructive feedback
675   on the paper.

676

References

678

679 Allendorf, F.W., Luikart, G., & Aitken, S.N. (2013). *Conservation and the genetics of*
680 *populations*. 2nd ed. John Wiley and Sons, Hoboken, NJ, USA.

681

682 Armendáriz-Toledano, F., & Zuñiga, G. (2017). Illustrated key to species of genus *Dendroctonus*
683 (Coleoptera: Curculionidae) occurring in Mexico and Central America. *Journal of Insect*
684 *Science*, 17, 34.

685

686 Ascunce, M.S., Yang, C.-C., Oakey, J., Calcaterra, L., Wu, W.-J., Shih, C.-J., Goudet, J., Ross,
687 K.G., & Shoemaker, D. (2011). Global invasion history of the fire ant *Solenopsis invicta*.
688 *Science*, 331, 1066–1068.

689

690 Batista, P.D., Janes, J.K., Boone, C.K., Murray, B.W., & Sperling, F.A.H. (2016). Adaptive and
691 neutral markers both show continent-wide population structure of mountain pine beetle
692 (*Dendroctonus ponderosae*). *Ecology and Evolution*, 6, 6292–6300.

693

694 Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., & Cavalli-Sforza, L.L.
695 (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*,
696 368, 455–457.

697

698 Bradburd, G.S., Coop, G.M., & Ralph, P.L. (2018). Inferring continuous and discrete population
699 genetic structure across space. *Genetics*, 210, 33–52.

700

701 Cavalli-Sforza, L.L., & Bodmer, W.F. (1971). *The genetics of human populations*. W.H.
702 Freeman, San Francisco, CA, USA.

703

704 Cavalli-Sforza, L.L., & Edwards, A.W.F. (1967). Phylogenetic analysis: models and estimation
705 procedures. *American Journal of Human Genetics*, 19, 233–257.

706

707 Chakraborty, R., De Andrade, M., Daiger, S.P., & Budowle, B. (1992). Apparent heterozygote
708 deficiencies observed in DNA typing data and their implications in forensic applications. *Annals*
709 *of Human Genetics*, 56, 45–57.

710

711 Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms
712 ascertaining spatial population structure: a new computer program and a comparison study.
713 *Molecular Ecology Notes*, 7, 747–756.

714

715 Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M., & Corander, J. (2013). Hierarchical and
716 spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and*
717 *Evolution*, 30, 1224–1228.

718

719 Corander, J. Waldmann, P., & Sillanpää, M.J. (2003). Bayesian analysis of genetic
720 differentiation between populations. *Genetics*, 163, 367–374.

721

722 Critchfield, W.B., & Little, E.L. Jr. (1966). Geographic distribution of the pines of the world.
723 USDA Forest Service, Miscellaneous Publication 991, Washington, DC.

724

725 Dieringer, D., & Schlötterer, C. (2003). Microsatellite analyser (MSA): a platform independent
726 analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, 3, 167–169.

727

728 Dodds, K.J., Aoki, C.F., Arango-Velez, A., Cancelliere, J., D'Amato, A.W., DiGirolomo, M.F.,
729 & Rabaglia, R.J. (2018). Expansion of southern pine beetle into northeastern forests:
730 Management and impact of a primary bark beetle in a new region. *Journal of Forestry*, 116,
731 178–191.

732

733 Duchesne, P., & Turgeon, J. (2012) FLOCK provides reliable solutions to the "number of
734 populations" problem. *Journal of Heredity*, 103, 734–743.

735

736 Earl, D.A., & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for
737 visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
738 *Resources*, 4, 359–361.

739

740 Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., & Mitchell,
741 S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity
742 species. *PLoS ONE*, 6, e19379.

743

744 Epps, C.W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: disentangling
745 historical and contemporary influences and inferring change. *Molecular Ecology*, 24, 6021–
746 6040.

747

748 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
749 using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611–2620.

750

751 Excoffier, L. & Ray, N. (2008). Surfing during population expansions promotes genetic
752 revolutions and structuration. *Trends in Ecology and Evolution*, 23: 347–351.

753

754 Frantz, A.C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian
755 methods to determine the genetic structure of a continuously distributed population: clusters or
756 isolation by distance? *Journal of Applied Ecology*, 46, 493–505.

757

758 Garrick, R.C. (2011). Montane refuges and topographic complexity generate and maintain
759 invertebrate biodiversity: recurring themes across space and time. *Journal of Insect*
760 *Conservation*, 15, 469–478.
761
762 Garrick, R.C., Caccone, A., & Sunnucks, P. (2010). Inference of population history by coupling
763 exploratory and model-driven phylogeographic analyses. *International Journal of Molecular*
764 *Sciences*, 11, 1190–1227.
765
766 Gernandt, D.S., Dugua, X.A., Vázquez-Lobo, A., Willyard, A., Letelier, A. M., de la Rosa,
767 J.A.P., Piñero, D., & Liston, A. (2018). Multi-locus phylogenetics, lineage sorting, and
768 reticulation in *Pinus* subsection Australes. *American Journal of Botany*, 105, 711–725.
769
770 Guillot, G., Estoup, A., Mortier, F., & Cosson, J.-F. (2005). A spatial statistical model for
771 landscape genetics. *Genetics*, 170, 1261–1280.
772
773 Guo, S.W., & Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion
774 for multiple alleles. *Biometrics*, 48, 361–372.
775
776 Haasl, R.J., & Payseur, B.A. (2011). Multi-locus inference of population structure: a comparison
777 between single nucleotide polymorphisms and microsatellites. *Heredity*, 106, 158–171.
778
779 Havill, N.P., Cognato, A.I., del-Val, E., Rabaglia, R.J., & Garrick, R.C. (2019). New molecular
780 tools for *Dendroctonus frontalis* (Coleoptera: Curculionidae: Scolytinae) reveal an east-west
781 genetic subdivision of Early Pleistocene origin. *Insect Systematics and Diversity*, 3, 2.
782
783 Havill, N.P., Griffin, B.P., Andersen, J.C., Foottit, R.G., Justesen, M.J., Caccone, A., D'Amico,
784 V., & Elkinton, J.S. (In press). Species delimitation and invasion history of the balsam woolly
785 adelgid, *Adelges* (Dreyfusia) *piceae* (Hemiptera: Aphidoidea: Adelgidae), species complex.
786 *Systematic Entomology*.
787
788 Hain, F.P., Duehl, A.J., Gardner, M.J., & Payne, T.L. (2011). Natural history of the southern
789 pine beetle. In: Coulson RN & Klepzig KD (eds.) *Southern Pine Beetle II*. General Technical
790 Report SRS-140. U.S. Department of Agriculture Forest Service, Southern Research Station,
791 Asheville, NC, USA.
792
793 Havill, N.P., Shiyake, S., Lamb Galloway, A., Foottit, R.G., Yu, G., Paradis, A., Elkinton, J.,
794 Montgomery, M.E., Sano, M., & Caccone, A. (2016). Ancient and modern colonization of North
795 America by hemlock woolly adelgid, *Adelges tsugae* (Hemiptera: Adelgidae), an invasive insect
796 from East Asia. *Molecular Ecology*, 25, 2065–2080.
797

798 Hewitt, G. (1996). Some genetic consequences of ice ages, and their role in divergence and
799 speciation. *Biology Journal of the Linnaean Society*, 58, 247–276.

800

801 Hewitt, G. (1999). Post-glacial recolonization of European biota. *Biology Journal of the*
802 *Linnaean Society*, 68, 87–112.

803

804 Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary.
805 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 183–195.

806

807 Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal*
808 *of Statistics*, 6, 65–70.

809

810 Ibrahim, K.M., Nichols, R.A., & Hewitt, G.M. (1996). Spatial patterns of genetic variation
811 generated by different forms of dispersal during range expansion. *Heredity*, 77, 282–291.

812

813 James, P.M.A., Cooke, B., Brunet, B.M.T., Lumley, L.M., Sperling, F.A.H., Fortin, M.-J.,
814 Quinn, V.S., & Sturtevant, B.R. (2015). Life-stage differences in spatial genetic structure in an
815 irruptive forest insect: implications for dispersal and spatial synchrony. *Molecular Ecology*, 24,
816 296–309.

817

818 Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., &
819 Andrew, R.L. (2017). The $K = 2$ conundrum. *Molecular Ecology*, 26, 3594–3602.

820

821 Janes, J.K., Worth, J.R.P., Batistia, P.D., & Sperling, F.A.H. (2018). Inferring ancestry and
822 divergence events in a forest pest using low-density single-nucleotide polymorphisms. *Insect*
823 *Systematics and Diversity*, 2, 3.

824

825 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
826 *Bioinformatics*, 24, 1403–1405.

827

828 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components:
829 a new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.

830

831 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,
832 Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A.
833 (2012). Geneious Basic: an integrated and extendable desktop software platform for the
834 organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649.

835

836 Kinn, D.N., Perry, T.J., Guinn, F.H., Strom, B.L., & Woodring, J. (1994). Energy reserves of
837 individual southern pine beetles (Coleoptera: Scolytidae) as determined by a modified
838 phosphovanillin spectrophotometric method. *Journal of Entomological Science*, 29, 152–163.

839

840   Kolbe, J.J., Glor, R.E., Schettino, L.R., Lara, A.C., Larson, A., & Losos, J.B. (2004). Genetic
841   variation increases during biological invasion by a Cuban lizard. *Nature*, 431, 177–181.
842

843   Landguth, E.L., Fedy, B.C., Oyler-Mccance, S.J., Garey, A.L., Emel, S.L., Mumma, M.,
844   Wagner, H.H., Fortin, M.-J., & Cushman, S.A. (2012). Effects of sample size, number of
845   markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology*
846   *Resources*, 12, 276–284.
847

848   Landguth, E.L., & Schwartz, M.K. (2014). Evaluating sample allocation and effort in detecting
849   population differentiation for discrete and continuously distributed individuals. *Conservation*
850   *Genetics*, 15, 981–992.
851

852   Lesk, C., Coffel, E., D'Amato, A.W., Dodds, K., & Horton, R. (2017). Threats to North
853   American forests from southern pine beetle with warming winters. *Nature Climate Change*, 7,
854   713–718.
855

856   Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide
857   polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, 6, S26.
858

859   Lombaert, E., Guillemaud, T., Cornuet, J.-M., Malausa, T., Facon, B., & Estoup, A. (2010).
860   Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE*, 5,
861   e9743.
862

863   Lynch, M., & Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers.
864   *Genetics*, 152, 1753–1766.
865

866   Mantel, N. (1967). The detection of disease clustering and a generalized regression approach.
867   *Cancer Research*, 27, 209–220.
868

869   Meglecz, E., Pech, N., Gilles, A., Dubut, V., Hingamp, P., Trilles, A., Grenier, R., & Martin, J.F.
870   (2014). QDD version 3.1: a user-friendly computer program for microsatellite selection and
871   primer design revisited: Experimental validation of variables determining genotyping success
872   rate. *Molecular Ecology Resources*, 14, 1302–1313.
873

874   Meurisse, N., Rassati, D., Hurley, B.P., Brockerhoff, E.G., & Haack, R.A. (2019). Common
875   pathways by which non-native forest insects move internationally and domestically. *Journal of*
876   *Pest Science*, 92, 13–27.
877

878   Miller, J.M., Cullingham, C.I., & Peery, R.M. (2020). The influence of a priori grouping on
879   inference of genetic clusters: simulation study and literature review of the DAPC method.
880   *Heredity*, 125, 269–280.
881

882   Narum, S.R., Banks, M., Beacham, T.D., Bellinger, M.R., Campbell, M.R., Dekoning, J., Elz,
883   A., Guthrie III, C.M., Kozfkay, C., Miller, K.M., Moran, P., Phillips, R., Seeb, L.W., Smith,
884   C.T., Warheit, K., Young, S.F., & Garza, J.C. (2008). Differentiating salmon populations at
885   broad and fine geographical scales with microsatellites and single nucleotide polymorphisms.
886   *Molecular Ecology*, 17, 3464–3477.
887

888   Nei, M. (1972). Genetic distance between populations. *American Naturalist*, 106, 283–292.
889

890   Nichols, R.A., & Hewitt, G.M. (1994). The genetic consequences of long distance dispersal
891   during colonization. *Heredity*, 72, 312–317.
892

893   Peakall, R., & Smouse, P.E. (2006). GenAlEx 6: genetic analysis in Excel. Population genetic
894   software for teaching and research. *Molecular Ecology Notes*, 6, 288–295.
895

896   Peakall, R., & Smouse, P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population
897   genetic software for teaching and research – an update. *Bioinformatics* 28, 2537–2539.
898

899   Perdereau, E., Bagnères, A.-G., Bankhead-Dronnet, S., Dupont, S., Zimmermann, M., Vargo,
900   E.L., & Dedeine, F. (2013). Global genetic analysis reveals the putative native source of the
901   invasive termite, *Reticulitermes flavipes*, in France. *Molecular Ecology*, 22, 1105–1119.
902

903   Price, T.S., Dogget, H.C., Pye, J.M., & Smith, B. (1997). *A history of southern pine beetle*
904   *outbreaks in the southeastern United States through 2004*. Georgia Forestry Commission,
905   Macon, GA, USA.
906

907   Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
908   multilocus genotype data. *Genetics*, 155, 945–959.
909

910   Pritchard, J. K., & Wen, W. (2003). Documentation for STRUCTURE Software: Version 2.
911   Chicago: University of Chicago Press. Available at:
912   http://web.stanford.edu/group/pritchardlab/software/structure2_1.html.
913

914   Prugnolle, F., & De Meeûs, T. (2002). Inferring sex-biased dispersal from population genetic
915   tools: a review. *Heredity*, 88, 161–165.
916

917   Pye, J.M., Holmes, T.P., Prestemon, J.P. & Wear, D.N. (2011). Economic impacts of the
918   southern pine beetle. In: Coulson RN & Klepzig KD (eds.) *Southern Pine Beetle II*. General
919   Technical Report SRS-140. U.S. Department of Agriculture Forest Service, Southern Research
920   Station, Asheville, NC, USA.
921

922   R Core Team (2020) R: a language and environment for statistical computing. Version 3.6.1,
923   Vienna, Austria. URL https://www.R-project.org/.
924

925   Ritland, K. (2000). Marker-inferred relatedness as a tool for detecting heritability in nature.
926   *Molecular Ecology*, 9, 1195–1204.
927

928   Roques, L., Garnier, J., Hamel, F., & Klein, E.K. (2012) Allee effect promotes diversity in
929   traveling waves of colonization. *Proceedings of the National Academy of Sciences of the United
930   States of America*, 109, 8828–8833.
931

932   Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J.K., & Feldman, M.W.
933   (2005). Clines, clusters, and the effect of study design on the inference of human population
934   structure. *PLoS Genetics*, 1, e70.
935

936   Rousset, F. (2008). Genepop'007: a complete re-implementation of the Genepop software for
937   Windows and Linux. *Molecular Ecology Resources*, 8, 103–106.
938

939   Rousset, F. & Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics*, 140,
940   1413–1419.
941

942   Russello, M.A., Kirk, S.L., Frazer, K.K., & Askey, P.J. (2012) Detection of outlier loci and their
943   utility for fisheries management. *Evolutionary Applications*, 5, 39–52.
944

945   Schrey, N.M., Schrey, A.W., Heist, E.J., & Reeve, J.D. (2008). Fine-scale genetic population
946   structure of southern pine beetle (Coleoptera: Curculionidae) in Mississippi forests.
947   *Environmental Entomology*, 37, 271–276.
948

949   Schrey, N.M., Schrey, A.W., Heist, E.J., & Reeve, J.D. (2011). Genetic heterogeneity in a
950   cyclical forest pest, the southern pine beetle, *Dendroctonus frontalis*, is differentiated into east
951   and west groups in the southeastern United States. *Journal of Insect Science*, 11, 110.
952

953   Selkoe, K.A., & Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using
954   and evaluating microsatellite markers. *Ecology Letters*, 9, 615–629.
955

956 Séré, M., Thévenon, S., Belem, A.M.G., & De Meeûs, T. (2017). Comparison of different
957 genetic distances to test isolation by distance between populations. *Heredity*, 119, 55–63.
958
959 Shirk, A.J., Landguth, E.L., & Cushman, S.A. (2017). A comparison of individual-based genetic
960 distance metrics for landscape genetics. *Molecular Ecology Resources*, 17, 1308–1317.
961
962 Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele
963 frequencies. *Genetics*, 139, 457–462.
964
965 Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., & Soltis, P.S. (2006). Comparative
966 phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15, 4261–4293.
967
968 Sullivan, B.T. (2011). Southern pine beetle behavior and semiochemistry. In: Coulson RN &
969 Klepzig KD (eds.) *Southern Pine Beetle II*. General Technical Report SRS-140. U.S. Department
970 of Agriculture Forest Service, Southern Research Station, Asheville, NC, USA.
971
972 Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and*
973 *Evolution*, 15, 199–203.
974
975 Szücs, M., Vahsen, M.L., Melbourne, B.A., Hoover, C., Weiss-Lehman, C., Hufbauer, R.A., &
976 Schoener, T.W.(2017). Rapid adaptive evolution in novel environments acts as an architect of
977 population range expansion. *Proceedings of the National Academy of Sciences of the United*
978 *States of America*, 114, 13501–13506.
979
980 Turchin, P., & Thoeny, W.T. (1993). Quantifying dispersal of southern pine beetles with mark–
981 recapture experiments and a diffusion model. *Ecological Applications*, 3, 187–198.
982
983 van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., & Shipley, P. (2004). Micro-Checker:
984 software for identifying and correcting genotyping errors in microsatellite data. *Molecular*
985 *Ecology Notes*, 4, 535–538.
986
987 van Strien, M.J., Holderegger, R., & Van Heck, H.J. (2015). Isolation-by-distance in landscapes:
988 considerations for landscape genetics. *Heredity*, 114, 27–37.
989
990 Wang, I.J., & Summers, K. (2010). Genetic structure is correlated with phenotypic divergence
991 rather than geographic isolation in the highly polymorphic strawberry poison-dart frog.
992 *Molecular Ecology*, 19, 447–458.
993
994 Weir, B.S., & Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population
995 structure. *Evolution*, 38, 1358–1370.

996

997 Whitney, K.D., & Gabler, C.A. (2008). Rapid evolution in introduced species, 'invasive traits'

998 and recipient communities: Challenges for predicting invasive potential. *Diversity and*

999 *Distributions*, 14, 569–580.

1000

1001 Williams, J.W., & Jackson, S.T. (2007). Novel climates, no-analog communities, and ecological

1002 surprises. *Frontiers in Ecology and the Environment*, 5, 475–482.

1003

1004 Zimmermann, C. (1868). Synopsis of the Scolytidae of America north of Mexico. *Transactions*

1005 *of the American Entomological Society*, 2, 141–149.

Characterization of microsatellite loci for *D. frontalis* from each of two local populations.

As in Havill et al. (2019), a pig-tail (5′-GTTT-3′) was added to the 5' end of each reverse primer (shown below). Abbreviations are: Eastern North America (ENA); observed (H_O) and expected (H_E) heterozygosity; forward (F) and reverse (R) primers. H_O values with an asterisk denote loci that showed significant departure from Hardy-Weinberg Equilibrium at the 0.05-level.

1 **Table 1**. Characterization of microsatellite loci for *D. frontalis* from each of two local populations. As in Havill et al. (2019),
2 a pig-tail (5′-GTTT-3′) was added to the 5' end of each reverse primer (shown below). Abbreviations are: Eastern North
3 America (ENA); observed (H_O) and expected (H_E) heterozygosity; forward (F) and reverse (R) primers. H_O values with an
4 asterisk denote loci that showed significant departure from Hardy-Weinberg Equilibrium at the 0.05-level.

Locus name	Primer sequences 5' to 3' (fluorescent label)	Repeat motif	All ENA populations		Sicily Island, Louisiana			Tombigbee, Mississippi		
			Allele size range (bp)	No. of alleles	H_O	H_E	No. of alleles	H_O	H_E	No. of alleles
SPB4422	F: (FAM)-ATCGACTTCGCACGCAAAAC R: GTTTCCGCTTTCACTCACTTTAATCAT	AC	226-246	7	0.259	0.236	5	0.179	0.165	3
SPB903595	F: (NED)-TTTATGTCTATGCCGGATGG R: GTTTGGACATTGACAAAATCGGAC	CAG	262-278	7	0.320	0.269	2	0.250	0.433	3
SPB180144	F: (PET)-ACTAATATTTCAGGTCCGCC R: GTTTGAGCTACTGAAAATTGCGAC	AC	172-182	6	0.185*	0.294	4	0.429	0.443	4
SPB265317	F: (FAM)-AAACATGTCGGGGAATCTAC R: GTTTGTTCATTAGCAGCAGGGATA	AT	376-395	10	0.593	0.707	5	0.593	0.713	5
SPB979494	F: (NED)-TGACATATGCGACATAAGGG R: GTTTGAAGTGTTTATTGTGCTCGG	ATC	192-225	8	0.630	0.502	4	0.679	0.561	5
SPB4155	F: (PET)-GATGCAGTGAAAGTGGCGTG R: GTTTGCCGATCTTTACCAACTCAAGC	ATG	95-116	8	0.481	0.519	4	0.593	0.562	4
SPB3702	F: (FAM)-AACGCTTCACATTTGCACCG R: GTTTCATCGGATAATCCTGCGGGA	CAC	93-117	10	0.185	0.173	4	0.321	0.284	4
SPB1278	F: (VIC)-TCAGATCTGAGACGACAAGAAAGA R: GTTTCCGGTCTGCAAATACGAGGT	AT	104-120	7	0.259	0.233	4	0.357*	0.390	5
SPB1534	F: (PET)-CGGGTGAAAGAGTTAGGGGA R: GTTTGCCCTTACGATCACAGGTACT	(CCA) GAA(C CA)	109-133	10	0.615*	0.641	6	0.643	0.645	6

MICRO-CHECKER assessment of null alleles and estimates of their frequency (r).

Twenty cases of putative null alleles were detected (marked by an asterisk) across 33 loci and seven outbred population (abbreviations follow Figure 1). For each locus, $r_{cumulative}$ is the sum of r values from each population with a significant excess of homozygotes, and this was used to rank-order loci from worst (1st) to best (equal 14th).

1   **Table 2**. MICRO-CHECKER assessment of null alleles and estimates of their frequency (r). Twenty cases of putative null
2   alleles were detected (marked by an asterisk) across 33 loci and seven outbred population (abbreviations follow Figure 1).
3   For each locus, $r_{cumulative}$ is the sum of r values from each population with a significant excess of homozygotes, and this
4   was used to rank-order loci from worst (1st) to best (equal 14th).

Locus name	Estimated null allele frequency (r) per population							$r_{cumulative}$	Locus rank
	AL	FL	GA	LA	MS-Hol	MS-Tom	PA		
SPB2727	0.000	0.000	0.000	0.116	0.273	0.000	0.000	0.000	14
Dfr-09	0.000	0.043	0.000	0.021	0.000	0.071	0.000	0.000	14
Dfr-16	0.001	0.000	0.096	0.000	0.008	0.000	0.000	0.000	14
SPB3731	0.000	0.000	0.140	0.008	0.000	0.104	0.000	0.000	14
Dfr-10	0.055	0.000	0.000	0.029	0.000	0.011	0.000	0.000	14
SPB2313	0.080	0.046	0.033	0.000	0.156	0.005	0.002	0.000	14
SPB0138	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	14
SPB1983	0.055	0.000	0.015	0.000	0.000	0.000	0.000	0.000	14
Dfr-24	0.179*	0.043	0.000	0.016	0.139*	0.000	0.011	0.318	5
Dfr-14	0.093	0.000	0.000	0.000	0.000	0.082	0.000	0.000	14
Dfr-17	0.010	0.013	0.082	0.000	0.000	0.050	0.000	0.000	14
SPB2613	0.033	0.327*	0.021	0.000	0.000	0.012	0.000	0.327	4
SPB1230	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	14
Dfr-18	0.032	0.000	0.024	0.000	0.021	0.000	0.000	0.000	14
SPB1875	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	14
SPB3013	0.042	0.000	0.000	0.000	0.000	0.000	0.136	0.000	14
SPB1272	0.017	0.000	0.000	0.000	0.000	0.000	0.043	0.000	14
SPB1284	0.000	0.041	0.000	0.000	0.248*	0.052	0.013	0.248	7
SPB1242	0.044	0.000	0.000	0.044	0.000	0.142	0.000	0.000	14
SPB2480	0.431*	0.134	0.000	0.208	0.324*	0.246*	0.404*	1.405	1
Dfr-06	0.048	0.044	0.046	0.054	0.069	0.099	0.054	0.000	14
SPB1507	0.162*	0.089	0.235*	0.085	0.197*	0.094	0.249*	0.843	2
SPB2187	0.014	0.000	0.071	0.058	0.000	0.148*	0.064	0.148	12
SPB1270	0.161*	0.106	0.000	0.000	0.156	0.000	0.000	0.161	10
SPB4422	0.000	0.000	0.000	0.000	0.000	0.000	0.239*	0.239	8
SPB903595	0.000	0.000	0.106	0.000	0.053	0.268*	0.000	0.268	6

SPB180144	0.024	0.025	0.000	0.228*	0.083	0.016	0.012	0.228	9
SPB265317	0.054	0.094	0.149*	0.088	0.000	0.092	0.083	0.149	11
SPB979494	0.000	0.000	0.078	0.000	0.000	0.000	0.068	0.000	14
SPB4155	0.053	0.000	0.426*	0.037	0.076	0.000	0.000	0.426	3
SPB3702	0.006	0.006	0.000	0.000	0.065	0.000	0.223	0.000	14
SPB1278	0.000	0.000	0.000	0.000	0.073	0.044	0.000	0.000	14
SPB1534	0.145*	0.000	0.000	0.021	0.026	0.002	0.069	0.145	13

Characterization of a proxy for microsatellite mutation per locus.

From each directly sequenced allele, the number of contiguous repeat units was recorded from the longest uninterrupted microsatellite region (imperfect or compound repeats were seen in three loci, marked with †), and allele length was determined by trimming sequences at the ends of primer annealing sites (alternative priming sites were inferred for two loci, marked with #). For each locus, all allele sizes observed across Eastern North American (ENA) populations (n = 255 individuals) were used to calculate the median allele size, and the associated number repeat units was extrapolated and then used to rank-order loci from most (1st) to least (33rd) potential for homoplasy.

1 **Table 3**. Characterization of a proxy for microsatellite mutation per locus. From each directly sequenced allele, the
2 number of contiguous repeat units was recorded from the longest uninterrupted microsatellite region (imperfect or
3 compound repeats were seen in three loci, marked with †), and allele length was determined by trimming sequences at
4 the ends of primer annealing sites (alternative priming sites were inferred for two loci, marked with #). For each locus, all
5 allele sizes observed across Eastern North American (ENA) populations (n = 255 individuals) were used to calculate the
6 median allele size, and the associated number repeat units was extrapolated and then used to rank-order loci from most
7 (1st) to least (33rd) potential for homoplasy.

Locus name	One directly sequenced allele			ENA population screening		Locus rank
	NCBI accession	No. of contiguous repeat units	Allele length (bp)	Median allele length (bp)	Median no. of contiguous repeat units	
SPB2727	PRJNA493650	7	144	141.5	6.88	19
Dfr-09	EF126297	10	109	104.5	9.59	5
Dfr-16	EF126300	11 †	191	183	10.54	3
SPB3731	PRJNA493650	6	145	142	5.88	31
Dfr-10	EF126298	13	158	149.5	12.30	2
SPB2313	PRJNA493650	9	296	292.5	8.89	9
SPB0138	PRJNA493650	6	115 #	116	6.05	27
SPB1983	PRJNA493650	7	143	142	6.95	18
Dfr-24	EF126305	15 †	151	150	14.90	1
Dfr-14	EF126299	9	188	190	9.10	8
Dfr-17	EF126301	9	122	126	9.30	6
SPB2613	PRJNA493650	7	283	291.5	7.21	14
SPB1230	PRJNA493650	7	219	227.5	7.27	13
Dfr-18	EF126302	7	130	136.5	7.35	12
SPB1875	PRJNA493650	6	110	118	6.44	22
SPB3013	PRJNA493650	6	154	156.5	6.10	25
SPB1272	PRJNA493650	9	263	259.5	8.88	10
SPB1284	PRJNA493650	8	130	126	7.75	11
SPB1242	PRJNA493650	6	146	144	5.92	30
SPB2480	PRJNA493650	6	162	164.5	6.09	26
Dfr-06	EF126295	10	139	139	10.00	4

SPB1507	PRJNA493650	7	146	149	7.14	15
SPB2187	PRJNA493650	6	196	194	5.94	29
SPB1270	PRJNA493650	9	243	246	9.11	7
SPB4422	PRJNA493650	7	244 #	230	6.60	21
SPB903595	PRJNA493650	5	275	272	4.95	33
SPB180144	PRJNA493650	7	174	177	7.12	16
SPB265317	PRJNA493650	6	386	389	6.05	28
SPB979494	PRJNA493650	7	215	205.5	6.69	20
SPB4155	PRJNA493650	7	107	108	7.07	17
SPB3702	PRJNA493650	6	109	103.5	5.70	32
SPB1278	PRJNA493650	6	108	110	6.11	24
SPB1534	PRJNA493650	6 †	116	120.5	6.23	23

Global F_{ST} values for each microsatellite locus, calculated across nine sampling sites (n = 255 individuals), and their rank-ordering from highest (1st) to lowest (33rd).

1 **Table 4**. Global F_{ST} values for each microsatellite locus, calculated across nine sampling sites (n = 255 individuals), and
2 their rank-ordering from highest (1st) to lowest (33rd).

Locus name	Global F_{ST}	Locus rank
SPB2727	0.025	11
Dfr-09	0.027	9
Dfr-16	0.023	15
SPB3731	0.035	4
Dfr-10	0.022	18
SPB2313	0.014	31
SPB0138	0.006	33
SPB1983	0.018	25
Dfr-24	0.024	12
Dfr-14	0.019	22
Dfr-17	0.028	8
SPB2613	0.018	24
SPB1230	0.012	32
Dfr-18	0.017	27
SPB1875	0.022	19
SPB3013	0.028	7
SPB1272	0.027	10
SPB1284	0.028	6
SPB1242	0.019	23
SPB2480	0.064	1
Dfr-06	0.015	29
SPB1507	0.017	26
SPB2187	0.036	3
SPB1270	0.023	16
SPB4422	0.023	14
SPB903595	0.031	5
SPB180144	0.017	28
SPB265317	0.023	17
SPB979494	0.023	13

SPB4155	0.039	2
SPB3702	0.021	20
SPB1278	0.015	30
SPB1534	0.019	21

Outcomes of genotypic clustering using STRUCTURE (i.e., non-spatial, individual-based), BAPS (i.e., spatial, group- or individual-based) and DAPC (i.e., non-spatial, individual-based).

For STRUCTURE analyses, the number of clusters (K) was inferred using both a quantitative (Evanno et al. 2005) and a qualitative (Pritchard & Wen 2003) approach. The maximum membership coefficient (Q) was used to identify putative "ghost clusters" in cases where the best-fit $K>1$. Means were calculated from either 20 (STRUCTURE) or 10 (BAPS) replicates. For DAPC analyses, abbreviations are Bayesian Information Criterion (BIC) and Principal Components (PCs), and Not Applicable (N/A). Dataset abbreviations are as follows: augmented dataset (ADS; 33 loci), low noise dataset (LNDS; 24 loci), high sensitivity datasets (HSDS; either 10 or 19 loci), males and females (m+f), males only (m), and females only (f).

1 **Table 5**. Outcomes of genotypic clustering using STRUCTURE (i.e., non-spatial, individual-based), BAPS (i.e., spatial,
2 group- or individual-based) and DAPC (i.e., non-spatial, individual-based). For STRUCTURE analyses, the number of
3 clusters (K) was inferred using both a quantitative (Evanno et al. 2005) and a qualitative (Pritchard & Wen 2003)
4 approach. The maximum membership coefficient (Q) was used to identify putative "ghost clusters" in cases where the
5 best-fit $K>1$. Means were calculated from either 20 (STRUCTURE) or 10 (BAPS) replicates. For DAPC analyses,
6 abbreviations are Bayesian Information Criterion (BIC) and Principal Components (PCs), and Not Applicable (N/A).
7 Dataset abbreviations are as follows: augmented dataset (ADS; 33 loci), low noise dataset (LNDS; 24 loci), high
8 sensitivity datasets (HSDS; either 10 or 19 loci), males and females (m+f), males only (m), and females only (f).

Microsatellite dataset	STRUCTURE					BAPS				DAPC			
	Evanno et al. (2005)		Pritchard & Wen (2003)		Maximum Q in each inferred cluster	Groups		Individuals		Best-fit K	BIC	No. of PCs retained	No. of clusters per site
	Best-fit K	Delta K	Best-fit K	Mean LnP(K)		Best-fit K	Mean log(ml)	Best-fit K	Mean log(ml)				
ADS m+f	3	10.28	3	-19161.63	0.31, 0.33, 0.94	1	-19891.68	1	-19891.68	2	577.66	80	2
ADS m	2	48.32	2	-13949.30	0.39, 0.92	1	-14533.15	1	-14533.15	2	419.22	120	2
ADS f	2	20.73	2	-5204.64	0.43, 0.87	1	-5575.48	1	-5575.48	1	162.61	N/A	N/A
LNDS m+f	3	5.19	1	-11792.72	0.45, 0.46, 0.47	1	-12153.79	1	-12153.79	3	451.90	80	3
LNDS m	2	87.50	2	-8534.31	0.37, 0.93	1	-8861.98	1	-8861.98	3	329.35	40	3
LNDS f	3	6.34	1	-3201.78	0.38, 0.38, 0.43	1	-3408.28	1	-3408.28	1	128.88	N/A	N/A
HSDS10 m+f	3	4.98	3	-5345.57	0.57, 0.67, 0.98	1	-5549.75	1	-5549.75	9	242.41	45	8 - 9
HSDS10 m	5	2.59	5	-3943.26	0.46, 0.47, 0.55, 0.73, 0.86	1	-4086.40	1	-4086.40	7	183.72	35	5 - 7
HSDS10 f	2	0.10	1	-1444.36	0.71, 0.80	1	-1529.06	1	-1529.06	4	71.63	20	3 - 4
HSDS19 m+f	3	20.73	3	-10933.57	0.35, 0.35, 0.94	1	-11338.45	1	-11338.45	4	432.41	80	4

HSDS19 m	4	2.28	1	-7939.17	0.37, 0.39, 0.45, 0.89	1	-8242.20	1	-8242.20	3	314.69	60	3
HSDS19 f	2	2.01	1	-3035.54	0.61, 0.74	1	-3234.04	1	-3234.04	2	125.00	25	2

Outcomes of Mantel tests of the significance of correlation between geographic and genetic distance.

Abbreviations are as follows: augmented dataset (ADS; 33 loci), low noise dataset (LNDS; 24 loci), high sensitivity datasets (HSDS; either 10 or 19 loci), males and females (m+f), males only (m), and females only (f). Significant *P*-values are marked with an asterisk (†, linearized).

1 **Table 6**. Outcomes of Mantel tests of the significance of correlation between geographic
2 and genetic distance. Abbreviations are as follows: augmented dataset (ADS; 33 loci),
3 low noise dataset (LNDS; 24 loci), high sensitivity datasets (HSDS; either 10 or 19 loci),
4 males and females (m+f), males only (m), and females only (f). Significant *P*-values are
5 marked with an asterisk (†, linearized).

Microsatellite dataset	Isolation-by-distance analysis *P*-values					
	Population-based metrics			Individual-based metrics		
	D_c	D_s	F_{ST}†	D_{kf}	D_{ps}	$\hat{r}$
ADS m+f	0.304	0.406	0.500	0.176	0.105	0.001*
ADS m	0.488	0.343	0.420	0.133	0.128	0.001*
ADS f	0.130	0.403	0.150	0.127	0.053	0.282
LNDS m+f	0.259	0.271	0.356	0.414	0.342	0.001*
LNDS m	0.379	0.280	0.253	0.343	0.465	0.001*
LNDS f	0.152	0.455	0.264	0.294	0.191	0.443
HSDS10 m+f	0381	0.380	0.357	0.330	0.243	0.001*
HSDS10 m	0.177	0.360	0.393	0.222	0.240	0.013*
HSDS10 f	0.419	0.451	0.276	0.457	0.287	0.406
HSDS19 m+f	0.464	0.390	0.451	0.332	0.505	0.001*
HSDS19 m	0.452	0.475	0.397	0.226	0.308	0.001*
HSDS19 f	0.287	0.481	0.174	0.283	0.154	0.031*

Figure 1

Map of eastern North America showing the locations of nine *D. frontalis* sampling sites (numbered yellow dots).

Pie charts associated with each site represent the proportion of males (black) and females (white) in the set of genotyped beetles. The geographic range of the species (green) is based on the distributions of suitable pine host tree species based on Critchfield and Little (1966), modified to reflect the current northern limits of *D. frontalis*.

1. Sicily Island, Louisiana (LA)
2. Homochitto, Mississippi (MS-Hom)
3. Holly Springs, Mississippi (MS-Hol)
4. Tombigbee, Mississippi (MS-Tom)
5. Talladega, Alabama (AL)
6. Woolford, Maryland (MD)
7. Goat Hill, Pennsylvania (PA)
8. Warwick, Georgia (GA)
9. Ponte Vedra, Florida (FL)
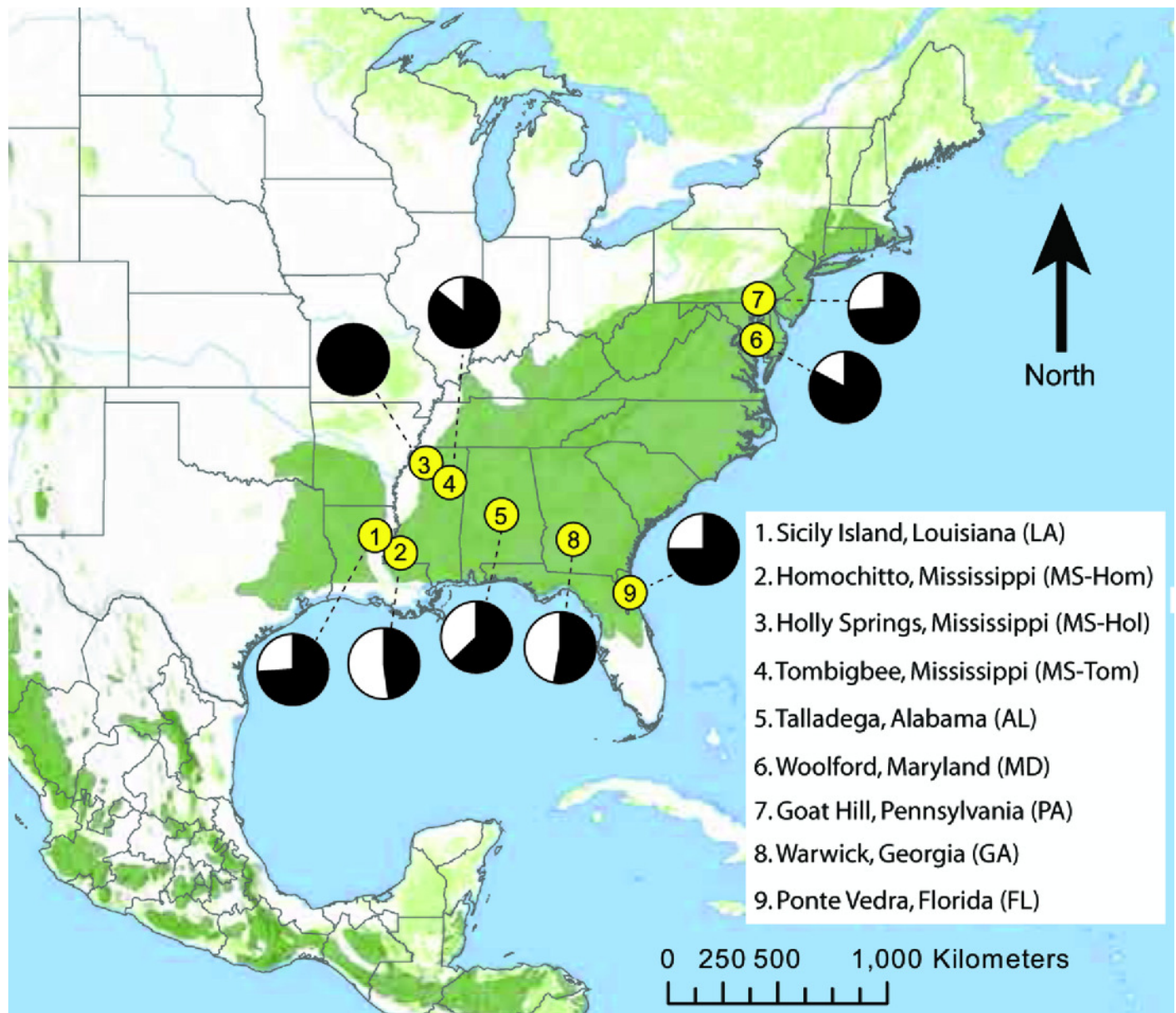
Figure 2

Plot showing the magnitude of reduction in F_{ST} between adjacent pairs of 27 rank-ordered loci, used to visually identify inflection points (two open circles) to guide 28 choice of thresholds for the number loci included in the high sensiti