New Molecular Tools for Dendroctonus frontalis (Coleoptera: Curculionidae: Scolytinae) Reveal an East–West Genetic Subdivision of Early Pleistocene Origin

Nathan P. Havill,1,6, Anthony I. Cognato,2 Ek del-Val,3 Robert J. Rabaglia,4 and Ryan C. Garrick5,

1USDA Forest Service, Northern Research Station, 51 Mill Pond Road, Hamden, CT 06514, 2Department of Entomology, Michigan State University, East Lansing, MI 48824, 3Instituto de Investigaciones en Ecosistemas y Sustentabilidad- Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México, Morelia, Michoacán, Mexico, 4USDA Forest Service, State & Private Forestry, Forest Health Protection, Washington, DC 20250, 5Department of Biology, University of Mississippi, Oxford, MS 38677, and 6Corresponding author, e-mail: nathan.p.havill@usda.gov

Abstract

Southern pine beetle, Dendroctonus frontalis Zimmermann, is a major native pest of pine trees in the southeastern United States, Mexico, and Central America. The species’ range has recently expanded north for the first time in recorded history. Accordingly, information about the timing of population divergence and past geographic range occupancy may provide an important yardstick for understanding rapid range expansions. Using 16 new and eight existing microsatellite loci, together with mitochondrial (COI) and nuclear (EF-1α and 28S) DNA sequence data, we characterized broad-scale patterns of genetic variation in D. frontalis, and estimated divergence times for the entire D. frontalis species complex. Molecular dating suggested a middle Miocene (ca. 12 million years ago [Mya]) origin of the D. frontalis species complex, and an early Pleistocene (ca. 2.2 Mya) divergence between eastern and western D. frontalis populations that are separated by the Tamaulipan mezquital ecoregion in southern Texas and northeastern Mexico. In the western D. frontalis group, there was additional differentiation between populations from Michoacán versus Arizona, suggesting that additional genetic structure could be uncovered in this region. In the eastern group, there was high genetic diversity, but little structure. There was no pattern of isolation by distance, and only weak population differentiation that distinguished populations from Georgia and Florida from the other eastern populations. Overall, our results suggest that eastern D. frontalis originated as a distinct group well before the last glacial period, but additional markers may be necessary to fully describe its contemporary rapid range expansion.

Key words: southern pine beetle, DNA sequence, genetic diversity, microsatellite, population genetics
Mexico up to Arizona, moving north in to California (Coleman et al. 2014); and 4) Soapberry Borer, *A. prionanus* Chevolat, native in Mexico, moving north into Texas (Billings et al. 2014). Owing to the native origins of these species, such cases cause a conundrum for classification of pest status and prediction of future impacts (Dukes et al. 2009, Nackley et al. 2017, Aoki et al. 2018, Tong et al. 2018).

Inferences about the timing of population divergence and past geographic range occupancy provide an important yardstick for the identification of rapid range expansions (Parmesan and Yohe 2003). Molecular data can contribute insights into these historical events as well as contemporary population processes that shape the spatial distribution of genetic diversity, and by extension, future range expansion and evolutionary potential (e.g., Ascunce et al. 2011, Janes et al. 2014, Havill et al. 2016). However, to date, few molecular tools exist for *D. frontalis*, and genetic variation across the species’ entire geographic range has not been assessed with modern DNA-based methods. Here we begin to redress this knowledge gap.

Existing phylogenetic and population genetic inferences for *Dendroctonus* species have mostly been based on DNA sequence data, plus some studies involving isozymes (Anderson et al. 1979, Nakmoong et al. 1979, Roberts et al. 1987), Amplified Fragment Length Polymorphism markers (Mock et al. 2007, Allender et al. 2008), microsatellites (Maroja et al. 2007, Schrey et al. 2007, 2008, Davis et al. 2009, Tærum et al. 2016), and single nucleotide polymorphism loci (Janes et al. 2014, Dowlle et al. 2017, Bracewell et al. 2018, Trevoy et al. 2018). Sequence data from the mitochondrial cytochrome oxidase subunit I (COI) gene have been used to reconstruct phylogenetic relationships among *Dendroctonus* species (Kelley and Farrell 1998, Reeve et al. 2012, Victor and Zuñiga 2016), to justify the recognition of new taxa (Armendáriz-Toledano et al. 2015), test hypotheses about the origin of *D. valens* LeConte introduction into China (Cognato et al. 2007), and to understand the phylogeographic histories of *D. pseudotsugae*, (Ruiz et al. 2010), *D. ponderosae* (Cullingham et al. 2012), and *D. mexicanus* Hopkins (Anducho-Reyes et al. 2008). Other genes, such as elongation factor 1 alpha (EF-1α) and the nuclear large ribosomal subunit (28S), have been sequenced for some *Dendroctonus* species but only in the context of broader Curculionidae or Scolytinae phylogenies (e.g., Sequeira et al. 2000, Jordal et al. 2011, McKenna et al. 2015). Taken together, these studies indicate that *Dendroctonus* species tend to be genetically variable with some population clusters corresponding to morphologically cryptic species (Avtris et al. 2012).

There is subtle morphological variability among *D. frontalis* and other closely related North and Central American *Dendroctonus* species (hereafter, the *D. frontalis* species complex; Lanier et al. 1988), which has led to taxonomic uncertainty and repeated revision. *Dendroctonus frontalis* was first described by Zimmermann (1868). However, Dietz (1890) later synonymized *D. frontalis* and *D. brevicornis* LeConte, and described a new species, *D. approximatus*. Shortly thereafter, Blandford (1897) described a new species, *D. adjunctus*, from specimens collected in Guatemala. *Dendroctonus frontalis* and *D. brevicornis* were then resurrected by Hopkins (1902), who also described two new species: *D. arizonicus* from Arizona and New Mexico, and *D. mexicanus* from Mexico (Hopkins 1905). These latter two species were synonymized with *D. frontalis* by Wood (1963), who later resurrected *D. mexicanus*, and described a new species, *D. vitei*, from Guatemala (Wood 1974). Most recently, Armendáriz-Toledano et al. (2014) reported variation in DNA sequences, morphology, karyotypes, and mating behavior within *D. frontalis*, and subsequently erected a new species, *D. mesoamericanus* Armendáriz-Toledano & Sullivan, which has a geographic range from southern Mexico to Nicaragua (Armendáriz-Toledano et al. 2015). Thus, the *D. frontalis* species complex currently contains seven species: *D. frontalis*, *D. brevicornis*, *D. mexicanus*, *D. vitei*, *D. approximatus*, *D. adjunctus*, and *D. mesoamericanus*.

The geographic range of *D. frontalis* (Fig. 1) broadly consists of two regions, a western area that extends from Honduras to central Arizona (Cognato 2011), and an eastern area that extends from east Texas to southern New England (Lanier et al. 1988, Dodds et al. 2018). The shortest distance between the two ranges, between north-eastern Mexico and east Texas (Moser and Macías-Sámano 2000), is separated by the Tamaulipan mezquital ecoregion in southern Texas and northeastern Mexico, which consists of desert and xeric shrublands inhospitable to pine hosts. In the western range, pine hosts of *D. frontalis* are predominantly distributed at high elevations, creating spatially isolated ‘sky islands’ (Salinas-Moreno et al. 2004, Williams et al. 2008). This broad geographic range with many dis-junct habitats may limit gene flow and ultimately promote allopatric speciation; accordingly, we hypothesized that population structure exists between eastern and western populations of *D. frontalis*.

Here we report the development of new microsatellite loci for *D. frontalis*. We use these loci together with those previously developed by Schrey et al. (2007), as well as mitochondrial and nuclear DNA sequence data, to characterize broad-scale patterns of genetic variation within the species. We also reconstructed a dated phylogeny of the entire *D. frontalis* species complex in order to make inferences about the timing of lineage divergence to place these patterns into historical context.

## Methods

### Population Sampling and DNA Extraction

Beetles were collected from 11 sites (Fig. 1; Table 1) in Mexico and the United States using funnel traps baited with frontal in and alpha-pinene, then stored in ethanol at −20°C. Beetles were identified as *D. frontalis* using the key in Armendáriz-Toledano and Zuñiga (2017). The head and pronotum were removed from each individual, retained as a voucher, and deposited at the Yale Peabody Museum in New Haven, Connecticut. The pronotum was retained because it contains species-specific characters (Armendáriz-Toledano and Zuñiga 2017), and allows sex determination via the presence (female) or absence (male) of a mycangium. For each population, additional entire specimens were also deposited as vouchers. The rest of the insect was ground with a pestle, and genomic DNA was extracted using the E.Z.N.A. Tissue DNA Extraction kit (Omega Bio-Tek, Norcross, GA) following the manufacturer's protocol.

### Microsatellite Development

DNA from one male beetle collected from Hormochitito National Forest, Franklin County, Mississippi in September 2016 was used to prepare a genomic library using Ion Xpress Plus Library and PGM Sequencing 400 kits (ThermoFisher) at the Functional Genomics Laboratory at UC Berkeley. Sequencing was performed with an Ion Torrent Personal Genome Machine Sequencer (ThermoFisher) using an Ion 318 sequencing chip at Yale University’s DNA Analysis Facility on Science Hill. The raw sequences are available from NCBI Bio-Tek, Norcross, GA) following the manufacturer's protocol.

Microsatellite discovery and primer design were performed using QDD 3.1.2 (Meglecz et al. 2014) with default parameters. Loci that contained pure microsatellites with at least six uninterrupted repeats were selected for further examination. Seventy-two primer pairs, including the eight described by Schrey et al. (2007), were assessed...
for amplification success in a test panel of eight *D. frontalis* samples: one individual each from Ponte Vedra, Florida; Woolford, Maryland; Sicily Island, Louisiana; and Tombigbee, Mississippi; and two each from Flagstaff, Arizona, and Nuevo San Juan Parangaricutiro, Michoacán (see Table 1 for locality information). The 5′ end of each reverse primer was modified with a pig-tail sequence (5′-GTTT-3′; Brownstein et al. 1996), and the 5′ end of each forward primer was modified with an M13 tail (5′-TCCCAGTCACGACGT-3′; Schuelke 2000) to allow incorporation of a 6-FAM labeled M13 primer during polymerase chain reaction (PCR). Amplifications were performed in 10 µl volumes containing 1X PCR Buffer, 1.0 µl dNTPs (10 mM each), 0.8 µl MgCl₂ (25 mM), 0.1 µl BSA (10 mg/ml), 0.025 µl of forward primer (10 mM), 0.25 µl of reverse primer (10 mM), 0.05 µl of 6-FAM labelled M13 primer (100 mM), 0.10 µl Go Taq DNA polymerase (Promega), and 1.0 µl sample DNA. The following touchdown thermocycler profile was used: 95°C for 2 min (1 cycle), 95°C for 45 s, 61°C decreasing 2°C for each cycle for 30 s, and 72°C for 45 s (6 cycles), 95°C for 45 s, 51°C for 30 s, and 72°C for 45 s (30 cycles), and final extension of 72°C for 2 min (1 cycle). PCR products were run with a Liz 500 internal size standard (Gel Company) on an ABI 3730 sequencer (Life Technologies) at Yale University’s DNA Analysis Facility on Science Hill. Genotypes were then scored using the microsatellite plugin in Geneious 10.0.5 (Kearse et al. 2012).

Microsatellite loci that yielded amplification products with detectable peaks for at least three of the four eastern U.S. samples were further characterized with directly-labelled fluorescent forward primers, by screening 24 beetles per site collected from the following four locations: Flagstaff, Arizona; Nuevo San Juan Parangaricutiro, Michoacán; Sicily Island, Louisiana; and Tombigbee, Mississippi. For each location, Arlequin 3.5 (Excoffier et al. 2005) was used to calculate alleles per locus, observed and expected heterozygosity, and to test for departures from Hardy–Weinberg equilibrium (HWE) and linkage equilibrium. The potential for false positives owing to multiple comparisons was accounted for using the method of Benjamini and Hochberg (1995) with a false discovery rate of 0.05. Loci that were variable, amplified reliably, displayed clear peaks, and were consistent with a Mendelian inheritance patterns for a diploid marker were retained for range-wide analyses.

### Range-Wide Microsatellite Analyses

For analysis of range-wide population structure within *D. frontalis*, 24–39 individuals from each of 11 local populations (Table 1) were genotyped using directly-labelled fluorescent forward primers as described above. As detailed in the Results section, 18 loci were deemed suitable for all populations, and 24 loci were suitable only for eastern U.S. populations (i.e., 11 newly developed [this study] plus seven previously published loci [Schrey et al. 2007] vs 16 new plus eight existing loci, respectively). Accordingly, separate analyses were performed for all 11 populations using 18 loci, and for the nine eastern U.S. populations using 24 loci. Geo-referenced microsatellite genotypes are available from the USDA Forest Service Research Data Archive, doi: https://doi.org/10.2737/RDS-2019-0002.
Table 1. Locality information and number of individuals sequenced for mitochondrial DNA (COI gene) and genotyped with nuclear microsatellites

<table>
<thead>
<tr>
<th>Locality</th>
<th>State</th>
<th>Country</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Date</th>
<th>Collector(s)</th>
<th>COI</th>
<th>Microsatellites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuevo San Juan Parangaricutiro</td>
<td>Michoacán</td>
<td>Mexico</td>
<td>19.4437</td>
<td>−102.1561</td>
<td>Mar. to May 2015</td>
<td>Ek del-Val</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>Sicily Island</td>
<td>Louisiana</td>
<td>United States</td>
<td>31.8649</td>
<td>−91.7398</td>
<td>18 April 2017</td>
<td>Jim Meeker</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Homochitto</td>
<td>Mississippi</td>
<td>United States</td>
<td>31.3818</td>
<td>−91.1523</td>
<td>20 Sept. 2016</td>
<td>JoAnne Barrett</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Holly Springs</td>
<td>Mississippi</td>
<td>United States</td>
<td>34.413</td>
<td>−89.3497</td>
<td>31 Mar. 2017</td>
<td>Jim Meeker</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>Tombigbee</td>
<td>Mississippi</td>
<td>United States</td>
<td>34.0376</td>
<td>−88.9318</td>
<td>10 April 2017</td>
<td>Jim Meeker</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>Talladega</td>
<td>Alabama</td>
<td>United States</td>
<td>32.7747</td>
<td>−86.984</td>
<td>20 July 2016</td>
<td>Larry Spivey</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>Woolford</td>
<td>Maryland</td>
<td>United States</td>
<td>38.5313</td>
<td>−76.2157</td>
<td>June 2016</td>
<td>Heather Disque</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Goat Hill</td>
<td>Pennsylvania</td>
<td>United States</td>
<td>39.7266</td>
<td>−76.0763</td>
<td>May–June 2017</td>
<td>Gina Peters, Paul Smith</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Warwick</td>
<td>Georgia</td>
<td>United States</td>
<td>31.7713</td>
<td>−83.894</td>
<td>3 April 2017</td>
<td>Chip Bates</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>Ponte Vedra</td>
<td>Florida</td>
<td>United States</td>
<td>30.2364</td>
<td>−81.3883</td>
<td>29 Aug. 2013</td>
<td>Jiri Hulcr</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

Arlequin was used to test for departures from HWE within local populations, and assess levels of differentiation among populations, as measured by $F_{st}$, using the infinite-allele model and 1,000 permutations. A neighbor-net network of $F_{st}$ values was reconstructed using the program SplitsTree 4.14.2 (Huson and Bryant 2006) to evaluate relationships among populations. Evidence for isolation by distance (IBD) among eastern U.S. populations was assessed via a Mantel test with 1,000 randomizations implemented in the adegenet R library (Jombart 2008). For the IBD test, genetic distances were generated using the Reynolds et al. (1983) method and Euclidian geographic distances were calculated using population coordinates. Population genetic structure was also analyzed using Bayesian clustering implemented in Structure (1983) method and Euclidian geographic distances were calculated using population coordinates. Population genetic structure was assessed via a Mantel test with 1,000 randomizations and graphical displays were created using Distruct 1.1 (Rosenberg 2004).

Range-Wide DNA Sequence Analyses

The 5' end of the mitochondrial COI gene was amplified using the primers LepF1 and LepR1 (Hebert et al. 2004). The 3' end of COI was amplified with primers Mod-TL2-N-3104 (5'-TCCA TTGCACTTTTCTGCCATTTTA-3'), modified from TL2-N-3014 (Simon et al. 1994), and Mod2-Cl-J-2183 (5'-CAACATYTTTGTATTTTTGG-3'), modified from Mod-Cl-J-2183 (Armendáriz-Toledano et al. 2014). Recovery of true mtDNA (cf. nuclear-mitochondrial pseudogenes) was assessed by translating DNA to amino acids and checking for open reading frames, and comparing to reference sequences in NCBI's GenBank database. A portion of the nuclear 28S rRNA gene was amplified using newly-designed primer Dfr-28S-Rev (5'-CAACGCTTTTTAGTTTGG-3') and Dfr-28S-For (5'-TGGAGGGAACCAAGCTACTA-3'), modified from 28S-b (Whiting et al. 1997). A portion of the nuclear EF-1α was amplified using newly-designed primers Dfr-EF1a-For (5'-TGGAGGGAAGAGCCAGGCAAAAGTAC-3') and Dfr-EF1a-Rev (5'-AACCAACCAGGCAATCTACATA-3'). The following thermocycler profile was used for all reactions: 95°C for 2 min (1 cycle), 95°C for 45 s, 48°C for 45 s, and 72°C for 1 min (30 cycles), and final extension of 72°C for 2 min (1 cycle). Design of new primers and modifications of existing primers were based on alignments of Dendroctonus sequences available in GenBank using Primer3 2.3.7 (Untergasser et al. 2012). Sequencing reactions were performed using the BigDye Terminator kit (Applied Biosystems, Foster City, CA) and analyzed on an Applied Biosystems 3730 automated sequencer at Yale University's DNA Analysis Facility on Science Hill. DNA sequences from each gene region were edited and aligned using Geneious, and deposited in GenBank under the following accession numbers: COI: MH997057–MH997371, EF-1α: MH997372–MH997387, and 28S: MH997388–MH997403.

The optimal values of $K$, and graphical displays were created using Distruct 1.1 (Rosenberg 2004).

...continued...
To estimate relationships among *D. frontalis* 5′ COI haplotypes, 24–39 samples from each of 11 local populations (Table 1) were analyzed using statistical parsimony (Templeton et al. 1992), implemented in TCS 1.21 (Clement et al. 2000) with a 90% connection limit. Sequences from all three genes, including concatenated 5′ and 3′ COI sequences, were used to reconstruct relationships among members of the *D. frontalis* species complex and estimate divergence times using the StarBEAST2 template in BEAST 2.5.0 (Drummond and Rambaut 2007, Ogilvie et al. 2017). Two or three *D. frontalis* individuals from each of seven local populations were included. Given the low genetic variation among individuals within a population as shown by microsatellite and COI data (see Results), it is unlikely that including additional individuals would change the resulting species relationships. Additional sequences from members of the *D. frontalis* species complex and from outgroup species were obtained from GenBank (Supp. Table S1). *Dendroctonus frontalis* sequences from GenBank represented samples collected in North Carolina (McKenna et al. 2015), Arizona (Duan et al. 2004), and Chiapas, Mexico (Armendáriz-Toledo et al. 2014). Outgroup species included two *Dendroctonus* species that are not part of the *D. frontalis* species complex, *D. jeffreyi* Hopkins, *D. ponderosa* and two members of Scolytinae that are not *Dendroctonus*, *Hylurgus ligniperda* (E.), and *Tomicus piniperda* (L.). Eastern and western *D. frontalis* were treated as separate species based on divergences evident in exploratory phylogenetic analyses of each gene separately using MrBayes 3.2.6 (Ronquist and Huelsenbeck 2003; Supp. Figs. S1–S3). Species tree analysis used unlinked site models for each codon position in COI and EF-1α, and for the EF-1α intron and 28S, with parameters determined using PartitionFinder 2.1.1 (LaNfear et al. 2017). The analysis used relaxed clocks with log-normal models of rate variation, and Yule tree prior with estimated birth rate. The clock and tree models were unlinked among genes. The MCMC chain was run for 50,000,000 generations, sampled every 10,000 generations, with the first 25% of sampled trees discarded as burn-in. Dates were calibrated using results from the fossil-calibrated phylogeny of bark and ambrosia beetles reported by Gohli et al. (2017). Accordingly, the *Tomicus-Hylurgus-Dendroctonus* root node was calibrated with a log-normal distribution centered at 35.45 million years ago (Mya), with mean (M) = 3.568 and SD = 0.05, and the *Hylurgus-Dendroctonus* node centered at 27.9 Mya (M = 3.329, SD = 0.05). This analysis was performed two times with different starting seeds to assess consistency of the estimated species trees.

**Results**

**Microsatellite Development**

Genomic sequencing resulted in 5,365,779 reads (median length: 262-bp) that formed the basis for microsatellite discovery. A total of 51,880 reads contained microsatellite repeats. Comparison and alignment of reads containing microsatellites resulted in 2,833 sequences that occurred a single time in the library and 689 unique consensus sequences that were aligned from multiple reads. In silico primer design was possible for 1,267 of these 3,522 putative loci. A final set of 64 high-priority primer pairs were selected that each had a single uninterrupted microsatellite sequence, at least eight repeat units, and primers greater than 20 bp away from the microsatellite motif. The first round of assessment indicated that 35 of these loci, plus the eight loci from Schrey et al. (2007; i.e., 43 loci in total), produced scoreable genotypes for at least three of the samples from the eastern United States that were included in the test panel. Thirty-three of these loci were polymorphic and further characterized in four populations. Twenty-four loci, that were most variable and in HWE for both eastern U.S. populations, were chosen for further analyses. These included 14 dinucleotides, nine trinucleotides, and one tetranucleotide, and exhibited three to 23 alleles per locus (Table 2). Eighteen of these 24 loci were suitable for application to all four populations. The six loci that were of more limited utility could not be scored for individuals from both Nuevo San Juan Parangaricutiro, Michoacán and Flagstaff, Arizona (loci SPB1230, SPB1507, and SPB2480), or could not be scored for Nuevo San Juan Parangaricutiro, Michoacán only (loci Dfr-18, SPB1875, and SPB3013). None of these loci deviated significantly from HWE in the two eastern populations after controlling for false discovery rate. For the 18 broadly applicable markers, four loci deviated in the Nuevo San Juan Parangaricutiro, Michoacán population and three in the Flagstaff, Arizona population, with observed heterozygosity lower than expected by HWE, perhaps suggesting the presence of null alleles (Table 2).

**Range-Wide Microsatellite Analyses**

A total of 324 individuals were genotyped for analysis of species-wide genetic structure using the 18 loci suitable for all populations. Mean pairwise *F*st between Nuevo San Juan Parangaricutiro, Michoacán and eastern U.S. populations was 0.314 (all 9 comparisons significantly different from zero), between Flagstaff, Arizona and the eastern populations was 0.333 (all nine comparisons were significantly different from zero), and among all eastern U.S. populations was 0.008 (11 of 36 comparisons significantly different from zero) (Supp. Table S2). The plots used to evaluate the number of clusters evident in the data are shown in Supp. Fig. S1. Structure analysis of all populations, suggested that genotypes clustered most prominently into *K* = 2 groups which separated Nuevo San Juan Parangaricutiro, Michoacán and Flagstaff, Arizona from the eastern U.S. populations. Analysis using *K* = 3 clusters revealed additional differentiation between the two western populations (Fig. 2A).

A total of 255 individuals from the eastern United States were genotyped using the 24 loci suitable for this geographic region. Mean pairwise *F*st among populations was 0.008. Eighteen of 36 comparisons were significantly different from zero (Supp. Table S3). Structure analysis of eastern U.S. populations suggested that genotypes clustered into *K* = 3 groups (Fig. 2B). The plots used to evaluate the number of clusters in the data are shown in Supp. Fig. S2. Here, weak genetic structure was evident via discrimination of Warwick, Georgia and Ponte Vedra, Florida, versus all remaining eastern populations. Neighbor-net network analysis (Fig. 3) also showed that these two populations were closely connected to each other, presumably due to ongoing gene flow, but were more distantly connected to remaining populations. Tombigbee, Mississippi, and Goat Hill, Pennsylvania also diverged somewhat from the remaining populations. The Mantel test did not show evidence of IBD in the eastern United States (*P* = 0.48), indicating moderate to strong genetic connectivity in this region.

**Range-Wide DNA Sequence Analyses**

The 5′ end of the COI gene was sequenced from 315 individuals. Sequences were A+T-biased and there were no indels or stop codons when translated to amino acids, indicating recovery of true mitochondrial DNA. For parsimony-based network analysis, the 90% connection limit was 16 mutational steps. Consistent with the microsatellite results, this haplotype network (Fig. 4) exhibited two major clusters: a western cluster that included samples from Nuevo Parangaricutiro, Michoacán and Flagstaff, Arizona, and an eastern cluster that included samples from Nuevo San Juan Parangaricutiro, Michoacán.
<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer sequences 5’ to 3’ (Fluorescent label)</th>
<th>Repeat motif</th>
<th>Size range</th>
<th>No. alleles</th>
<th>(H_O)</th>
<th>(H_E)</th>
<th>No. alleles</th>
<th>(H_O)</th>
<th>(H_E)</th>
<th>No. alleles</th>
<th>(H_O)</th>
<th>(H_E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dfr-06</td>
<td>F: GGAACCATCCTGATGGCCAAC (6-FAM) R: TAAAGGAACTGCTAATCCTGGGCTG (VIC)</td>
<td>GT</td>
<td>126–151</td>
<td>13</td>
<td>7</td>
<td>0.744</td>
<td>0.769</td>
<td>11</td>
<td>0.933</td>
<td>0.870</td>
<td>7</td>
<td>0.704</td>
</tr>
<tr>
<td>Dfr-09</td>
<td>F: TTAATCTTTTCTGCGGTAGC (VIC) R: TAAAGGAACTGCTAATCCTGGGCTG</td>
<td>GT</td>
<td>97–118</td>
<td>18</td>
<td>11</td>
<td>0.744</td>
<td>0.859</td>
<td>15</td>
<td>0.967</td>
<td>0.914</td>
<td>11</td>
<td>0.815</td>
</tr>
<tr>
<td>Dfr-10</td>
<td>F: AAATGCGTCGCTTTCGGCCAAC (6-FAM) R: GGGATAGGAAATTACATCCTGCAT</td>
<td>GT</td>
<td>136–165</td>
<td>23</td>
<td>9</td>
<td>0.744</td>
<td>0.718</td>
<td>10</td>
<td>0.833</td>
<td>0.866</td>
<td>11</td>
<td>0.815</td>
</tr>
<tr>
<td>Dfr-14</td>
<td>F: CAAAGTCCTTTAATCTGGCATC (VIC) R: GGGATAGGAAATTACATCCTGCAT</td>
<td>GT</td>
<td>181–213</td>
<td>9</td>
<td>7</td>
<td>0.692</td>
<td>0.633</td>
<td>4</td>
<td>0.700</td>
<td>0.638</td>
<td>5</td>
<td>0.704</td>
</tr>
<tr>
<td>Dfr-16</td>
<td>F: TTCGGACATAAGATAGCTCTC (NED) R: AACGCGTAACTCAGCTGTCGAG</td>
<td>GT</td>
<td>172–192</td>
<td>11</td>
<td>3</td>
<td>0.385</td>
<td>0.386</td>
<td>4</td>
<td>0.533</td>
<td>0.532</td>
<td>5</td>
<td>0.630</td>
</tr>
<tr>
<td>Dfr-17</td>
<td>F: GAGCGGACGGATCTAAATGCTGTG (VIC) R: CGGTCTATTTATTGCTAACCCCAAAC</td>
<td>GT</td>
<td>116–136</td>
<td>10</td>
<td>7</td>
<td>0.769</td>
<td>0.764</td>
<td>5</td>
<td>0.833</td>
<td>0.775</td>
<td>7</td>
<td>0.889</td>
</tr>
<tr>
<td>Dfr-24</td>
<td>F: TGGCACGGTGTCGCTTCTGCATC (PET) R: CCATGTAGGTGCAAACTC</td>
<td>GT</td>
<td>126–174</td>
<td>18</td>
<td>0.897</td>
<td>0.907</td>
<td>6</td>
<td>0.567</td>
<td>0.654</td>
<td>12</td>
<td>0.778</td>
<td>0.818</td>
</tr>
<tr>
<td>SPB0138</td>
<td>F: TGTACATCGTAGTATGGCTGCA (VIC) R: TGCCGTTCAGCAGATGGACGTAAT</td>
<td>AT</td>
<td>107–122</td>
<td>7</td>
<td>2</td>
<td>0.051</td>
<td>0.051</td>
<td>4</td>
<td>0.467</td>
<td>0.464</td>
<td>3</td>
<td>0.667</td>
</tr>
<tr>
<td>SPB1242</td>
<td>F: GAATGACCGACGACAAGGAGA (PET) R: TTTGCGCTGAAACCTGAAGGA</td>
<td>ATC</td>
<td>136–152</td>
<td>8</td>
<td>3</td>
<td>0.180</td>
<td>0.168</td>
<td>2</td>
<td>0.100</td>
<td>0.097</td>
<td>3</td>
<td>0.259</td>
</tr>
<tr>
<td>SPB1270</td>
<td>F: TGGACCCGATGTCGTTGCTGTG (PET) R: TGGACCCGATGTCGTTGCTGTG</td>
<td>AC</td>
<td>236–258</td>
<td>9</td>
<td>2</td>
<td>0.205</td>
<td>0.226</td>
<td>2</td>
<td>0.033*</td>
<td>0.033</td>
<td>6</td>
<td>0.407</td>
</tr>
<tr>
<td>SPB1272</td>
<td>F: CGTCGGCGTCAATAGGCTCCA (PET) R: TTCAGCGGCTTCCTGACAGAT</td>
<td>AG</td>
<td>254–267</td>
<td>8</td>
<td>6</td>
<td>0.447*</td>
<td>0.789</td>
<td>4</td>
<td>0.400</td>
<td>0.689</td>
<td>6</td>
<td>0.741</td>
</tr>
<tr>
<td>SPB1284</td>
<td>F: CTCCCAGGTAGGACGGAC (6-FAM) R: TGGGAAGATGACGCTGAAGGA</td>
<td>AC</td>
<td>122–130</td>
<td>6</td>
<td>4</td>
<td>0.282</td>
<td>0.330</td>
<td>4</td>
<td>0.633</td>
<td>0.566</td>
<td>5</td>
<td>0.481</td>
</tr>
<tr>
<td>SPB1983</td>
<td>F: GTGATAACCCGCGGAAGGAGA (NED) R: GTCGTTGCTCTTCTGCAAGA</td>
<td>AAC</td>
<td>131–160</td>
<td>16</td>
<td>11</td>
<td>0.718</td>
<td>0.721</td>
<td>7</td>
<td>0.367</td>
<td>0.358</td>
<td>7</td>
<td>0.630</td>
</tr>
<tr>
<td>SPB2187</td>
<td>F: ACACTGTCCTGCTGGTCTAC (PET) R: TGGTGACTGCTCAAGGACAGG</td>
<td>AAAT</td>
<td>197–205</td>
<td>9</td>
<td>4</td>
<td>0.180*</td>
<td>0.295</td>
<td>3</td>
<td>0.133*</td>
<td>0.434</td>
<td>4</td>
<td>0.407</td>
</tr>
<tr>
<td>SPB2313</td>
<td>F: ATGACCGAGATTCACATC (6-FAM) R: AGTGGTCCTGCATCAATCCA</td>
<td>AGG</td>
<td>279–306</td>
<td>10</td>
<td>5</td>
<td>0.454</td>
<td>0.639</td>
<td>2</td>
<td>0.467</td>
<td>0.499</td>
<td>6</td>
<td>0.704</td>
</tr>
<tr>
<td>Locus</td>
<td>Primer sequences 5’ to 3’ (Fluorescent label)</td>
<td>Repeat motif</td>
<td>Size range</td>
<td>No. alleles</td>
<td>( H_0 )</td>
<td>( H_e )</td>
<td>No. alleles</td>
<td>( H_0 )</td>
<td>( H_e )</td>
<td>No. alleles</td>
<td>( H_0 )</td>
<td>( H_e )</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SPB2613</td>
<td>F: TTGGTCGGAATGAGCTTGGG (NED) R: CAGCTCAACGGAACATCACATG (6-FAM)</td>
<td>AGC</td>
<td>266–308</td>
<td>15</td>
<td>11</td>
<td>0.795</td>
<td>0.853</td>
<td>8</td>
<td>0.833</td>
<td>0.805</td>
<td>8</td>
<td>0.741</td>
</tr>
<tr>
<td>SPB2727</td>
<td>F: TCAGTTAAACAAACAGTTGCGGT (NED) R: TTCAACATTCAATTATGCTTGTGT</td>
<td>AGT</td>
<td>134–149</td>
<td>5</td>
<td>2</td>
<td>0.180*</td>
<td>0.473</td>
<td>1</td>
<td>--</td>
<td>--</td>
<td>3</td>
<td>0.185</td>
</tr>
<tr>
<td>SPB3731</td>
<td>F: CAAATTAGCACAAAGACCGTATCA (PET) R: TGCAACATATTTCCGTTATCTGCT</td>
<td>AAC</td>
<td>139–145</td>
<td>3</td>
<td>3</td>
<td>0.154*</td>
<td>0.3094</td>
<td>2</td>
<td>0.233</td>
<td>0.413</td>
<td>2</td>
<td>0.481</td>
</tr>
<tr>
<td>SPB1230</td>
<td>F: TGGTCGCAATTCGCTTTGT (VIC) R: TTTACGCACCACATCGCC</td>
<td>AAT</td>
<td>217–239</td>
<td>10</td>
<td>16</td>
<td>0.633*</td>
<td>0.796</td>
<td>10</td>
<td>0.852</td>
<td>0.842</td>
<td>10</td>
<td>0.857</td>
</tr>
<tr>
<td>SPB1507</td>
<td>F: TCCTTTTGGCGCTGCTAAGTGA (NED) R: TAGCTCAACGCGATCATGTGTT</td>
<td>AT</td>
<td>140–229</td>
<td>16</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SPB1875</td>
<td>F: GCAATACGCTTTATGACCGC (PET) R: GACTCGAACCTTATAAGTCTTTTGTTA</td>
<td>AG</td>
<td>109–125</td>
<td>8</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SPB2480</td>
<td>F: GCCGTCCTTCTCATGATACCAAGGA (PET) R: AGAGCTGATCGGAAGTGATGT</td>
<td>AAG</td>
<td>156–174</td>
<td>8</td>
<td>2</td>
<td>0.269</td>
<td>0.419</td>
<td>3</td>
<td>0.286</td>
<td>0.481</td>
<td>3</td>
<td>0.286</td>
</tr>
<tr>
<td>SPB3013</td>
<td>F: GTAAGAGGAGGAGAAGTATG</td>
<td>AAG</td>
<td>148–201</td>
<td>6</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SPB1188</td>
<td>F: GACCGCAAGAGGCGCTGAA (FAM) R: ATTTGAGCTTCACAGGCTCTC</td>
<td>AC</td>
<td>94–104</td>
<td>5</td>
<td>3</td>
<td>0.182</td>
<td>0.246</td>
<td>3</td>
<td>0.375</td>
<td>0.377</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>SPB2417</td>
<td>F: ACACTATGGACAGAAAGACCA (FAM) R: CGGCGCGTGGAAGAAGTCAG</td>
<td>AT</td>
<td>295–299</td>
<td>3</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SPB3828</td>
<td>F: CAGTAGGTTGGTGAACATCAGCA (NED) R: CGGCTGCGAGCTTTGGTATG</td>
<td>ACG</td>
<td>125–152</td>
<td>9</td>
<td>7</td>
<td>0.476</td>
<td>0.539</td>
<td>4</td>
<td>0.304*</td>
<td>0.337</td>
<td>5</td>
<td>0.383</td>
</tr>
<tr>
<td>SPB4040</td>
<td>F: GAAATCATTGACAGAGTGGTGTGTGT (FAM) R: CCAAATCCAGGCTGTCAGG</td>
<td>AAG</td>
<td>113–119</td>
<td>3</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SPB4992</td>
<td>F: GTGCGCGGAGATGACCATCATT (FAM) R: TGGTTCGCAATCTCTCTCAGT</td>
<td>AT</td>
<td>179</td>
<td>1</td>
<td>1</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>Locus</td>
<td>Primer sequences 5′ to 3′ (Fluorescent label)</td>
<td>Repeat motif</td>
<td>Size range (bp)</td>
<td>No. $H_0$ alleles</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
<td>$H_0^*$</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------</td>
<td>--------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>SPB307</td>
<td>F: GGCTGGTCTGGATGCTGAC</td>
<td>AAG</td>
<td>159–183</td>
<td>6</td>
<td>0.337</td>
<td>0.471</td>
<td>5</td>
<td>0.435</td>
<td>0.487</td>
<td>1</td>
<td>0.033</td>
<td>0.033</td>
</tr>
<tr>
<td>SPB2348</td>
<td>R: AGCGGCCGAAAGACGATGGT</td>
<td>AT</td>
<td>157–166</td>
<td>6</td>
<td>0.337</td>
<td>0.471</td>
<td>5</td>
<td>0.435</td>
<td>0.487</td>
<td>1</td>
<td>0.033</td>
<td>0.033</td>
</tr>
<tr>
<td>SPB4902</td>
<td>F: ACCATGGATGACTGAGCAG</td>
<td>AC</td>
<td>175–187</td>
<td>7</td>
<td>0.391</td>
<td>0.621</td>
<td>6</td>
<td>0.409</td>
<td>0.762</td>
<td>5</td>
<td>0.100</td>
<td>0.429</td>
</tr>
<tr>
<td>SPB4574</td>
<td>R: CCAATCCACGCGGTTCGTC</td>
<td>AAT</td>
<td>183–215</td>
<td>7</td>
<td>0.391</td>
<td>0.621</td>
<td>6</td>
<td>0.409</td>
<td>0.762</td>
<td>5</td>
<td>0.100</td>
<td>0.429</td>
</tr>
</tbody>
</table>

Table 2. Continued

San Juan Parangaricutiro, Michoacán, plus Flagstaff, Arizona, and an eastern cluster that included samples from the nine populations located in the eastern United States. Samples from Nuevo San Juan Parangaricutiro, Michoacán included 19 haplotypes from 42 individuals, and Flagstaff, Arizona had four haplotypes from 30 individuals. Nuevo San Juan Parangaricutiro sequences were separated from Flagstaff sequences by at least 10 mutational steps. In contrast, the eastern cluster contained 59 haplotypes from 243 individuals and there was no discernable geographic localization of sequences.

The data set for the multi-gene species tree analysis included 3,265 nucleotides (COI: 1,495bp, 28S: 790bp, and EF-1α: 980bp), containing 492 (32.9%), 150 (19.0%), and 223 (22.8%) variable sites, for COI, 28S, and EF-1α, respectively. In preliminary phylogenetic trees estimated from each gene separately, samples of *D. frontalis* from the eastern United States formed a separate clade from samples collected in Arizona and Michoacán (Supp. Figs. S3–S5). The *D. frontalis* COI sequences generated in our study formed a well-supported clade with the other *D. frontalis* sequences from GenBank, which was reciprocally monophyletic with the morphologically similar *D. mesoamericanus* (Supp. Fig. S3). The phylogenetic placement of the *D. frontalis* sequences confirmed our initial species determination based on morphology. Furthermore, for COI, beetles from Arizona were in a separate clade from those collected in Michoacán. Notably, however, one sequence from GenBank (ID# Dfr-Duan2004; Supp. Table S1) from an individual collected in Flagstaff, Arizona in 2002 (Duan et al. 2004) grouped with the samples from Michoacán, suggesting some admixture between these groups (Supp. Fig. S3). As shown in the time-calibrated phylogeny (Fig. 5), the origin of the *D. frontalis* species complex was dated to 9.0 (95% CI: 5.9–12.5) Mya. *Dendroctonus mesoamericanus* was sister to *D. frontalis*, with an estimated divergence time of 4.7 (95% CI: 2.4–7.2) Mya, which was concurrent with a split between *D. mexicanus* and *D. vitei* estimated at 4.5 (95% CI: 1.8–7.7) Mya. Samples of *D. frontalis* from Arizona and Mexico diverged from populations in the eastern United States during the Pleistocene or earlier, estimated as 2.2 (95% CI: 1.0–3.4) Mya.

**Discussion**

In this study, we used newly developed microsatellite markers and broad geographic sampling to make inferences about the magnitude and spatial configuration of population structure in *D. frontalis*. These analyses of nuclear genotypic data were complemented by assessment of genetic structure embedded in more slowly evolving, maternally-inherited mitochondrial COI sequences to enable cross-validation. Finally, to place these results into historical context, a dated phylogeny of the *D. frontalis* species complex was reconstructed by generating DNA sequence data from additional loci from a representative subset of our samples and integrated with preexisting data from other species within the complex. Below, we highlight major findings, and place these into the context of earlier studies.

Our data show a distinct east–west split in *D. frontalis* as evidenced by divergent microsatellite genotypes and allele frequencies, and differences in mitochondrial and nuclear DNA sequences. The inability to amplify microsatellite loci in western samples for 10 of the 33 primer pairs that were designed using eastern *D. frontalis* genomic DNA provides additional evidence for lineage divergence, given that this is likely due to group-specific nucleotide substitutions in primer sites. Other studies have also hinted at significant genetic divergence between eastern and western *D. frontalis*. Anderson et al. (1979), using isozyme electrophoresis, reported that populations from Mexico and Arizona were significantly different from each
other and from eastern U.S. populations sampled in Virginia, Texas, and Georgia, which formed a single, largely undifferentiated group. Namkoong et al. (1979), also using isozymes, found that a population from Arizona was more divergent from eastern populations sampled in Georgia, Texas, Virginia, North Carolina, and Louisiana than these eastern populations were among each other. Also consistent with this pattern, Lanier et al. (1988) showed that mating crosses among *D. frontalis* from Arizona, Mexico, and the eastern United States were far more divergent than those with eastern United States populations.
States all produced fertile offspring, but there was reduced fertility in female offspring with one parent from Mexico and the other parent from Arizona or the eastern United States.

After tripling the number of microsatellite markers developed by Schrey et al. (2007) we still found little *D. frontalis* genetic structure in the eastern United States. Our analysis of population structure with additional loci provides confidence that this pattern is real and not simply the result of stochastic processes that can affect a small number of loci and thus potentially mislead population genetic inferences. Previously, the eight loci described by Schrey et al. (2007) did not detect genetic differentiation among forest stands <500 km apart in southern Mississippi (Schrey et al. 2008). However, they did detect weak genetic structure, roughly associated with eastern and western sides of the Appalachian Mountains (Schrey et al. 2011). Similarly, we found that genetic differentiation (*F_{ST}* among populations in the eastern United States, estimated using all 24 microsatellite loci, was very low (Fig. 3; Supp. Tables S3 and S4), and there was not a pattern of isolation by distance. Likewise, the COI haplotype network (Fig. 4) did not show evidence of spatial clustering. We did find some evidence using microsatellites of differentiation in the southeastern part of the range which may be a remnant signal of a Pleistocene glacial refugium for the beetles and their host trees in this region. Overall, we expect microsatellite data to be most informative over short to intermediate timescales, while DNA sequences should more readily retain signatures of deeper-time historical events and processes (Sunnucks 2000). The lack of strong genetic differentiation

---

**Fig. 4.** Network of mitochondrial COI haplotypes obtained from *D. frontalis* samples from 11 sites in Mexico and the United States. The area of each circle is proportional to the number of samples with that haplotype. Small black dots represent un-sampled (or extinct) haplotypes separating observed haplotypes. Haplotypes are color-coded to indicate the geographic population from which they were sampled (Table 1).
on the basis of both types of genetic data lends further evidence to the ability of *D. frontalis* to travel long distances, promoting genetic admixture. Detailed analysis of the timing and rate of *D. frontalis* expansion at the northern edge of its range in the eastern United States may therefore require additional markers and expanded geographic sampling.

The above findings have a number of parallels with those reported for the mountain pine beetle, *D. ponderosae*, for which genetic investigations have been extensive. For example, contemporary range expansions have been documented in both species, and genetic structure is characterized by the existence of at least two major groups, each spatially localized (although the axis of separation is oriented north-south in *D. ponderosae* versus east-west in *D. frontalis*). Furthermore, relatively weak substructure exists within clusters, indicative of long-distance dispersal and subsequent gene flow over distances in excess of 1,000 km (Cullingham et al. 2019, and references therein). An understanding of demographic history and range expansion dynamics in *D. ponderosae* has benefited from the development and application of large SNP data sets (Janes et al. 2018, and references therein), and we expect that additional insights into population structure and gene flow of *D. frontalis* will be possible with continued development of molecular markers for this species.

Our phylogeny of the entire *D. frontalis* species complex (Fig. 5) is consistent with other studies (Kelley and Farrell 1998, Sequeira et al. 2000, Reeve et al. 2012, Armendáriz-Toledano et al. 2014; Victor and Zuñiga 2016), but is the first to include a complete taxon sample, multiple loci, and to estimate divergence times in the whole group. It is not possible to analyze strict patterns of co-speciation between *Dendroctonus* species and their pine (*Pinus*) hosts because most *Dendroctonus* are generalists, using multiple host species (Kelley and Farrell 1998), and *Pinus* phylogeny is complicated by hybridization and introgression (Gernandt et al. 2018). Regardless, we comment on the apparent concordance of their evolutionary histories. Pines are thought to have first moved into the Mexican mountain ranges from the north during the middle Miocene (10–15 Mya; Perry et al. 1998, Graham 1999) when there was a marked decline in global temperature (Flower and Kennett 1994). This coincides with the origin of the *D. frontalis* species complex, which we dated to ca. 12 Mya. The preferred pine hosts of the *D. frontalis* species complex (Wood 1982, Salinas-Moreno et al. 2004, Armendáriz-Toledano and Zuñiga 2017) are among the North American hard pines (*Pinus* section *Triflora*) in subsections *Ponderosae* and *Australes* (Gernandt et al. 2005). The middle to late Miocene was an active time of diversification for these groups of pines (Leslie et al. 2012, Hernández-León et al. 2013, Saladin et al. 2017), and for the beetles (Fig. 5). It is interesting to note that there are two pairs of concurrent splits in the *D. frontalis* species complex that may have occurred allopatrically in response to the same climatic event. One pair of splits was dated to ca. 9 mya: (*D. adjunctus* (*D. approximata* + *D. brevicomis*)) and (*D. mexicanus* + *D. vitei* (*D. mesoamericanus* + *D. frontalis*)), and the other pair was dated to ca. 4.5 mya: (*D. mexicanus* + *D. vitei*) and (*D. mesoamericanus* + *D. frontalis*) (Fig. 5). The latter divergence events could be associated with the warming period that followed the late Miocene cooling from 5.4 to 7 Mya (Herbert et al. 2016).

Our molecular dating likely benefited from the integration of multiple independent DNA sequence markers, given that inherent variance in coalescences times can be more accurately characterized, and accounted for (Edwards and Beerli 2000). This analysis places the divergence between eastern and western *D. frontalis* early in the Pleistocene (ca. 2.2 Mya). This date is older than the one postulated by Moser and Macías-Sámano (2000) who suggested a very recent, 4,000- to 5,000-yr-old divergence between eastern and western *D. frontalis* based on similarities in morphology, behavior, and phoretic mite communities. They suggested that *D. frontalis* could have expanded into the eastern United States from Mexico at this time, when pines in the Mexican highlands may have been growing at lower elevations and pines in the eastern United States reached east Texas after advancing from a refugium in the southeastern

---

**Fig. 5.** Species tree of the *D. frontalis* species complex reconstructed using mitochondrial COI, and nuclear EF-1α, and 28S DNA sequence data (3258-bp total) using BEAST. Node labels are given as: mean height (million years ago) (95% confidence interval); node posterior probability. Geologic time scale bar indicates millions of years.
United States following the last glacial maximum. Our considerably older estimate for the split within *D. frontalis* is associated with a more ancient glacial period in the early Pleistocene (Halffter 1987, Mastretta-Yanes et al. 2015). Therefore, eastern *D. frontalis* was already well differentiated from the western group during the range contraction and expansion of its pine hosts during the last glacial period (Schmidtling 2007, Eckert et al. 2010).

The set of markers that we developed will be useful for continued exploration of broad-scale genetic structure in *D. frontalis*. Analysis of additional western populations, sampled from the sky islands of Arizona through the mountain ranges spanning Mexico to Nicaragua, is likely to uncover additional genetic structure in this region. It will be especially important to sample in the Sierra Madre Oriental in eastern Mexico which has a distinct biotic community from the Sierra Madre Occidental in western Mexico (Contreras-Medina et al. 2007, Corona et al. 2007), where we sampled in Michoacán. Beetles in eastern Mexico are likely to be more closely related to those in the eastern United States than those in western Mexico, and their inclusion might help refine inference of the biogeographic history of the species.

Ultimately, we hope to understand recent range expansion dynamics in the context of the species’ long-term population history. The deep divergence between eastern and western *D. frontalis* might correspond to biological differences that could make western populations an imperfect proxy for understanding spread dynamics in the eastern United States. Conversely, the lack of strong population structure in the eastern United States associated with long-distance dispersal and frequent admixture and might predict little difference between established populations and those in the recently expanded range. Analysis of additional populations, coupled with other approaches such as those that drawn on occurrence records and outbreak chronology, can provide a basis for managing susceptibility and mitigating loss of pine forests owing to pest damage (Cullingham et al. 2019).

**Supplementary Data**

Supplementary data are available at *Insect Systematics and Diversity* online.

**Acknowledgments**

This project was supported with funding from the U.S. Department of Agriculture Forest Service Northern Research Station. E.D.V. received research funding from Comisión Nacional Forestal, Consejo Nacional de Ciencia y Tecnología (CONAFOR-CONACYT)-234547. We are grateful to DeAdra Newman for assistance in the laboratory, and to Richard Hoffstetter, Jim Meeker, JoAnne Barrett, Larry Spivey, Heather Disque, Houping Liu, Gina Peters, Paul Smith, Brittany Barnes, Kamal Gandhi, Chip Bates, and Jiri Hulce for providing samples. We would also like to thank Dr. Jasmine Janes and two anonymous reviewers whose comments greatly improved this manuscript.

**Data Availability Statement**

Data from this study are available from the USDA Forest Service Research Data Archive: https://doi.org/10.2737/RDS-2019-0002

**References Cited**


interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J. Biogeogr. 42: 1586–1600.


